PUBLIC HEALTH IN THE 21ST CENTURY

GLOBAL ISSUES IN PUBLIC HEALTH

MADELON L. FINKEL, EDITOR

How to go to your page

This eBook contains three volumes. Each volume has its own page numbering scheme, consisting of a volume number, and a page number, separated by a colon.
For example, to go to page 5, of Volume 1, type Vol1:5 in the “page #” box at the top of the screen and click “Go.” To go to page 5, of Volume 2, type Vol2:5 in the "page #" box… and so forth.

Public Health in the 21st Century

This page intentionally left blank

Public Health in the 21st Century

Volume 1: Global Issues in Public Health

Madelon L. Finkel, Editor

Foreword by David J. Skorton, MD

PRAEGER

AN IMPRINT OF ABC-CLIO, LLC
Santa Barbara, California • Denver, Colorado • Oxford, England

Library of Congress Cataloging-in-Publication Data

Public health in the 21st century / Madelon L. Finkel, editor ; foreword by David J. Skorton.
p. ; cm.
Public health in the twenty-first century
Includes bibliographical references and index.
ISBN 978-0-313-37546-0 (hard copy : set : alk. paper) — ISBN 978-0-313-37548-4 (hard copy : vol. 1 : alk. paper) — ISBN 978-0-313-37550-7 (hard copy : vol. 2 : alk. paper) — ISBN 978-0-313-37552-1 (hard copy : vol. 3 : alk. paper) — ISBN 978-0-313-37547-7 (ebook : set) — ISBN 978-0-313-37549-1 (ebook : vol. 1) — ISBN 978-0-313-37551-4 (ebook : vol. 2) — ISBN 978-0-313-37553-8 (ebook : vol. 3)
1. Public health. 2. World health. 3. Health policy.
I. Finkel, Madelon Lubin, 1949– II. Title: Public health in the twenty-first century.
[DNLM: 1. Public Health. 2. Disease Management. 3. Health Policy. 4. World Health. WA 100]
RA424.8.P83 2011
362.1—dc22 2010039064

ISBN: 978-0-313-37546-0
EISBN: 978-0-313-37547-7

15 14 13 12 11 1 2 3 4 5

This book is also available on the World Wide Web as an eBook.
Visit www.abc-clio.com for details.

Praeger
An Imprint of ABC-CLIO, LLC

ABC-CLIO, LLC
130 Cremona Drive, P.O. Box 1911
Santa Barbara, California 93116-1911

This book is printed on acid-free paper ♾

Manufactured in the United States of America

To my husband Arnold,
whose ideas, insight, and, most of all,
support and love are so important to me.

This page intentionally left blank

Contents

This page intentionally left blank

Foreword

As we enter the second decade of the 21st century, myriad issues compete for the world's attention, from the continuing stresses of the deep and widespread recession, to personal and national security, to climate change. But no issue looms larger than that of public health. Within this general rubric fall multiple issues critical to the individual, the country, and the world. Infectious diseases, including malaria, multiple drug–resistant tuberculosis, pandemic influenza, and HIV/AIDS pose even greater threats because of enormously increased international travel. In the developing world, the traditional diseases of poverty—communicable disease, especially infectious diarrhea and other waterborne diseases, malnutrition, and inadequate maternal and child health care—and displacement and violence, which are the sequelae of political instability, increasingly are being joined by the ailments of excess such as obesity, diabetes, and atherosclerotic cardiovascular disease, making the burden infinitely greater. Long overdue recognition of the worsening plight of women, particularly but not only in resource-poor environments, compounds the ongoing dilemmas of maternal-child health. Even within rich societies, such as the United States, shocking health disparities stubbornly continue.

Despite these daunting challenges, the tools of public health are more robust than ever. In addition to the traditional tools of medicine and the social sciences, the use of molecular genetics techniques and advanced statistical analysis presents new opportunities for the student and practitioner of public health. This comprehensive work on public health thus appears at a most opportune time.

Including a carefully assembled combination of original work and important recent literature and covering a huge sweep of relevant problems, *Public Health in the 21st Century* succeeds admirably in bringing together much of the broad field into one work that should find its place as a reference for public health workers and academics as well as policy makers and those in the private sector, whether health care providers, insurers, or drug or device manufacturers. Dr. Madelon Finkel, an experienced and recognized expert in several aspects of public health and, importantly, in the pedagogy of public health, has assembled a most impressive group of writers on a huge variety of public health topics, covering everything from global population health, to special needs cohorts, to

health care policy, to the often-ignored topic of public health teaching strategies and tactics. Readers from across the spectrum of public health concerns will find thought-provoking material of great value.

I commend Professor Finkel and her many colleagues on bringing to fruition a work that undoubtedly will receive wide use.

David J. Skorton, MD
President, Cornell University

Acknowledgments

This three-volume set could not have been produced without the contributions of the authors who so generously took the time to research and write their respective chapter. Most of the authors are my friends and colleagues who gladly agreed to accept my invitation to be included in the effort. I thank each of the authors for their time, effort, and especially their friendship.

My editorial assistant, Sophia Day, was tremendously helpful in organizing the huge volume of material and keeping track of missing information. Editorial reviews of many of the chapters were graciously and professionally done by Dr. Rebecca Finkel, an author and former editor who also happens to be my accomplished daughter. Technical computer work and support was provided by Jean Policard of the Department of Public Health at Weill Cornell Medical College. His assistance was invaluable to me.

Many thanks to my editor, Debbie Carvalko of Praeger, who invited me to write this multivolume text. Her support and faith in my being able to deliver the goods on time was reassuring.

This page intentionally left blank

Introduction

Madelon L. Finkel, PhD

Compiling topics for inclusion in a multivolume text on public health at first seemed like a simple task. Because so many significant advances have been made in disease prevention and health promotion, and so many public health initiatives have been put in place over the years to improve health and well-being, deciding which topics to select proved more difficult than anticipated. Which ones should be included? Which ones are the most relevant, the most important to highlight? Narrowing the focus, but being as comprehensive and inclusive as possible, seemed the most prudent way to proceed. And, therein lay the problem. How was I to select from such a wide array of public health issues to produce a comprehensive text on current public health topics? In an effort to be both comprehensive and inclusive, I endeavored to select as many important and timely subjects as possible for these three volumes. For fear of overwhelming the readers with chapters on every conceivable public health issue, a careful selection was made to highlight topics that represent and reflect the field of public health's breadth and scope. As such, the three volumes include chapters on topics reflecting advances and progress in knowledge and practice as well as challenges that remain. Naturally, many more topics could have been included. The essays selected for inclusion, many written specifically for this multivolume set and others reprinted from the published literature, represent a broad overview of important public health issues in the 21st century.

Charles-Edward A. Winslow, a bacteriologist and professor of public health at the Yale School of Medicine from 1915 to 1945, proposed a definition of public health as

> the science and art of preventing disease, prolonging life, and promoting physical health and efficiency through organized community efforts for the sanitation of the environment, the control of community infections, the education of the individual in principles of personal hygiene, [and] the organization of medical and nursing service for the early diagnosis and preventive treatment of disease.[1]

To a large extent, his definition has not been changed or amended over the ensuing decades. Public health's focus, then and now, is to safeguard the public's health and to handle threats to public health.

Public health has its roots in antiquity. It has long been recognized that polluted water and air, inadequate waste disposal, overcrowding and a concomitant lack of hygiene, and lifestyle behavior contributed to the spread of disease. Moving away from the miasma theory of disease, which argues that most diseases are caused by miasma (Greek for "pollution," that is, a noxious form of "bad air"), to the germ theory of disease was an important step in disease prevention. The 19th century witnessed so many discoveries and advances in the fields of medicine and public health. Essentially, pre-20th-century efforts focused on the eradication of infectious diseases and improvements in hygiene and sanitation, which led to a dramatic increase in average life expectancy. For example, the science of epidemiology probably dates from Dr. John Snow's identification of polluted public water wells as the source of the 1854 cholera outbreak in London. Hungarian physician Ignaz Semmelweis, for example, successfully reduced infant mortality at a Vienna hospital by instituting a disinfection procedure. His findings were published in 1850, but his work was ill received by his colleagues, who unwisely discontinued the procedure. Disinfection did not become widely practiced until British surgeon Joseph Lister "discovered" antiseptics in 1865, helped significantly by the work of the French chemist and microbiologist Louis Pasteur and German physician and bacteriologist Robert Koch.

The early 20th century expanded the scope and complexity of public health concerns. High rates of infant mortality led to the establishment of maternal and child health programs that emphasized nutrition. The disgraceful state of the food processing industry was notably depicted in Upton Sinclair's book, *The Jungle*. The book dealt with conditions in the U.S. meat-packing industry, causing a public uproar that partly contributed to the passage of the Pure Food and Drug Act and the Meat Inspection Act in 1906. High rates of occupational injuries and occupational-related diseases led to programs for industrial hygiene and occupational health, but it was not until 1970 that the U.S. Occupational Safety and Health Administration (OSHA) was created by Congress under the Occupational Safety and Health Act. Its mission is to prevent work-related injuries, illnesses, and occupational fatality by issuing and enforcing rules called standards for workplace safety and health.

These and other public health efforts contributed substantially to a dramatic decrease in mortality. From 1900 to 1940, for example, mortality rates in the U.S. fell by 40 percent, and life expectancy at birth increased from 47 years to 63 years.[2] No other period in American history showed such a dramatic decline in overall death rates. Nearly all of this decrease can be accounted for by reductions in infectious diseases, which in 1900 accounted for 44 percent of deaths. Contributing to the decrease in infectious diseases was the implementation of clean water technologies, one of the most important public health interventions of the early 20th century. At the turn of the century, waterborne diseases accounted for one-quarter of reported infectious disease deaths in urban areas. By 1936, less than 20 percent of deaths were due to infectious diseases. Perhaps the greatest public health feat was the worldwide eradication of smallpox, a

highly contagious, serious viral disease that was a worldwide scourge. The last case was recorded in 1977 in Somalia, and the eradication was certified by the World Health Organization (WHO) in 1979.

By the mid- to late-20th century, achievements saw a shift in focus from acute infectious diseases to the treatment and prevention of the growing burden of noninfectious, chronic diseases. English physician-researcher Sir Richard Doll and English epidemiologist and statistician Sir A. Bradford Hill pioneered the randomized clinical trial, and together were the first to demonstrate the connection between cigarette smoking and lung cancer. The focus on individual behaviors and risk factors (for example, antismoking campaigns) was an important step in addressing the *social determinants of disease*. The *new* public health sought to address the burden of chronic disease in a more comprehensive way by focusing on the effects of disease on vulnerable populations (for example, the elderly, the young, and the disabled), how health status differs among population groups (health inequalities), and how health care systems are organized and financed. Indeed, the challenges facing modern public health in the 21st century must be broad and inclusive, and focus on improvement in population health through the reduction of preventable diseases, both communicable and noncommunicable.

Looking back over the last century, public health is credited with adding 30 years to the life expectancy of people in the United States over the course of the 20th century; 25 years of this gain are attributable to advances in public health.[3] The Centers for Disease Control and Prevention (CDC) cataloged 10 of what it considered to be the most notable public health achievements based on the opportunity for prevention and the impact on death, illness, and disability.[4] These include (not ranked by order of importance) the following:

- Vaccination programs (as a result of widespread vaccine use, many of the infectious diseases that once killed so many have been almost eliminated);
- Fluoridation of drinking water (fluoride was first added to the public water system in 1945; tooth decay and tooth loss has declined substantially as a result);
- Occupational safety policies (since 1980, the rate of fatal occupational injuries has decreased by 40 percent);
- Access to safe, improved family planning and contraceptive services;
- Control of infectious diseases as a result of antibiotics, clean water, and improved sanitation;
- Food safety (safer and more healthful foods can be attributed to decreased microbial contamination and increased nutritional;
- Recognition of tobacco use as a health hazard;
- Motor vehicle safety (safety belts, child safety seats, motorcycle helmets, and engineering improvements in both vehicles and highways have helped reduce fatal motor vehicle accidents);
- Decline in deaths from coronary heart disease and stroke (lifestyle modifications and pharmaceuticals have led to a decline in deaths for these diseases); and
- Healthier mothers and babies as a result of better hygiene, prenatal health care and nutrition.

The 21st century presents new challenges. Largely preventable infectious diseases, such as tuberculosis (TB), polio, measles, and cholera continue to

plague millions of people around the globe, especially children. HIV/AIDS, which appeared on the scene in the mid-1980s, continues to be a major public health problem, although antiretroviral medications have done wonders in terms of extending life. Malaria, multidrug-resistant TB, and global outbreaks of viral diseases, most recently the H1N1 swine flu pandemic of 2009, continue to challenge public health efforts. That being said, chronic diseases such as diabetes and heart disease are now prevalent around the world. Obesity is not just a problem of the wealthy nations, as the increase in adult and childhood obesity in the developing world threatens to jeopardize progress. Statistics compiled by WHO show that chronic diseases are the largest cause of death in the world today and that global prevalence of all the leading chronic diseases is increasing, with the majority occurring in developing countries. Cardiovascular disease is already the leading cause of mortality in the developing world.[5] The increased burden of chronic diseases in countries that also have a high burden of infectious diseases is creating both a tremendous economic and public health strain. Furthermore, the recognition that health is affected by many factors, including genetics, economics, ethnicity and race, and geography, has necessitated a shift in focus in thinking. Public health in the 21st century must address these health inequalities to reduce the incidence of disease and improve health and well-being. Malnutrition, poverty, lack of access to health care, and so forth threaten to undermine the progress made in disease control and prevention.

In summary, over the past 150 years, much of the focus has been on disease control, understanding sources of contagion, and implementing programs to prevent the spread of disease. As scientific knowledge grew, public health's purview expanded to include maternal and child health care, health education, nutrition, aging of the population, the recognition of the role of behavioral factors in determining health, the impact of violence (domestic, civil, and international), health care disparities, and globalization. Indeed, increased globalization and technological advances have contributed to a worldwide economic, political, and social interdependence. In 1945, the United Nations Conference in San Francisco unanimously approved the establishment of a new, autonomous international health organization, the WHO, which came into being on April 7, 1948. The WHO was established as a specialized agency of the United Nations to serve as a coordinating authority on international public health issues.

Despite the progress made in improving the health, so much still remains to be done, not just in the United States, but also globally. In 2000, for example, 11.1 million children under the age of 5 died from preventable diseases such as diarrhea and acute respiratory infection.[6] These and other primarily preventable diseases kill more people each year than conflicts alone. Worldwide, poverty is one of the most significant causes of preventable mortality. Gender inequality persists and perhaps in some areas of the world actually has increased. Population growth remains a serious concern as the world's population has surged to 6.7 billion, most of the increase occurring within the last century. Environmental degradation and climate concerns have significant health implications. Each has a dimension that necessarily involves public health. As such, public health must be looked at in a global context if it is to be successful in fulfilling its mandate. Microbes have no boundaries, and we have seen over and over again,

localized outbreaks can quickly spread to national epidemics, and even worldwide pandemics.

Global health refers to health problems that transcend national borders and are of such magnitude that they have a global political, social, and economic impact. Assessing and measuring the impact of globalization on population health status should not be done in a vacuum; a global public health perspective needs to be integrated into health, social, and economic policies and programs to be effective. Reducing social and economic deprivation, reducing health inequalities, and improving health status go hand-in-hand. Domestic and international entities whose function and purpose is promulgating public health policy need to work together to achieve common goals. Collectively, progress can be made; individually, the effect is more muted. At their summit in 2000, heads of state of the G8 countries went on record as recognizing health as a global challenge and acknowledging that health is the "key to prosperity" and that "poor health drives poverty."[7] Following up on this challenge, G77 heads of state from 130 developing countries also expressed support for working toward the reduction of disease worldwide.[8] The motives and intent are laudable, but a decade later, the world still finds itself grappling with disease control and health prevention issues. Indeed, a WHO assessment of the capacity of 185 countries to prevent, conduct surveillance, and control disease showed that while health ministries had a high level of awareness of the issues, they had little or no allocation of significant resources to address the problems.[9]

Global nongovernmental organizations (NGOs) have played and continue to play a critical role in building capacity and sustainable development in specific areas of the world, although often the focus is narrow (for example, tobacco, TB, malaria, diet and nutrition, and so forth). Foundations, such as the Bill and Melinda Gates Foundation, have provided extraordinary sums of money and manpower to address pressing global health issues. Their importance and impact cannot be denied or ignored. These private investments in global health far exceed government assistance. The pharmaceutical industry also has the potential for being an effective player in the global health arena, but the industry is constantly criticized for not taking a greater role in the access to life-saving drugs, particularly in developing nations. The arguments are plentiful, ranging from focusing research and development efforts on health issues for rich countries to pricing drugs at unaffordable levels. That being said, despite the challenges that the pharmaceutical industry faces, it has been involved in a number of global health initiatives. Often, this means that companies are donating drugs, cutting prices, and developing partnerships with local governments and NGOs.

The World Bank and the United Nations play a major role in setting priorities for global health. The World Bank recognizes the negative effect of the increasing burden of disease, especially on the poor. Billions of dollars have been provided to countries for disease prevention. These efforts are crucially important as most developing countries have inadequate financing, lack of manpower, and poor infrastructure. Numerous UN organizations are specifically designed to focus on the global burden of disease. For example, the United Nations Population Fund, the United Nations Children's Fund, and many other UN agencies and organizations focus on providing assistance to the poorest

countries in an effort to "make a difference." The Millennium Development Goals (MDGs), also, are an agreed-upon set of goals that were developed in response to the world's main development challenges. They were drawn from the actions and targets contained in the Millennium Declaration that was adopted by 189 nations and signed by 147 heads of state and governments during the UN Millennium Summit held in September 2000. The MDGs are targeted to promote poverty reduction, education, maternal health, and gender equality, and to combat child mortality, AIDS, and other diseases. Poor countries pledged to govern better and invest in their people through health care and education. Rich countries pledged to support them, through aid, debt relief, and fairer trade. The MDGs represent a global partnership that has grown from the commitments and targets established at the world summits of the 1990s. The eight goals are to be achieved by 2015.[10]

Many stakeholders, public and private, are working toward similar goals, but it seems at times as if progress has taken one step forward and two steps back. While tremendous progress has been made over the past century, substantial challenges remain. Capacity development for prevention, treatment, and research remains weak; global economic factors impede progress; and the need for health systems change (delivery, financing, organization, and insurance coverage) remains unmet. The three volumes in this text were formulated to address public health issues from a national and global perspective. Volume 1 focuses on global population health issues, while volume 2 presents chapters on various aspects of determinants of health and disease, and volume 3 examines current public health policy issues, including ethics and human rights, public health education, and challenges we face as we enter the second decade of the 21st century. I made a concerted effort to include authors from around the world. Colleagues from Africa, Australia, Canada, China, Europe, India, Latin America, and New Zealand are well represented in this multivolume text. Their perspective and insight add a global dimension to the set.

VOLUME 1

Section 1 of volume 1 focuses on global population health issues. In their chapter on the global burden of disease, Kishore and Michelow carefully review the salient features of the global burden of disease, including its distribution and changing patterns over time. If current trends continue, diseases such as diarrhea, AIDS, TB, and malaria will become less important causes of morbidity and mortality as heart disease, cancer, diabetes, and traffic accidents increase in prevalence. Although the "burden" of a disease can be defined in a variety of senses, the consensus definition, particularly from the WHO, is a fairly specific one. The global burden of disease (GBD) as defined by the WHO is a comprehensive regional and global assessment of mortality and disability from 136 diseases and injuries and 19 risk factors. While useful, the thinking was that a better measurement of the GBD was needed, one that integrated morbidity, mortality, incidence, and prevalence into a single common metric that can be compared across time, space, and interventions. A new metric, the Disability-Adjusted Life Year (DALY), is a summary measure of population health, measured in units of time (years), combining estimations of both fatal and nonfatal health outcomes

(morbidity and mortality) to provide an estimate of the number of years of fully healthy life lost by an individual with a particular illness or condition. When DALYs are used to estimate the GBD, communicable diseases displace noncommunicable diseases as leading drivers of illness. The authors discuss the explosion of noncommunicable chronic diseases worldwide and the existing burden of communicable diseases, the combination of which poses a significant threat to the public's health. The challenge we face is how to best deal with the double burden of disease.

China and India together account for 37 percent of the world's population, about 6.8 billion. In 2025, India will surpass China in total population. India thus will have the distinction of being the world's most populous country. China's fertility rate is decreasing, whereas India's continues to increase. The United States is the third most populous country. Bongaarts focuses on population projections to the year 2050 for the world and major regions, and then identifies the demographic factors responsible for continued expansion of human numbers. Discussion focuses on policy options for slowing population growth in the developing world, where the growth continues to surpass that in the industrial world. Four main demographic factors contribute to future population growth: continued high fertility, declining mortality, young age structure, and migration. Bongaarts concludes by noting that the unprecedented pace at which the world's population has grown over recent decades has had an adverse impact on social and economic development, on health care, and on the environment. Despite substantial and partially successful efforts to reduce growth in the less developed countries, this expansion of human members is expected to continue at a rapid pace over the next decades with nearly all of this growth occurring in Africa, Asia, and Latin America. He advocates for three key strategies to reduce this growth rate: strengthen family planning programs to provide women with the knowledge and means to regulate their fertility; emphasize "human development," in particular education, gender equality, and child health; and encourage delays in subsequent childbearing.

The effect of urbanization on the public's health is discussed by Galea and Vlahov. The authors focus on the substantial change from how most of the world's population lives, reflecting on how the characteristics of the urban environment affect population health. The key factors affecting health in cities are considered within three broad themes: the physical environment, the social environment, and access to health and social services.

Continuing the theme of adverse effects of rapid population growth and urbanization, Brown and DeGaetano present a scholarly piece on the consequences of climate change on health. Concerns about recent changes in global climates and possible future trends on the health of the world's population are now considered important policy topics. With the election of Barack Obama, who has pledged a new era of leadership and responsibility to reduce the serious negative effects of climate degradation, the United States resumed its leading role in combating climate change and the adverse effects thereof. The United States is the world's largest source of cumulative emissions in the atmosphere, and as such, needs to lead the way for other nations to make a serious effort on climate change. Brown and DeGaetano make the case that climatic changes have, and will continue to have, direct negative health effects from altered weather patterns, but state

also that the indirect effects on agriculture and wider population systems are important factors for the GBD. Global warming (that is, melting of Arctic ice), extreme weather (for example, heat waves, cold spells), flooding (for example, Hurricane Katrina), erosion of ocean coastlines (that is, a result of extreme and heavy precipitation), and drought (for example, dust bowls) are leading to a disruption of food production and to disease. The authors note that exposure to infectious diseases has altered because of changes in temperature, humidity, rainfall, and sea-level rise. Specifically, some evidence of changes in the distribution of mosquito, tick, and bird vectors has been attributed to climate change. Mosquitoes, for example, can transmit diseases, such as malaria, dengue, yellow fever, and Japanese encephalitis, but their sensitivity to weather conditions can inhibit or enhance their efficacy as a vector. Malaria is spread by mosquitoes, which are inhibited from transmitting the parasite in cooler temperatures. Air quality and pollutants are affected by the weather and climate, and can cause negative health effects; the incidence of asthma has soared over the past decades. The authors caution that vulnerability to climate change will depend on responses to prevention, adaptation, adaptive capacity, mitigation, and future advances in disease control. It is clear that doing nothing will only make the situation worse than it already is.

The issue of global health and nutrition is a complex interplay of many factors ranging from politics to economics to food production policies to environmental degradation. Food is a basic human need. With roughly 1 billion humans suffering from overnutrition and a similar number unable to find enough food to subsist, no one seriously disagrees about the urgency of world hunger. One in six individuals does not get enough food to be healthy and to lead an active life. Hunger and malnutrition adversely affect physical and mental development; indeed, one might argue that hunger and malnutrition are leading risks to the health and well-being of individuals worldwide. Davison presents a comprehensive overview of the salient issues and focuses in particular on the interdependence of nutrition, economic development, and health. His discussion of the topic includes an assessment of the MDGs drafted to address the issue of alleviating hunger and malnutrition and a brief overview of some of the programs designed to eliminate global nutrition disparities, including the Millennium Village Projects, the Grameen enterprises and "microcredit" initiatives in resource-poor countries, and the role that foundations play in providing the financial means to reduce poverty and, in turn, to alleviate hunger and malnutrition.

No matter how one defines "health," prevention and treatment of disease is an essential prerequisite for achieving health and well-being. Implicit in this is that the right medicine be available at the time and place of need. Reidenberg presents an overview of the WHO Essential Medicines Program. By definition, an "essential drug" is a drug needed to satisfy the health needs of the majority of the population. The essential drugs concept of purchasing a limited list of essential drugs for a health service and making them generally available has been accepted by 156 countries and most, if not all, donor organizations. The essential medicines idea was developed to help limited resource countries make choices to use their medical resources for the greatest good for the greatest number. Thirty years after the WHO initiated the Essential Medicines Model List, four out of five countries have adopted a national essential medicines list. More than 100 countries have a

national drug policy in place or under development. Furthermore, a network of 83 countries provides global monitoring for adverse drug reactions and as well as for potential safety problems. Regarding pricing, 30 years ago, virtually no publicly available price information was available, and few countries actively encouraged generic substitutions. In the 21st century, at least 33 countries provide such information.

Prevention and safety have long been an integral component of public health. Section 2 of volume 1 presents several essays on the topic. By focusing on ways to control risks, public health works toward making the environment a safer and healthier place in which to live. Silverstein presents an historical account of occupational health and safety in the United States. The Occupational Safety and Health Act of 1970 (OSHAct) declared that every worker in America is entitled to a safe and healthful workplace, and that employers are responsible for work being free from recognized hazards. Now, forty years later, many of the promises of the OSHAct have yet to be met. Silverstein reviews the history of occupational health and safety in the United States and exposes the barriers to OSHA's success (predominantly linked to the statutory design of regulation, inspection, and education) and the challenges that remain in preventing injury and illness at the workplace.

Hupert, Wattson, and Xiong present a sophisticated analysis of the complexity of planning for and responding to public health emergencies. Using the example of a large-scale aerosol anthrax exposure over an urban locale, they explore key determinants of health outcomes and health system surge capacity using several modeling techniques (state transition, queuing network). They suggest that such models can provide valuable insights for forecasting the logistical and staffing needs of large-scale prophylaxis campaigns for a range of intentional and natural disease outbreaks, such as the 2009 influenza A (H1N1) pandemic. While all model-based studies have their potential limitations, they may serve many functions in emergency preparedness and planning that cannot be provided through other means.

Food safety has periodically seeped into the consciousness of the lay public, almost always after a public tragedy involving tainted food. In 2006, there were 1,270 reported U.S. foodborne disease outbreaks, resulting in 27,634 illnesses and 11 deaths. Since then, many other well-publicized incidents have involved the safety of food products, including food recalls (berries from California, grapes from Chile, and so forth), contaminated beef or poultry, and recalls due to contamination (peanuts, almonds, and pistachio salmonella outbreaks occurred in 2009). Furthermore, public concerns over the use of food additives as well as use of pesticides have spurred interest in organically grown food products. Perhaps in response to the recent food outbreaks in the United States, the Food and Drug Administration (FDA) recently named a highly qualified food safety expert to be deputy commissioner for foods at the FDA. The newly created position is the first to oversee all the agency's numerous food and nutrition programs, and setting safety standards for produce is a top priority.

The article on foodborne illnesses by Tauxe, Doyle, Kuchenmuller, Schlundt, and Stein focuses on this important topic. Foodborne diseases are caused by a broad variety of pathogens and toxins. In their comprehensive and scholarly

article, the authors review the epidemiological, microbiological, and public health aspects of foodborne diseases resulting from the ingestion of contaminated foods and food products, and discuss the evolving public health approaches to the global challenges of foodborne infections. The global challenge of safeguarding the world's food supply is complicated by growing international trade, migration, and travel. Through the globalization of food marketing and distribution, contaminated food products can and do affect the health of people in numerous countries at the same time.

Pharmaceutical safety in the United States is under the purview of the FDA. Haas presents an overview of how the benefits and risks of pharmaceuticals are managed and discusses the implications for global drug safety. As more pharmaceutical products are manufactured in countries around the world (including Canada, China, and India), concerns about safety of the products are quite valid. The FDA does not have the money or the manpower to inspect each plant overseas; yet, the final product is distributed and marketed in the United States. Haas provides an excellent historical overview of key issues in drug safety, which led to regulation of the marketing of approved drug products, most notably the Food and Drug Amendment Act of 2007. The act mandated that product label changes for safety were to be imposed and executed promptly. To ensure an acceptable benefit-risk balance, the FDA was empowered to require additional studies or trials, and it could stipulate specific conditions limiting the market availability of a product to ensure its safe use. The FDA was instructed to promptly communicate evolving product safety concerns even if the available information was limited. In addition, the act mandated that virtually all clinical trials, regardless of sponsorship, be registered and that efficacy and safety results be publicly posted in a timely manner. The act created a major new safety information system (the Sentinel System) that would complement spontaneous adverse event reporting. Despite efforts to tighten the mechanisms to ensure drug safety, the system is not fail-safe. The goal for drug safety is to have a flexible and responsive system able to recognize potential risks early, collect information efficiently, and take action that is appropriate in the context of both benefits and risks.

Focusing on the needs of vulnerable populations is an important component of public health. Section 3 of volume 1 highlights health care issues of special population groups. Karpur, Bjelland, and Bruyère from the Employment and Disability Institute of Cornell University highlight the role of public health in improving the health, well-being, and overall quality of life for people with disabilities through the consideration of epidemiological trends in disability prevalence, issues related to health disparities, the legal and regulatory environment affecting access to preventive and curative health services, methods of measuring and tracking the population of people with disabilities, and specific priorities in public health. The Institute of Medicine (IOM) refers to disability as "the nation's largest public health problem," one that affects not only the health of people with disabilities, but also their immediate families and the population at large. Key issues for people with disabilities requiring attention in the U.S. public health system need to be addressed at the global level. It is estimated that there are approximately 650 million people with disabilities in the world with about 80 percent living in developing countries. The authors discuss various models and strategies to improve

health and well-being for people with disabilities, focusing on health disparities for people with disabilities; addressing the unique considerations for youth, women, and the aging with disabilities; and working toward an equitable access to health care, health care insurance coverage, health promotion, and prevention of secondary conditions—universal concerns that public health systems in all countries should be taking into account in the development of their national strategy.

Immigrant health care traditionally has largely been ignored by health policy makers. Yet, in 2009, an estimated 16 million children lived in immigrant families in the United States, representing one of the fastest-growing segments of the population. Clearly, policies and programs are needed to support immigrant parents and children, but the reality is haphazard at best. Mohanty, Woolhandler, Himmelstein, Pati, Carrasquillo, and Bor present compelling data based on the 1998 Medical Expenditure Panel Survey (MEPS) and found that immigrants have less access to health care and less health care use than do U.S.-born individuals. They also found that per capita health care expenditures for immigrants were far lower than expenditures for the U.S.-born. The study convincingly showed that the widely held assumption that immigrants consume large amounts of scarce health care resources is not supported by the data. The authors conclude that the low expenditures of publicly insured immigrants also suggest that policy efforts to terminate immigrants' coverage would result in little savings.

The provision of health care (or lack thereof) to those incarcerated has a long, sordid history. Finkel presents statistical evidence to illustrate the spectrum of health problems in correctional facilities. Inmate health and medical conditions range the gamut from minor (colds or viruses) to the significant (HIV/AIDS, TB). In addition to the communicable diseases, the prevalence of mental health and psychiatric diseases and substance abuse is higher among the prison population than the general population. The public health consequences of not paying attention to the health of prisoners can be quite significant; infectious diseases transmitted or exacerbated in prisons have the potential to become full-blown public health problems when prisoners return to their communities. The scope of this chapter provides an overview of the state of health among prisoners, assesses the provision of health care to those incarcerated, examines the policies regulating care of prisoners, including the challenges governments face in their ability to provide health and medical care to inmates, and discusses the pros and cons of having the private sector (privatization) involved in prison health care delivery.

Lesbian, gay, bisexual, and transgender (LGBT) health care also has received marginal interest and attention among policy makers and in the medical school curriculum. Medical education in the United States, both during medical school and in residency, is often unlikely to include adequate cultural competency related to the care of sexual orientation and gender identity minorities. A survey conducted to assess curricula in U.S. medical schools found that less than 3.5 hours were dedicated to teaching about health issues related to homosexuality. Part of the problem is the paucity of data on population demographics and health status for this population. For public health departments and providers to plan appropriate services for this vulnerable population, it is essential to have reliable data. Until recently, many of the research studies conducted in the LGBT community were community-based

studies using nonprobability sampling techniques. Radix and Mayer discuss the barriers to access to care as well as the health issues prevalent among the LGBT community. Of course, each group has its own set of health care needs, as lesbian health care is necessarily different from gay health care. The authors make the point that LGBT individuals have specific health needs that require targeted and culturally appropriate interventions.

The first ever surgeon general's report on oral health in the United States was published in 2000. The report highlighted a "silent epidemic" of dental and oral diseases, especially among the poor, the elderly, and children. Globally, too, oral disease burden and disability, especially in poor and marginalized populations, is a huge unmet issue. Oral health is much more than the pains of a toothache. Oral diseases such as dental caries, periodontal disease, tooth loss, oral mucosal lesions, and oropharyngeal cancers are major public health problems worldwide. Poor oral health has a profound effect on general health and quality of life. The burden of oral diseases and conditions is greatest among the economically disadvantaged, which include a disproportionately large number of racial and ethnic minorities and underserved populations. The major risk factors for oral disease are known and they are common with other chronic diseases: diet, smoking, alcohol, and risky behaviors. Canto and Cruz provide an epidemiologic overview of the state of oral health care as well as discuss preventive measures initiated to reduce dental caries, including exposure to fluoride (community water fluoridation, for example, has done much to reduce dental caries), use of dental sealants, practice of good oral hygiene, and reduction in sugar intake. The unmet need for dental care is a serious problem that needs to be acknowledged and addressed.

Taking care of the health care needs of the growing geriatric population is complex, challenging, and, to some extent, costly. Adelman, Finkelstein, Mehta, and Greene present an overview of the challenges of providing high-quality care to a rapidly aging population. They examine the medical, psychological, and social components of older age and explore the needs of this heterogeneous cohort. Issues such as dementia and Alzheimer's disease, elder abuse, ageist bias, the risks of polypharmacy, and long-term care issues are discussed.

Section 4 of volume 1 focuses on population-based prevention strategies. Adolescent substance abuse (alcohol, drugs, tobacco) has been well studied over the past decades; yet, the problem remains. Prevention and control programs have received considerable attention over the past decades as well. Botvin, Griffin, and Murphy, leaders in adolescent substance abuse prevention and cessation studies, raise a number of important issues related to adolescent substance abuse, including prevalence trends and types of prevention-based program modalities used by schools, families, and communities. The authors state that the most effective approaches target salient risk and protective factors, are guided by psychosocial theories regarding the etiology of substance use and abuse, and are implemented over many years. Many school-based prevention programs, for example, focus on skill-building in the area of drug resistance as well as life-skills training. While progress has been made in the field of substance abuse prevention, continued efforts must be made in the area of skill-building to prevent adolescents and children from taking drugs in the first place.

The issue of violence has been the subject of numerous reports by private and public organizations over the past decade. The public health consequences of all forms of violence are considerable as violence is associated not just with fatalities, but also with substantial morbidity and costs. It is estimated that in 2006 the health costs of violence (both fatal and nonfatal) in the United States exceeded $70 billion. Anderson and Sidel take a global approach to the discussion of violence and its sequela, and lay out a public health approach to violence prevention. They posit that the goals of public health—to prevent disease, and injury and premature death and to promote healthy living conditions for all—are identical to the goals of violence prevention. The disciplines and methods of public health—analyzing the causes of diseases, injuries, and premature deaths and of poverty and despair and determining methods to counter them—can strengthen efforts to prevent violence. And the ability of public health workers to gain trust both nationally and internationally can bring new skills and vigor to violence prevention.

Women, especially women in resource-poor nations, are an especially vulnerable group in terms of economics and in health care. In the industrial and developing world, gender-based violence (GBV) is endemic. Not only is it a major public health and human rights problem, but also for the victims it can, and most often does, have devastating personal, health, societal, and economic consequences. Meshkat and Landes eloquently delineate the types of GBV ranging from sexual, psychological, and physical, and depict the global burden of the problem. In addressing the issue, it is important to understand that public health initiatives often are bound tightly to existing legal frameworks, and this holds especially true in the case of GBV. It is excellent that GBV is now recognized as a major global public health and human rights issue, but efforts to stem its practice still stymie those involved in the prevention and management of GBV. Much work remains both in the industrial and developing world to ensure the safety and well-being of women of all ages against all forms of GBV.

Few issues in public health have fostered as much controversy as contraception and abortion. Passions run high on both sides of the debate; religion, politics, and policy regularly clash. Henshaw, who has spent his career conducting research in this area, presents a comprehensive statistical report on the issue of contraception and abortion, and explores some of the barriers that inhibit or even prevent women from controlling their fertility. Focusing on unintended pregnancy, contraceptive use, and abortion in the United States, he clearly and concisely presents the statistical evidence showing trends and highlighting the barriers that exist to prevent women of all ages from controlling fertility. Regarding abortion, since the legalization of this procedure in 1973, it is estimated that 35 percent of women in the United States will have had at least one abortion by age 45. Regarding birth rates, recent figures show that in 2007 more babies were born in the United States than in any other year in U.S. history.[11] This increase reflects a larger population of women of childbearing age. Births to teenagers (ages 15 to 17), after declining for many years, increased, reasons for which are poorly understood. Mississippi has the nation's highest teen pregnancy rate, which was 60 percent higher than the national average.

Although tremendous advances have been made in the eradication of once-deadly diseases, the development of vaccines probably is the most significant

reason for the decline in morbidity and mortality from such diseases. Perhaps the world's greatest achievement in this area is the eradication of smallpox. Rosen takes a global look at disease prevention through vaccination, presenting an historical overview and then focusing on the challenges that remain. There is a staggering disparity in vaccination efforts worldwide; millions of children are needlessly dying from preventable infections. Closing the gap will require multinational efforts and significant amounts of manpower and financial resources. The WHO and the Global Alliance for Vaccines and Immunization are deeply involved in coordinating immunization plans, especially in the resource-poor nations.

VOLUME 2

Volume 2 focuses on the determinants of health and disease. Section 1 addresses the treatment and prevention of chronic diseases. Since the mid-20th century, there has been a huge explosion in the number of individuals diagnosed with diabetes mellitus. Endocrinologist Baker's chapter focuses on the global epidemic of diabetes and discusses the public health, medical, and economic implications of dealing with this disease as well as the consequences to patients and to society. Diabetes is a growing and serious disease that affects rich and poor alike. According to the WHO, diabetes is likely to be one of the most substantial threats to human health in the 21st century. In the United States. alone, the direct medical costs of treating diabetics will be $336 billion. This does not take into account the growing proportion of overweight children and teenagers who are at high risk for developing diabetes and does not factor in immigration or the growing population of ethnic minorities who also suffer from diabetes at much higher rates than the U.S. white population. Without significant changes in public or private strategies, the burden of treating diabetics will place a significant strain on an already overburdened health care system. Ironically, and perhaps tragically, diabetes is among the most preventable of major illnesses. Clearly, as Baker discusses, much more needs to be done to stem the epidemic domestically as well as internationally. For a chance of success, prevention efforts must include a partnership between the individual and the health care provider.

Cardiovascular disease, too, is among the leading causes of morbidity and mortality globally. In 2004, according to the WHO, 17.1 million people died from cardiovascular disease, which represents 29 percent of global deaths. By 2010, cardiovascular disease is predicted to be the leading cause of death in developing countries. Kassahun and Borden explore the surge in cardiovascular disease and its risk factors worldwide, the characteristics and implications of this growth, as well as public health initiatives that can stem or even reverse this trend. They discuss the social, environmental, and cultural determinants of cardiovascular health, such as obesity, tobacco use, and access to health care that need to be addressed globally to reduce the incidence of cardiovascular disease.

The surging prevalence of obesity in the United States and around the world is growing faster than that of any other public health condition. This trend is alarming from a medical and an economic perspective. The ever-increasing prevalence of obesity has been accompanied by a host of inherently associated comorbidities. As a result, obesity is fast becoming the major cause of premature death

in the industrial as well as the developing world. Cardiologists Bornstein and Cooper examine the implications of the huge explosion of overweight and obesity in the world in the 21st century. Over the past two decades, the number of overweight and obese adults, adolescents, and children has increased dramatically. Of great concern are the children and adolescents who already have early obesity-related degenerative diseases, such as hypertension, dyslipidemia, metabolic syndrome, and type-2 diabetes mellitus, as well as manifestations of early preclinical atherosclerotic cardiovascular disease that previously has not been observed in this age-group. The economic costs of obesity are examined as are preventive means of addressing the epidemic. For example, health care spending for obese American adults soared 82 percent between 2001 and 2006.[12] Health care costs related solely to obesity could easily total $344 billion in the United States by 2018, or more than one in five dollars spent on health care, if the trends continue. The central message is that if nothing is done to stem the rise in obesity, the economic, medical, and personal consequences will be even more difficult to deal with.

The global burden of asthma is explored by Shirtcliffe and Beasley from the Medical Research Institute of New Zealand. The rising global burden of chronic, noncommunicable diseases over recent decades has been labeled "the neglected epidemic." Over recent decades, asthma has become one of the most common chronic diseases in the world and is now the most common chronic disease of childhood in many countries. The authors present a comprehensive epidemiologic overview of the disease and address the probably causes of the increase of asthma worldwide, including climate change and urbanization. The economic burden of asthma is considerable both in terms of direct medical costs, such as the cost of pharmaceuticals and hospital admissions, and indirect medical costs, such as time lost from work and premature death. The GBD to governments, health care systems, families, and patients is substantial. Indeed, the authors argue that the burden of asthma in many countries is of sufficient magnitude to warrant its recognition as a priority disorder in government health strategies.

Mild to moderate hypertension is generally an asymptomatic disease. It aptly has been called the "silent killer" because it usually produces no symptoms and increases gradually and slowly over the years. People with high blood pressure usually have no idea that they have this problem, and they do not go to the doctor specifically because of elevated blood pressure. The detection and treatment of hypertension is thus a major public health challenge. Cheung and Ong focus on the growing burden of hypertension worldwide, and especially in the United States. Although hypertension is seldom curable, the more practical aim is to control the blood pressure. The authors present an epidemiologic overview of the disease and discuss the known risk factors for hypertension. Medical management and improving compliance with treatment are discussed.

Arthritis, especially osteoarthritis (OA), is a prevalent condition among most of the older population. As the baby boomers age, the number of new cases of OA are likely to increase. Perhaps not surprisingly, there has been a concomitant increase in the number of joint replacements being performed. Lyman and Nguyen focus on OA of the knee and the explosion in the number of total joint replacements being performed. While total knee replacement, in particular, is an elective procedure to treat severe arthritis of the knee, the increase in this surgery is driven

both by the aging of the population as well as the obesity epidemic. Weight loss interventions may be the single most efficacious method of prevention of knee OA, but barring that, surgical intervention is increasingly being used to treat OA and enhance quality of life. The authors discuss the economic burden of OA, which is substantial (direct and indirect costs associated with OA).

Section 2 focuses on advances in cancer screening and the challenges that remain. For years, the dominant view about screening was that early detection and aggressive treatment would lead to increased longevity. Screening for cancer targets healthy, asymptomatic individuals. The purpose is to detect the disease at an early stage to initiate treatment, which hopefully will extend life. A key principle is that the potential benefits of screening should outweigh the harms of testing. The physical, psychological, and economic sequelae of follow-up testing should the screening test be positive needs to be compared with the number of lives saved as a result of screening.

Cancer is the nation's number two killer behind heart disease and accounts for nearly a quarter of annual deaths. The good news, however, is that cancer death rates and the number of new cancer cases in the United States continue to decline.[13] The conclusion drawn was that early detection and new therapies are major contributors to this effect. Almost at the same time as this report was released, a new study on the effectiveness of mammogram screening also was released. An expert panel from the U.S. Preventive Services Task Force recommended that mammography screening to detect breast cancer should be scaled back. This bombshell recommendation, in direct conflict with the recommendations from the American Cancer Society and other medical groups, caused considerable confusion, distrust, and even anger. Studies evaluating the effectiveness of the prostate-specific antigen (PSA) test to screen for prostate cancer also have yielded questionable results, making a clear recommendation for or against this test almost impossible. The American Cancer Society, a staunch defender of most cancer screening, has said that the benefits of detecting many cancers in particular prostate cancer has been overstated. The PSA prostate cancer screening test has not been shown to prevent prostate cancer deaths. The dilemma for breast and prostate screening is that it is not usually clear which tumors need aggressive treatment and which can be left alone.

Some studies focusing on routine early cancer screening found that the screening did not save lives, thus calling into question why screening was being advocated in the first place. In some cases, widespread screening increased the detection and treatment of small, slow-growing tumors that may well never have caused harm. In some cases, the tumor might regress or even disappear. While almost all of the cancer screening tests in wide use are minimally invasive, fairly inexpensive, and generally accepted by the public, none are 100 percent accurate; positive test results require further workup, which often are invasive and costly and usually lead to overtreatment. Screening does come with medical risks. In many cases, disease is not evident, such as in the case of false-positive test results.

Trevena presents a scholarly overview of the benefits and risks of screening and early detection of disease. She then examines the evidence for screening for colorectal cancer. The issue of whether an individual benefits from early detection of cancer is not as straightforward as it may seem. For some diseases, a preclinical

phase may be so short that the disease is not likely to be detected by screening. Or, even if detected, options for cure may not exist. Also, not every preclinical case will progress to clinical disease. Trevena examines the screening options for colorectal cancer, including fecal occult blood testing, flexible sigmoidoscopy, colonoscopy, and a new screening option, CT colonography. She presents the pros and cons of each modality. Some countries recommend that a fecal occult blood test be used while others advocate for colonoscopy. The accuracy of the test, including false-positive results, needs to be weighed against the potential benefit in reducing colorectal mortality.

Elkin and Blinder of the Memorial Sloan Kettering Cancer Center in New York City focus on breast cancer screening. Mindful of the current mammography screening controversy, the authors present a comprehensive overview of the advances made in reducing and preventing breast cancer. Because so many of the risk factors for breast cancer are not modifiable, much attention has been devoted to other means of breast cancer prevention such as understanding the role of hormones in breast cancer etiology. Much of the chapter focuses on the current controversy in mammogram screening. The questions of when screening should be initiated, on whom, and how frequently remain controversial. The authors present a scholarly assessment of the evidence, including a discussion on the realities of false-positive results. The chapter concludes with a discussion of advances in breast cancer treatment, including surgical treatment, radiation therapy, systemic therapy, hormonal therapy, and chemotherapy.

Lung cancer, the leading cause of cancer mortality worldwide, typically exhibits symptoms only after the disease has spread to other organs, unfortunately making it difficult to cure patients with such advanced disease. The overall prognosis of this cancer is poor when compared with other cancers, such as breast or colon, and is dependent on where the cancer is located, the size and type of tumor, and the overall health status of the patient. The two types of lung cancers, small-cell lung cancer and non-small-cell lung cancer, grow and spread in different ways and also have different treatment options. To date, screening for lung cancer is not advocated for these reasons. Yet, we have known for decades that tobacco smoking is the leading cause of this cancer, and evidence is quite clear that if individuals stopped smoking (or never started), the incidence of lung cancer would be greatly reduced. Lung cancer can be prevented. Mazumdar's chapter on lung cancer prevention, screening, and treatment reviews the epidemiology of this cancer as well as focuses on the "effectiveness," "efficiency," and "efficaciousness" of treatment regimens. A national initiative for comparative effectiveness research (CER) for clinical decision making is described. A discussion of CER provides a review of ongoing research and initiatives in this area, and highlights the gaps in information and research. Overall, much works is needed to find a cure for lung cancer and in being able to bring the best possible care to patients of all race, gender, and socioeconomic status.

Controversy over prostate cancer screening and treatment options continues to play out in the lay and professional media. Nguyen and Kattan's chapter reviews the current status of prostate cancer screening and assess its benefits and potential deleterious effects, to determine ways to improve its predictive accuracy and efficacy. To better understand the controversy surrounding prostate cancer

screening and perhaps offer a solution, they provide a review of current screening modalities, assess their accuracy and utility in contemporary medical practice, and suggest future directions for improvement of prostate cancer screening.

Over the past decades, the incidence of skin cancer has increased substantially. The chapter by Berwick, Erdei, Gonzales, Torres, and Flores focuses on the epidemiology and genetics of skin cancer and illustrates the growing public health burden of this particular form of cancer. Advances in screening and treatment are discussed and preventive measures are explored. The incidence of melanoma, a potentially deadly form of skin cancer, has soared over the past few decades perhaps because of an increased interest in screening for the disease. Indeed, the increase might be due to a growing tendency to identify and treat benign lesions as malignant cancers. It is quite difficult, and sometimes impossible, to tell a malignant lesion from a melanocytic nevus, a type of benign mole. The authors discuss how to protect oneself from skin cancer and provide informative information on sunscreens, tanning beds, and genetic susceptibility. Although sun exposure is the major risk factor for skin cancer, it is also necessary for synthesis of vitamin D, necessary for bone and muscle health and a possible protective factor for many diseases, including colon cancer. Given the worry about sun exposure and skin cancer, the question remains: How does one achieve favorable vitamin D levels yet also practice skin cancer prevention?

Cervical cancer, so easily and inexpensively prevented, remains a major killer among women in resource-poor nations. Without screening intervention, morbidity and mortality will continue to increase. Sankaranarayanan, Thara, Ngoma, Naud, and Keita have published groundbreaking research on the topic, and in this chapter they present a comprehensive overview of the issue, including evidence convincingly showing that screening for human papillomavirus (HPV) can yield a significant reduction in the numbers of advanced cervical cancers and deaths from this disease. They review the current status and future prospects for controlling cervical cancer in developing countries in this chapter. Low-tech screening methods (often used because most rural areas cannot realistically conduct Pap smear screening) and a single round of screening for HPV can and does result in a dramatic reduction in the incidence of advanced cervical cancer. There is a huge potential to reduce the cervical cancer burden by means of HPV vaccination. The authors acknowledge that a recommendation for HPV vaccination for adolescence women for a disease that occurs during adulthood is a major paradigm shift in cervical cancer control. Although HPV vaccination holds great promise, and has been licensed for use in more than 100 countries, there are several challenges (notably cost) for its widespread implementation through national immunization programs in high-risk developing countries. Cervical cancer reflects striking global health inequity, resulting in deaths of women in their most productive years in developing countries, with a devastating effect on the society at large. It remains as the largest single cause of years of life lost to cancer in the developing world.

Section 3 of volume 2 focuses on the treatment and prevention of infectious diseases. So much has been written about HIV/AIDS over the past two decades and so much progress has been made in extending life expectancy among those with the disease. Demars takes a broad view of the epidemic, tracing its history and focusing on the global burden of the disease. While recent trend data indicate

that the incidence in Africa has appreciably slowed, dealing with the disease's sociopsychological sequela and ensuring that progress made is not eroded remain challenges both in the industrial world and the resource-poor, hard-hit part of the world.

The WHO estimates that more than 500 million individuals worldwide are infected with the hepatitis B or C virus. Hepatitis viruses are found in every part of the world and often cause infections ranging in severity from acute infections that are asymptomatic to fulminate, chronic infections, which in some instances can lead to cirrhosis and hepatocellular carcinoma or even death. Aden presents a focused discussion on the most prevalent hepatitis viruses (A, B, and C) and explains how these diseases remain an important public health concern in both the developing and the industrial world. Whereas hepatitis B is a more serious type of infection than hepatitis A, hepatitis C infection can result in serious liver damage; hepatitis C is one of the leading causes for liver transplantation. While hepatitis A and hepatitis B are vaccine preventable, no vaccine is available for hepatitis C. Risk factors, population at risk, and treatment modalities are presented.

The ebb and flow of sexually transmitted diseases (for example, chlamydia, gonorrhea, and syphilis) has long been a focus of public health practitioners. All three diseases are preventable, treatable with medication, and, in the early stages, curable. Torrone and Peterman of the CDC present an overview of the topic and focus on syphilis specifically. The authors discuss the challenges of sexually transmitted disease (STD) control focusing on trends, efforts at prevention and control, and the challenges that remain. STD control and eradication is possible, but certainly not easy.

Section 4 of volume 2 addresses the treatment and prevention of mental health illness and disease. The burden of mental health disorders in the United States is substantial with approximately half of the population meeting the criteria for one or more such disorders in their lifetime and almost one-quarter meeting the criteria in any given year.[14] Treatment costs for mental disorders are substantial, rising from $35 billion (in 2006 dollars) to nearly $58 billion, making it the costliest medical condition between 1996 and 2006.[15] The most prevalent class of disorders is anxiety disorders (for example, phobias, panic disorders, and the like) followed by impulse-control disorders, mood disorders (for example, major depressive disorders, bipolar disorders, and the like), and substance abuse disorders (for example, alcohol abuse or dependence, drug abuse or dependence, and the like). The most prevalent type of disorder is major depressive disorders. Most individuals with a lifetime mental disorder had their first onset in childhood or adolescence. Little is known about the epidemiology of child mental disorders and controversy exists about how best to treat children. Contributors to this section focus on specific mental disorders, such as depression, suicide, and substance abuse.

Depressive disorders are prevalent conditions among the general population, and the medical, public health, and economic consequences of depression are considerable. Tedeschini, Cassano, and Fava present an overview of depressive disorders and focus on the recognition, management, and treatment of these diseases. The authors stress that depression is underdiagnosed and undertreated as only half of all Americans with depression receive treatment of any kind. Despite the

availability of numerous effective treatments, many depressive disorders are often misdiagnosed. Several factors contributing to the poor recognition of depression have been identified, ranging from the stigma of depression itself to the relative lack of systematic ascertainment of depressive symptoms by physicians.

Barber and Miller focus on the topic of suicide both within the Untied States as well as globally. In their informative and scholarly piece, they review the salient aspects of the epidemiology of suicide and the challenges posed by a purely clinical approach to its prevention. They outline a public health approach to suicide prevention, with an emphasis on reducing a suicidal person's access to lethal means of suicide. Their thesis clearly illustrates that although suicide is a global problem, a public health approach to prevention is still in its infancy. Public health strategies, such as changing cultural attitudes, increasing social support, improving access to high-quality treatment, and perhaps most important, reducing access to lethal means are measures that can and should be implemented.

Griffin examines the data on substance use and abuse across the life span from early adolescence to late adulthood. There is great diversity in patterns of alcohol, tobacco, and other drug use over the life course, with some individuals abstaining from use throughout their lives and others facing ongoing battles with substance abuse and dependence. The focus is primarily on substance use rather than abuse, because substance use is more prevalent than abuse in the general population and therefore has a greater public health impact. A goal of the chapter is to examine the extent to which substance use can be thought of as a developmental phenomenon not only among young people, but also throughout the life course. The implications of a life span developmental perspective to guide substance use prevention efforts are discussed. Griffin highlights a future challenge: the anticipated increase in substance use problems among the elderly and among the baby boomers, the eldest of whom will be reaching age 65 in a few short years. By taking into account how age-related developmental factors can affect substance use, we may be better able to address these and other new prevention challenges in the future.

VOLUME 3

Volume 3 shifts focus to health policy issues. In section 1, Finkel provides a historical overview of comparative health care systems illustrating why and how other industrial nations moved toward universal health care and why the United States did not. The organization, administration, financing, and delivery of health care in several countries are presented in an in-depth analysis. A critique of how health care is delivered and financed in other countries provides a stark contrast to how health care is provided and paid for in the United States.

Quality and patient safety, in addition to cost management, is an important issue in health care policy. Lazar, Dawson, Hyman, Collins, Regan, Kaplan, Green, Cook, and Graham from the New York-Presbyterian Hospital present an overview of quality assurance, quality metrics, and quality evaluation techniques. Performance improvement management methodologies designed to reduce medical errors and safeguard a safe workplace. In 1999, the IOM published a seminal report entitled *To Err Is Human,* which catalyzed an enormous shift in the understanding of medical errors. The IOM report defined an error as an event in which

there is a failure of a process to achieve the intended outcome, or where an incorrect process of care was selected initially. An adverse event was defined as an injury to a patient caused by medical management rather than the patient's medical condition. The IOM report concluded that medical errors were responsible for as many as 98,000 deaths in the United States annually. Estimated annual costs of these errors were in the range of $17 billion to $29 billion. The report further opined that injuries caused by errors are inherently preventable. Lazar and colleagues state that achieving better outcomes for patients, lowering overall costs, and improving the patient experience will require the continued investment of time and money to spur innovation and create reliable effectiveness, safety, and efficiency in clinical settings. Measurement and continuous performance improvement are the mainstays of a robust organizational quality assurance program.

Until the 1980s, most people with private insurance in the United States were covered by traditional indemnity plans. As remains the case, the vast majority got their coverage through employment-based plans provided as a tax-exempt benefit. These indemnity plans delegated shopping decisions about what care to buy and where to buy it to individual consumers and their physicians and then relied on consumer cost sharing to contain costs. Specifically, plans used deductibles and coinsurance to create financial accountability for purchases; the notion was that responsibility for resulting out-of-pocket payments would create incentives for cost-conscious shopping. By design, health plans were relegated to a passive role of paying the bills, while providers were reimbursed fee-for-service on a cost basis. Such open-ended insurance schemes laid the foundation for rising costs, which the United States is now trying to reign in. Managed care was an attempt to contain costs, but has not succeeded in doing so. White discusses an alternative, consumer-oriented strategy (Consumer Directed Health Plans, CDHPs) to address the concerns and shortfalls of managed care. The basic nation of CDHPs is that by placing consumers at risk for paying for substantial amounts of care with their own money, this simultaneously will restore control over shopping decisions and increase consumers' motivation for cost-conscious shopping, while introducing savings options will mediate the accompanying increase in exposure to financial risk. White provides an in-depth discussion of CDHPs and their potential effectiveness in managing health care cost increases.

The use of health information technology (health IT) has become an exciting and important field in medicine. Ancker, Kern, Patel, Abramson, and Kaushal present a scholarly overview of the present and future uses of health IT. Health IT has been promoted widely as a potential solution to managing the massive amounts of data and information as well as serving as a cost management tool. The authors discuss the various types of health IT systems and explore the barriers to development and implementation of these technologies. Health IT offers particularly exciting possibilities for improving the quality and efficiency of health care delivery by making essential individual-level medical data more readily accessible at the point of care; improving communication among clinicians, patients, and public health agencies; and providing evidence-based clinical decision support to help clinicians practice according to optimal care guidelines.

Section 2 of volume 3 focuses on the difficult issue of health care disparities. Health status and health outcomes vary markedly among racial and ethnic groups.

According to an IOM report *Unequal Treatment: Confronting Racial and Ethnic Disparities* race and ethnicity remain a significant factor in determining whether an individual receives high-quality care and in determining health outcomes. Race has been shown to be a determinant of the characteristics and qualifications of physicians who patients see, the types of hospital to which a patient is admitted, and the types of procedures they will undergo. The explanations are complex.

Boutin-Foster focuses on diversity and the public health implications of a growing racially and ethnically diverse America. She examines the role academic medical centers can and should play in providing care to this multicultural population. An argument is made for the need to bring the issues of cultural diversity to the forefront of medical education. While progress has been made in increasing the proportion of racial and ethnic minorities in the health care field, the racial and ethnic composition of the health care workforce does not match that of the general population. Would systematic biases in treatment be reduced if the composition of the workforce resembles more closely that of the patient population? While no studies have been done to empirically answer this question, given the extent of disparities and unequal treatment (which have been researched), one could assume that it certainly would not hurt.

Section 3 of volume 3 discusses ethics and human rights issues. Atkinson explores why human rights is crucial to the work of public health, and argues that human rights is a necessary framework for public health. Her chapter explores why a human rights framework is crucial to the work of public health. The human rights framework—in concert with traditional medical ethics—articulates certain values and standards that specify how we should conduct ourselves. She presents an argument for an ethical and legal framework for moving forward the global public health agenda. She believes that the human rights framework offers us a reason to believe in the possibility of change.

Bioethicist de Melo-Martin addresses the ethically charged topic of genetic testing and public health. She presents some of the most significant ethical concerns that arise in relation to the use of genetic tests, discusses matters related to the analytic and clinical validity and utility of genetic tests, and explains how these aspects result in ethical quandaries. She then focuses on the concerns that the use of genetic tests, if such tests prove beneficial for the populations' health, might contribute to furthering existing health inequities. Finally, she discusses ethical issues related to obtaining, or omitting, informed consent and to protecting privacy and confidentiality. Ensuring that people are not unjustly discriminated against because of their genetic or health status requires careful attention to issues of privacy and confidentiality; yet, concerns about privacy need to be balanced against the legitimate public health needs. Focusing on these ethical concerns when making public-policy decisions about implementation of genetic testing and screening is necessary if we want to use these medical technologies in ways that will advance the public's health.

As chronic diseases, including cancer, surpass infectious diseases as the primary causes of death, and as individuals are living longer with their diseases, providing timely access to consistently high-quality end-of-life care has become an important international issue. How we manage the dying patient has both medical and ethical concerns. Tickoo and Glare present a comprehensive overview of the

palliative care movement both in the United States as well as in Australia, England, and India. At some point, all humans have to confront the inevitability of end of life. How one prepares for the eventuality of death is a personal and individual matter. What is necessary and important, however, is that end-of-life choices be made clear and available. Providing for end-of-life care is emotionally difficult, thus making it even more imperative that all patients have the option of timely access to palliative care services that are both appropriate and cost-effective.

Section 4 of volume 3 focuses on public health practice and education. Trushin and Bang present an interesting chapter on the role of epidemiology and biostatistics in health news reporting. In their thoughtful piece focusing on the role of uncertainty in science, they provide a comprehensive overview of the mechanisms of research and statistical analysis. The scientific method is based largely on common sense, and statistical thinking involves concepts that are accessible to all: an acceptance of chance and uncertainty, an appreciation of context, an ability to detect logical and factual flaws in information and ideas, and the realization that science is a fluid process whereby new empirical evidence is accumulated every day. The true spirit of science requires a healthy skepticism, which means suspended judgment and the use of reason to evaluate the validity of research results. Science thrives on these qualities, because they lead to a search for knowledge and ensure that the scientific method remains self-correcting.

Evidence-based medicine has been incorporated into the medical school curriculum, but it also has a role in public health practice. The design and use of public health actions that are effective in promoting health and preventing disease underlie the growing field of evidence-based public health (EBPH), which emerged in the 1990s to improve the *practice* of public health. Maylahn, Brownson, and Fielding describe the concepts and principles underlying EBPH, the analytic tools to enhance the adoption of evidence-based decision making, the dissemination and implementation in public health practice, and challenges and opportunities for more widespread use of EBPH, especially through state and local health departments. Unlike solving a math problem, significant decisions in public health must balance science and art, because rational, evidence-based decision making often involves choosing one alternative from among a set of rational choices. By applying the concepts of EBPH outlined in their chapter, the authors concluded that decision making and, ultimately, public health practice can be improved.

The American Association of Medical Colleges (AAMC) is the umbrella organization for U.S. medical schools. The AAMC's position on medical curriculum has far-reaching impact. Maeshiro of the AAMC presents an informative historical overview of the tensions and barriers to integrating the disciplines of clinical medicine and public health. The challenge of incorporating public health content into the standard medical curriculum is not new. Not surprisingly, the roots of this struggle are entwined with the historical events and trends that led to the separation, or "schism" as some have described, between the practice of medicine and the practice of public health in the United States. She relates how over time the disciplines have gradually moved toward an integrated whole both at the medical school curriculum level and at the postgraduate medical training level. The rise of

a specialty in preventive medicine, the development of a residency in preventive medicine, and the subsequent creation of board certification in this area are interrelated. The framework in which medical education exists (for example, accreditation criteria for both medical school and residency training, national examination content) acknowledge the need for physicians to have a population perspective.

Section 5 of volume 3 addresses some of the challenges public health faces as we move into the second decade of the 21st century. Few areas of biomedical science have aroused as much controversy as embryonic stem cell research. With advances in medical research and technology, stem cell research has proliferated around the world. Cauley addresses the stem cell debate, focusing on the medical, ethical, legal, and political aspects of the topic. He provides a scholarly overview of the short history of stem cell research and raises important questions that need to be addressed today and into the future.

Advances in computer science have opened a new area of research for global disease monitoring. McEntee, Castronovo, Jagai, and Naumova from Tufts University provide an overview of a number of advanced computational and analytical techniques that open new opportunities to examine the role of forecasting disease transmission and manifestation. They review applications of various remote sensing (RS) techniques and present the relatively nascent epidemiological applications of this technology. Public health applications of RS data are no longer new; spatial epidemiology is equally important as the strictly environmental applications for which RS was originally intended. This is not surprising because environmental studies and epidemiology are inextricably linked. Each provides information on human health conditions and the corresponding management of environmental resources. Climate and land-use change and variability can be measured remotely and corresponding effects of alterations in natural and built environments can be predicted. Their scholarly and thoughtful presentation of this new field illustrates the tremendous opportunities that can be tapped.

Smith's concluding remarks on thinking creatively about public health in the 21st century is an excellent historical wrap-up of key events in public health over the centuries. He provocatively asks what the public health field needs to do to meet the challenges. How should public health be shaped for the 21st century, both for its own sake as a critical field for the world's well-being and for the sake of the local and global public it serves? The answers to these questions, of course, are multilevel and multifaceted. He advocates that 21st-century public health should begin to look more rigorously at the multiple factors in a society that predict health outcomes. These factors include economics, housing, nutrition, sports and recreation, education, spirituality, family structure, gender relations, childcare, transportation, and whatever other factors make up an integrated human life. Those concerned with the improvement of health on a local, national, and global scale need to work collectively rather than in isolation. Health, after all, is a product of the multiple facets of society, and, as such, requires a multifaceted approach to health promotion, the prevention and treatment of disease, and, most important, the improvement of the quality of life for all people.

This is an exciting time for public health. As public health practitioners continue to work toward improvement in the health and well-being of populations around the world and focus on disease eradication and the prevention and control

of diseases, injury, and disability, this increasingly is being achieved in a global context for the potential benefit of all. It is the aim of this multivolume reference text to identify and analyze the diversity of the work being conducted in the contemporary public health landscape.

The tremendous effort that went into creating this multivolume text could not have been done without the generosity of the contributing authors. My appreciation for their time, their enthusiasm, and their scholarship, and especially their friendship, cannot be underestimated.

Madelon L. Finkel, PhD
New York City

REFERENCES

1. Winslow C-EA. The untilled fields of public health. *Science*. 1920; 51:23.
2. D. Cutler D, Miller G. The Role of public health improvements in health advances: the 20th century United States. http://www.nber.org/papers/w10511. National Bureau of Economic Research working paper w10511. Published May 2004.
3. Bunker JP, Frazier HS, Mosteller F. Improving health: measuring effects of medical care. *Milbank Quarterly*. 1994; 72:225–258.
4. Ten great public health achievements—United States, 1900–1999. *MMWR*. 1999; 48(12):241–243. http://www.cdc.gov/mmwr/.
5. World Health Organization. *The World Health Report 2003-Shaping the Future*. Geneva, Switzerland: World Health Organization; 2003.
6. Health: Confronting the Challenges. http://www.usaid.gov/our_work/global_health/home/confrontingfactsheet.html.
7. G8. G8 communiqué. http://www.g8.utoronto.ca/summit/2000okinawa/finalcom.htm. Okinawa 2000.
8. Group of Seventy-Seven. Final communiqué adopted by the thirty-fourth meeting of chairmen/coordinators of the chapters of the Group of 77. http://www.g77.org/news/pr062703.htm. June 26–27, 2003; Geneva, Switzerland.
9. Alwan A, MacLean D, Mandil A. *Assessment of National Capacity for Non-Communicable Disease Prevention and Control*. Geneva, Switzerland: World Health Organization; 2001.
10. United Nations Development Programme. Millennium Goals. http://www.undp.org/mdg.
11. National Vital Statistics Report. Births, Marriages, Divorces, and Deaths. Vol. 57, No. 12. March 18, 2009. www.cdc.gov/nchs/data.
12. Trends in health care expenditures by body mass index (BMI) category for adults in the US civilian noninstitutionalized population. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st247/stat247.pdf.
13. Annual report to the nation on the status of cancer: 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. In BK Edwards, eds. *Cancer*. Lawrenceville, GA: Surveillance Research Program, National Cancer Institute.
14. Kessler RC, Wang PS. The descriptive epidemiology of commonly occurring mental disorders in the United States. *Annu Rev Public Health*. 2008; 29:115–129.
15. The five most costly conditions, 1996 and 2006: estimates for the US civilian noninstitutionalize population. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st248/stat248.pdf.

This page intentionally left blank

SECTION 1

GLOBAL POPULATION HEALTH

This page intentionally left blank

CHAPTER 1

The Global Burden of Disease

Sandeep P. Kishore, MSc, and Marilyn D. Michelow, BA

At the turn of the 20th century, the average woman born in the United States could expect to live for 51 years.[1] Fast-forward 100 years and the average life expectancy for an American woman was closer to 80 years. The nearly 30 years that American women have gained in life expectancy is a testament to the great strides that have been made in modern medicine and public health. This progress, however, has hardly been uniform, even within our country. Today, life expectancy for a Native American male in South Dakota, for instance, is a mere 58 years. On average, a nearly 35-year gap in life expectancy exists between the most and least healthy populations in the United States.[2] Looking beyond U.S. borders and comparing the health status among countries, the data reveal striking gaps in the distribution of health globally. In the 21st century, a girl born in Sierra Leone can expect to live less than an American girl could 100 years ago; life expectancy in Sierra Leone is a shocking 41 years. The political, social, and economic determinants of health all drive this vastly heterogeneous, checkered, and complex global burden of disease (GBD).

More than 99 percent of the burden of maternal and early childhood diseases is concentrated in developing countries; at the same time, communities around the globe increasingly are burdened with a silent epidemic of noncommunicable chronic diseases. Emerging nations, typified by India and China, rapidly are urbanizing and maturing their economies and, in turn, driving a new global picture of disease burden. Highly processed and fast food, staples of Western Hemisphere life, have arrived in the far corners of the earth, from Delhi to Djibouti, making high-fat, cheap foods the easy choice for billions of people. The effect of this and other related trends is predictable: more and more people will suffer from diseases traditionally associated with the Western Hemisphere. Already, cardiovascular disease, often thought of as a disease restricted to affluent populations, is the leading cause of death in the world—with almost a full 80 percent of these deaths concentrated in developing countries.[3] When the more recent epidemic of noncommunicable disease is added to the persistent plagues of communicable diseases,

including malaria, tuberculosis (TB), hookworm, and HIV/AIDS, the task of assuaging disease rationally and strategically becomes most confusing, if not downright daunting.

The focus of this chapter is to highlight the salient features of the GBD, including its distribution and changing patterns over time. The reader seeking more detailed or specific information is encouraged to consult the Disease Control Priorities Project of the World Health Organization (WHO).[1] We hope to give the reader a perspective on both historical and contemporary views of the burden of disease, outline the current data on disease prevalence and impact, and introduce a fuller discussion of future directions for the study of the global burden on disease.

HOW TO DEFINE THE 'BURDEN'

Although the "burden" of a disease can be defined in a variety of senses, the consensus definition, particularly from the WHO, is a fairly specific one. "GBD" is defined by the WHO as a comprehensive regional and global assessment of mortality and disability from 136 diseases and injuries and 19 risk factors. It accounts for the morbidity and mortality to an individual that is caused by a specific disease.[4] This information is aggregated into country level data to form the "burden," which can be viewed as the gap between current health status and an ideal situation in which everyone lives to old age free of disease and disability.

Economic and social determinants and the effect of individual disease on an individual's community and society (apart from "ill health") are not included in this definition. As the currency in the public health literature is limited to the parameters of the WHO definition, burden of disease in this chapter is defined in the above sense. We do return, however, to alternative, if not complementary, ways to capture this "burden" later.

THE EPIDEMIOLOGIC AND ECONOMIC TRANSITION

The GBD can be divided into three distinct buckets (see figure 1.1).[3] The first set of diseases includes the communicable, maternal, perinatal, and nutritional conditions, which, unsurprisingly, predominate in lower-income countries (for example, those in Sub-Saharan Africa and Southeast Asia). As a country develops, a well-documented epidemiologic transition takes place. Here, traditional risk factors for infectious diseases (poor food, limited access to good water, inadequate sanitation) are supplanted by risk factors for more chronic, lifestyle diseases (for example, workplace-associated pollution, smoking, high-fat diets). In addition, with a reduction in

1. See the Disease Control Priorities Project Web site, http://www.dcp2.org/main/Home.html, or the WHO Global Burden of Disease Web site http://www.who.int/topics/global_burden_of_disease/en/, for up-to-date information and raw data. In addition, the following texts on the Global Burden of Disease will be of particular use to readers:
Lopez AD, Disease Control Priorities Project. *Global Burden of Disease and Risk Factors.* New York, Washington, DC: Oxford University Press; World Bank; 2006.
Jamison DT, World Bank, Disease Control Priorities Project. *Disease Control Priorities in Developing Countries*. New York, Washington, DC: Oxford University Press; World Bank; 2006.

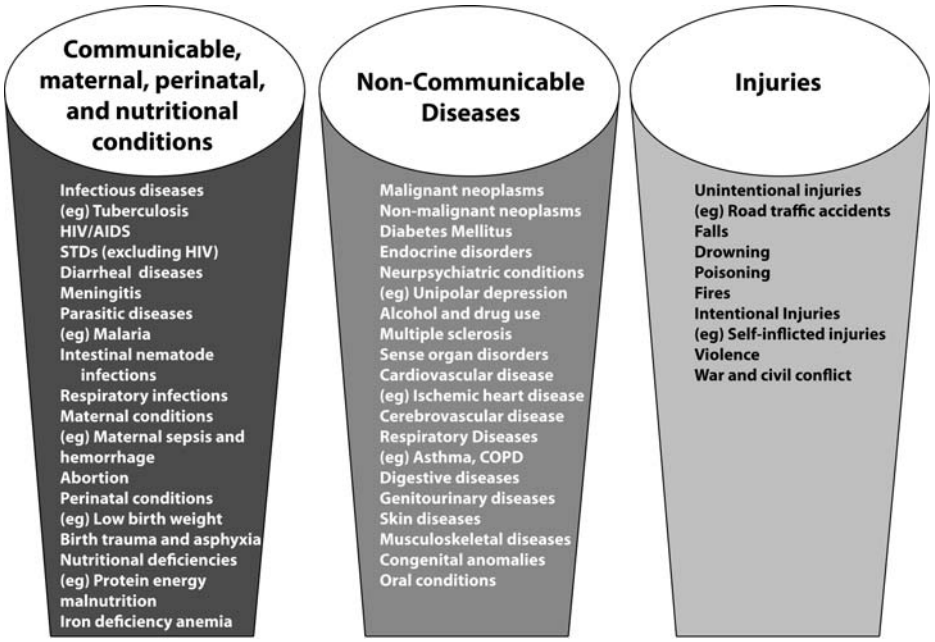

Figure 1.1 ''Buckets'' of diseases surveyed globally.

childhood mortality, populations collectively age, meaning that they experience disease risks for longer periods of time, particularly in urban environments.[5]

The result is as would be expected: people from countries with transitional economies begin to develop noncommunicable diseases, including diabetes, cancer, and heart disease. A final category of the burden, often neglected by health practitioners, is injuries—either intentional (violence, war) or unintentional (road traffic accidents).

The epidemiological and economic transition began in countries around the North Atlantic in the late 19th century. Since this time, the transition has diffused globally, fundamentally altering the demographic structure of countries.[6, 7] Maximum life expectancy has increased 2.5 years per decade for the past 160 years, along with declines in fertility and death rates.[4] The change in life expectancy affects economic growth dramatically; it is estimated that each additional year of life expectancy per person raises the gross domestic product (GDP) per capita by 4 percent in the long run.[8] Naturally, increasing longevity has resulted in an explosion of diseases associated with longer lives.[7] It is projected that by 2030, two-thirds of the global burden will be chronic diseases, and that communicable diseases will decline from affecting 41 percent to 20 percent of the global population.[9] The combination of persistent risk factors for communicable disease together with the emergent risks for noncommunicable diseases, including tobacco use, lack of physical activity, and poor diet, drive a so-called dual burden of disease in countries going through the epidemiologic transition (see figure 1.2).[10]

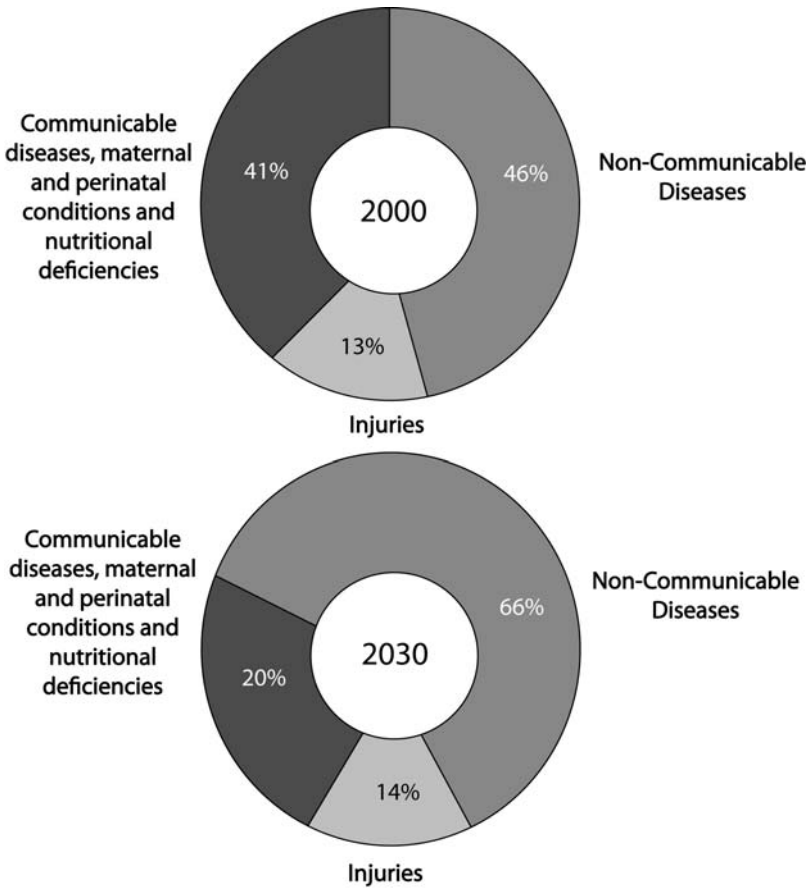

Figure 1.2 The Global Burden of Disease: 2000–2030.

HOW TO MEASURE THE BURDEN: THE OLD, THE NEW, AND THE UNEXPLORED

The most straightforward way to measure and present data on the burden of disease is to use statistics that quantitatively describe the proportion of a population with a particular illness, or the number of individuals negatively affected by a given risk factor. However, demographers, policy makers, and epidemiologists have recognized that aggregate statistics are difficult to compare across time, space, and differing conditions. Separate measures for mortality, morbidity, incidence, and prevalence make policy evaluation and intervention analysis problematic.

Perhaps a better measurement of the GBD is one that integrates morbidity, mortality, incidence, and prevalence into a single common metric that can be compared across time, space, and interventions. In 1990, recognizing the need for such a common metric, the World Bank, in concert with the WHO and the Harvard School of Public Health, launched a study to assess and quantify the GBD. Out of this study came a new metric, the Disability-Adjusted Life Year (DALY), which is a summary measure of population health, measured in units of

time (years), combining estimations of both fatal and nonfatal health outcomes (morbidity and mortality) to calculate the number of years of fully healthy life lost by an individual with a particular illness or condition.[11]

The DALYs for a specific cause are calculated as the sum of the years of life lost due to premature mortality (YLL) resulting from that cause, and the years of healthy life lost as a result of disability (YLD), with the disability weights factored in for incident causes of the health condition as follows:

$$\mathrm{DALY} = \mathrm{YLL} + \mathrm{YLD}$$

YLL is the number of cause specific deaths, multiplied by the standard life expectancy at age of death in years, and YLD is the number of incident cases multiplied by the average duration of the disease weighted by a disability factor.[12]

The disability factors, a way to include morbidity in estimations of burden of disease, are a source of much debate among critics of the DALY.[2] The DALYs are calculated in such a way that years lived at older and younger ages are given less weight. With these calculations, a death in infancy corresponds to 33 DALYs, and deaths at ages 5–20 account for around 36 DALYs.[4] To quote Christopher Murray, the architect of the DALY: "A disease burden of 3,300 DALYs is equivalent to 100 deaths in infancy or 5,500 people aged 50 living for one year with blindness."[13] Measurements in units of time, such as years, are practical because they can easily be converted to economic estimates for cost-effectiveness analysis, for example, to decide that effective interventions are those that cost a certain percentage of GDP per capita per DALY averted.

In addition to measuring the burden of DALYs of each specific disease, much recent work has been done on developing a better understanding of the burden of disease attributable to major risk factors, such as smoking, malnutrition, or environmental factors. The standard measurement to quantify the contribution of a risk factor to a disease is the population attributable fraction (PAF). The PAF measures the estimated reduction in disease or mortality that would result, in the absence of the specific risk factor. For example, PAF for tobacco use would be the percentage reduction in DALYs to be expected if nobody in that population was exposed to tobacco smoke. It should be noted that risk factors are not necessarily additive—one risk factor can cause multiple diseases, and diseases can be caused by multiple risk factors; therefore, the risk factor PAFs for a given disease often add up to more than 100 percent.[9] The DALY—along with summary health statistics on morbidity, mortality, life expectancy, and PAF for risk factors—are now viewed as important

2. Disability weights arose out of a notion that there is a cross-cultural single and quantifiable negative effect for each adverse, nonfatal health event. While proponents of the DALY have argued that this is the most equitable way to measure the burden of disease, because the loss of an eye, for example, is not worth less or more to an individual depending on where the individual lives, others have argued that it is impossible to equate health states in such varying environments. For example, the argument goes, paraplegia in Australia is altogether different from paraplegia in Cameroon given the vastly different infrastructure and social service support available in the two countries, and it is therefore inequitable to use uniform disability weights.[12] The debate continues, but for the moment, disability weights do take into account age and gender but do not adjust for environmental factors, such as where an individual lives.

measures of burden of disease.[3] Are there "problems" with a DALY index? Yes, of course. As this metric gains greater acceptance, it is important to understand that there are alternative ways, albeit largely untested, to measure the burden.

WHO MONITORS AND UPDATES THE BURDEN?

The World Bank and the WHO have been largely responsible for the current information available on the GBD. They have teamed up to conduct a series of surveys synthesizing information available globally in 1996, 2001, and 2004 to estimate mortality, morbidity, incidence, prevalence, and DALYs for 136 diseases and injuries over seven economic and geographic groups. In addition, mortality and population attributable risk is estimated for 26 separate risk factors.[9, 11, 14] A fourth global survey of the burden of disease is planned for 2010. The most up-to-date statistics on the burden of disease are available at the WHO Web site.[4] Information for the estimation of the GBD is drawn from a vast variety of sources, including regular reporting information submitted to the WHO; WHO- and United Nations Children's Fund (UNICEF)-funded country-level surveys; in-country surveillance systems; ministries of health and country census data; U.S. government–funded demographic and health surveys; verbal autopsy data; academic epidemiological studies; and specifically developed modeling programs to extrapolate information available from incomplete or outdated data sources.[14, 15]

THE GLOBAL BURDEN OF DISEASE

This section compares the burden in terms of mortality (death), DALYs (death and disability = "ill health") and causes of death across all incomes, using the most updated WHO GBD data from 2004, published in 2008.[9, 14][4] A summary of key findings is provided in box 1.1.

Mortality Patterns

Historically, mortality has been used to estimate the GBD, with reporting on the patterns of global deaths in adults and children, and by geographic region. In 2004, there were an estimated 58.8 million deaths or 1 percent global mortality. Overall, of every 10 deaths globally, six are due to noncommunicable diseases; three to communicable diseases, reproductive, and nutritional conditions; and one to injuries. Cardiovascular diseases are the leading cause of death in the world, followed by infectious and parasitic diseases and cancers. HIV/AIDS remains high on the mortality list, and it is a leading cause of death in adults ages 15 to 59 in

3. In high-income countries, an alternative measurement of disease burden, the Quality-Adjusted Life Year (QALY), is often used, especially for cost-effectiveness analysis. The QALY is a measurement of the number of years of life in perfect health gained by avoiding a given adverse condition, as compared to a DALY, which measures the number of years of life *lost* by having that same condition. QALYs, like DALYs, assess both the quality and quantity of life lived. While DALYs are calculated with specific disability weights for specific diseases, QALYs are based on the measurement of overall health status, and so the two are not easily compared.[4]
4. Available on the WHO Web site, http://www.who.int/topics/global_burden_of_disease/en/.

Box 1.1. Key Points for the Global Burden of Disease

1. Around 10 million children under the age of five die each year. Ninety-nine percent of under-five deaths occur in lower- and middle-income countries. Of these deaths, undernutrition is the underlying cause for at least 30 percent of all children under age five. Seven out of every 10 child deaths (under age 18) are in Africa and Southeast Asia.
2. Complications of pregnancy account for almost 15 percent of deaths in women of reproductive age worldwide.
3. The disease burden is uneven across regions. Southeast Asia and Africa together bore 54 percent of the total global burden of disease while accounting for only 40 percent of the world's population. Ninety percent of DALYs lost worldwide are in low- and middle-income countries; 44 percent in low income countries alone.
4. Rapidly industrializing countries are suffering a dual burden of disease with high DALYs lost from both communicable and noncommunicable diseases. Forty percent of deaths in lower-income countries are caused by category 1 diseases (infection, undernutrition, maternal complications), while noncommunicable diseases such as cardiovascular disease and stroke account for another 50 percent of deaths.
5. Cardiovascular diseases are the leading cause of death in the world. Eighty percent of all deaths are in developing countries. High-fat diets, inadequate physical activity, and smoking account for this trend.
6. Population aging is contributing to the rise in cancer; lung cancer is the most common cause of death from cancer in the world. Tobacco use, it was estimated, is a major driver of mortality, accounting for 1 out of every 10 deaths globally.
7. Mental disorders such as depression along with hearing loss and vision problems are among the top 20 leading causes of disability worldwide.
8. Injuries from road traffic accidents are a top 10 cause of death globally and expected to be the fifth leading cause of death by 2050.

Africa. Almost three-quarters of the deaths in the developing regions of Asia and Western Pacific are now due to noncommunicable chronic diseases. Malaria and self-inflicted injuries such as suicides each account for approximately 1.4 percent of global mortality. Figure 1.3 shows the breakdown of global mortality for adults and children.

Eighteen percent of all deaths are in children less than five years of age, with more than 99.9 percent of these deaths occurring in developing countries, which constitute 85 percent of the global population. Nearly half of all deaths in Africa were in children age 15 and under (largely due to communicable diseases, malnutrition, and poverty) and only 20 percent of deaths were in people age 60 years and over. In contrast, in high-income countries, only 1 percent of deaths were in children under 15 years and 84 percent of deaths were in people age 60 years and older. Overall, the importance of child mortality to the disease burden is underscored by factoring in age of death; when years of life lost is used as a mortality metric, the leading cause of death shifts to perinatal conditions, while noncommunicable diseases decrease significantly in their contribution to the global burden. Moreover, complications of pregnancy still account for almost 15 percent of deaths in women of reproductive age worldwide. Maternal morbidity and mortality rates vary among countries, and causes of high morbidity and mortality also vary.[16] Medical

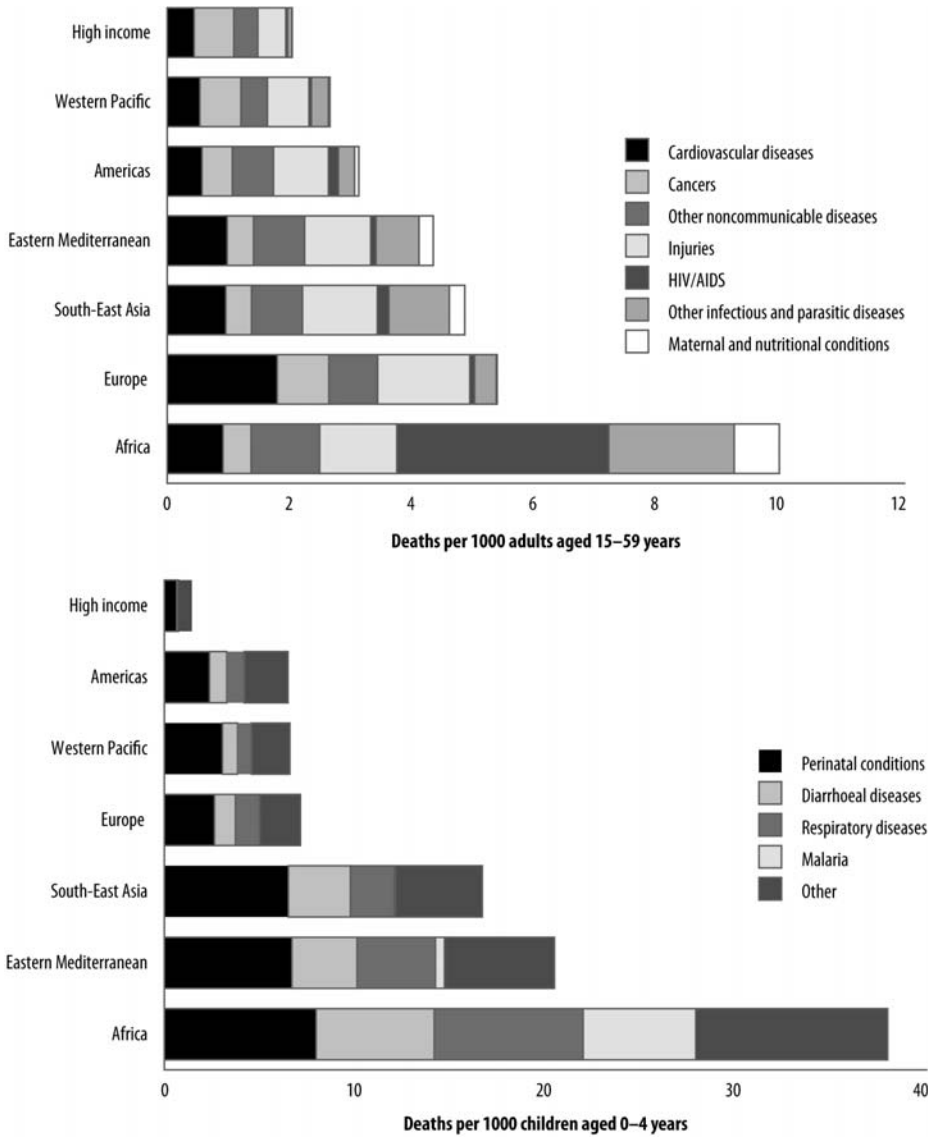

Figure 1.3 Adult (top) and child (bottom) mortality per 1000 persons. (Adapted from the 2004 WHO Global Burden of Disease.)

conditions such as hemorrhage, preeclampsia and eclampsia, obstructed labor, and complications after abortion are the primary causes of maternal death worldwide. In instances in which maternal mortality is greater than 20 per 100,000 live births, numerous nonmedical factors (such as socioeconomic, educational, and nutritional factors) usually compound the situation.[17] Although life expectancy gaps between countries, on average, have narrowed in the past 50 years, considerable variability remains in life expectancy within different social, economic, and cultural groups within countries.

Global Disease Burden

When DALYs are used to estimate the GBD, communicable diseases displace noncommunicable diseases as leading drivers of illness. Other nonfatal health outcomes emerge as important causes of disease burden: 60 percent of DALYs lost are due to premature mortality, while 40 percent are attributable to nonfatal health conditions. Globally, the two leading causes of DALYs lost are infectious diseases: lower respiratory infections and diarrheal diseases. The silent epidemic of unipolar depression is the third leading driver of DALYs lost worldwide, with the burden eighth in low-income countries and steadily rising in middle- and high-income countries. Ischemic heart disease and HIV/AIDS are the fourth and fifth leading drivers of the burden, respectively (see figure 1.3). Road traffic accidents are now the ninth leading overall cause of DALYs lost globally. The incidence of diseases related to tobacco smoking continue to rise, particularly in rapidly developing countries; chronic obstructive pulmonary disease (COPD) now ranks fifth as a cause of DALYs lost.

The greatest burden of disease is concentrated in Africa, where twice as many DALYs are lost as compared with any other region. More generally, across low-income settings, infectious diseases account for 8 of the top 10 diseases in these countries. A collection of infectious diseases including at least 13 parasitic, helminthic and bacterial infections (for example, lymphatic filaraisis, hookworm, and African river blindness) is a major cause of disability that often is not fatal but results in long-term disability and ill-health.

Risk Factors

The single leading global cause of health loss is undernutrition, responsible for 9.5 percent of the GBD, or more than 140 million DALYs. Underweight is almost exclusively a problem of low- and middle-income countries, where it is responsible for nearly 15 percent of the burden of disease in DALYs. Notably, an additional 15 percent of global DALYs lost are attributable to diet-related (overnutrition) risk factors, such as high blood pressure, high cholesterol, high Body Mass Index (BMI), heart disease, and low fruit and vegetable consumption. Among the other risk factors assessed (physical activity, addictive substances, environmental risks, and occupational risks), sexual and reproductive health is the second most important risk factor.

Unsafe sex, mainly linked to the prevalence of HIV-1 and other sexually transmitted diseases (STDs), is responsible for 6.3 percent of the GBD. The greatest burden of the risk of unsafe sex is concentrated in high-mortality, developing countries. The importance of undernutrition and unsafe sex as risk factors for disease is so staggering it is worth emphasizing; childhood and maternal underweight and unsafe sex in high-mortality developing regions of the world, which make up 38 percent of the global population, contribute as much to the loss of healthy life years as do all the injuries and diseases in the developed world combined.

In high-income countries, tobacco is a leading cause of the burden of disease (12.2 percent) along with other more proximal risk factors driving noncommunicable diseases and injuries: high blood pressure (10.9 percent), alcohol (9.2 percent), high cholesterol (7.6 percent), and high BMI (7.4 percent). Critically, these risk

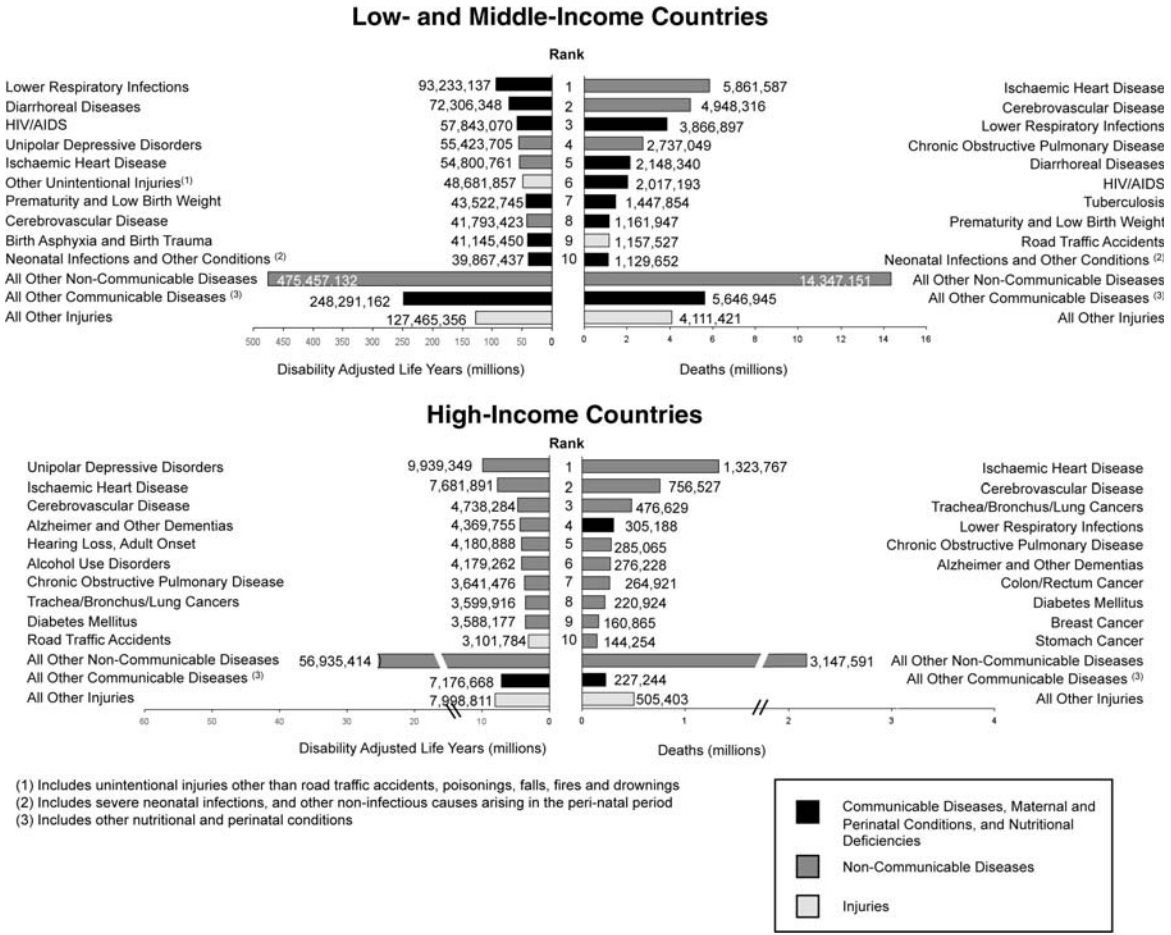

Figure 1.4 Burden of disease in detail.

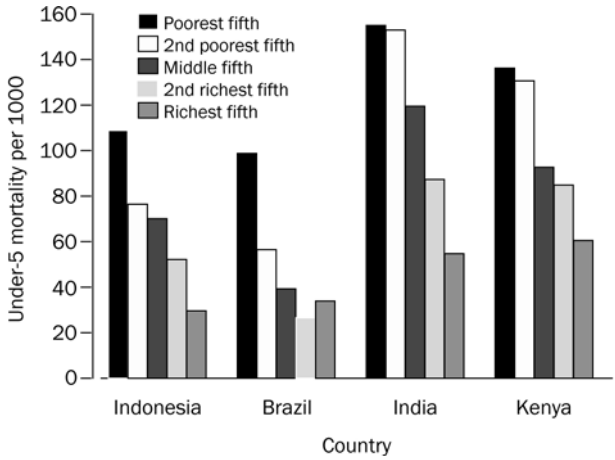

Figure 1.5 Under-five mortality rates by socioeconomic quintile of the household for selected countries. (Adapted from Victora CG et al., *Lancet*, 2003.)

factors alone contribute more to the GBD than the most common noncommunicable diseases (heart disease, depression, and stroke), highlighting the major health gains that can be realized through vigorous risk reduction and preventive efforts.

Burden of Diseases within Countries

Although most of the WHO burden data are reported on a country level and not disaggregated by wealth quintile or ethnic group, tremendous inequality often exists in the distribution of diseases within a given country. Within-country burden data are included because this information is largely missing from the country-aggregated data that most commonly are associated with the burden. It is important to appreciate the critical role that social and economic circumstance, in addition to geography and gender, play in determining the distribution and severity of disease states.

As we have discussed, developing countries shoulder a disproportionate burden of the global share of disease. Likewise, within individual countries, the burden of disease in general falls inequitably on the poorer and less advantaged populations.[18] Several groups, most notably the WHO Committee on the Social Determinants of Health, have been vocal in highlighting these within-country health inequities. As an example of within-country differences in the distribution of disease, figure 1.5 shows the rate of under-five mortality across four countries (Brazil, India, Indonesia, and Kenya). The mortality rate is over twice as high in the poorest quintile group than it is in the richest quintile group in all four study countries.[19] Such disparities, however, also are found in high-income countries such as the United States. In a landmark paper, McCord and Freeman found that mortality rates for African American males living in Harlem were at least twice the national average and, indeed, that these men were less likely than men living in Bangladesh to reach the age of 65.[20] Furthermore, diabetes and cardiovascular disease and their associated risk factors such as obesity show a general trend of higher incidence and prevalence in the lower socioeconomic quintiles.[21–23]

Indigenous populations may be at particular risk for a greater burden of disease than country averages. A recent study, for example, determined that indigenous Australians have a 60 percent higher disease burden (in DALYs) than the average for the white Australian population. Most of this additional burden is from increased susceptibility to noncommunicable diseases, to which indigenous Australians are 40 percent more susceptible than the white Australian average.[24]

PROJECTIONS OF THE BURDEN: WHAT WILL 2030 BRING?

Assuming that the trends continue, including enhanced control of communicable diseases and the increased diffusion of major risk factors for chronic diseases globally, WHO data have been used to project where the burden of disease will fall in the coming years.[3] As the recent pandemic threats of severe acute respiratory syndrome (SARS) and H1N1 swine flu highlight, projections do not account for emerging infectious threats. Furthermore, these projections are rough and essentially are based on estimates of estimates. The data indicate that the collective burden of disease is projected to decline by 10 percent from 2004 to 2030 (1.53 billion DALYs to 1.36 billion) even with a population increase of roughly 25 percent over the same period. This represents a significant reduction in the global per capita burden. It is estimated that the decrease will be driven by global reduction in diarrheal diseases, lower respiratory infections, and HIV/AIDS (HIV/AIDS drops from fifth leading cause to ninth leading cause of death). In 2030, the three leading causes of DALYs lost are projected to be unipolar depressive disorders, ischemic heart disease, and road traffic accidents. This represents quite a change from 50 years ago.

In light of the increased risk factor development for chronic diseases, in terms of mortality, the four leading causes of death globally are projected to be ischemic heart disease, cerebrovascular disease, chronic obstructive pulmonary disease, and lower respiratory infections. Large declines are projected for the main maternal, perinatal, and nutritional causes, including HIV/AIDS, TB, and malaria. Global HIV/AIDS deaths are projected to rise from 2.2 million in 2008 to a maximum of 2.4 million in 2012 and then decline to 1.2 million in 2030 (assuming coverage with antiretrovirals continues at present rates). Aging of the population in low- and middle-income countries also will increase deaths due to noncommunicable diseases in 2030. Cancers will rise from 7.4 million deaths to 11.8 million; cardiovascular deaths will rise from 17.1 million to 23.4 million deaths.

THE BURDEN AND THE DALY: WHAT ARE WE REALLY MEASURING?

What are the issues with the DALY? A common critique, though difficult to address, is the limitation of the raw data, which is uneven, and often unavailable for many countries.[4] Another issue is the decision to use a single disability weight across all regions of the world (with the assumption that paraplegia in Australia and Cameroon mean the same thing) discussed in a footnote 2.[12] A third issue is the failure of the DALY to capture comorbidities. In the United States, for example, 61 percent of women and 47 percent of men ages 70 to 79 suffer from at least two

chronic conditions, and it is likely that these conditions together are a greater burden on the individual than each condition separately would be.[25] Likewise, in developing world populations polyparasitization can compound illness significantly. The main critique we explore below is that the DALY is more a measure of physical ill health than it is the actual burden of disease, which accounts for a much richer understanding of the social, economic, and communal aspects of the burden.

Anand and Hanson have argued that a real burden of disease metric should include calculations about the circumstances, stigmas, support services, incomes, family, and friends of individuals with the illness, rather than simply taking into account age and gender.[26] This expanded definition has important intervention allocation implications. For example, if the true economic cost of the lost productivity due to poor nutrition or parasitic diseases in early childhood were taken into account, de-worming or nutritional interventions targeting these vulnerabilities might acquire renewed significance on the global agenda. Many of the parasitic diseases, in particular, account for enormous losses in economic productivity, especially in agriculture, the most prevalent economic engine in rural parts of the globe. Pediatric infections in children with soil-transmitted helminthes (hookworm) are associated with a reduction in education and school performance and attendance with adverse effects on future earnings that in aggregate can be considerable.[27] Yet, these social and economic parameters are not part of the estimation of the GBD in official reports.

The recent work done by the WHO Commission on the Social Determinants of Health has been crucial in bringing to the forefront the important contributions that the physical and social environment makes to health. Recognizing the considerable importance and contribution of the current GBD work, we hope that the future for GBD research and reporting will herald an increased accounting of the wider burden of disease, as well as the social protection approaches that can address these causes, rather than a focus on physical illness alone.

LEVERAGING THE BURDEN DATA FOR POLICY CHANGE: HOW IS IT USED?

The GBD and especially the DALY have been used to highlight existing gaps and realign prioritization of health care expenditures—the mostly revealing of which is that, when compared with their global impact, noncommunicable diseases traditionally have been underfunded.

Overall, noncommunicable chronic diseases receive about $3 per annual death compared with $1,030 per death for HIV from the World Bank, Bill & Melinda Gates Foundation, the U.S. government, and the Global Fund, as shown in figure 1.6, constituting less than 0.01 percent of overseas donor assistance (ODA).[28, 29]

Injuries and violence, expected to become the fifth leading causes of death in 2030, receive less than 1 percent of WHO funds. Nearly 87 percent of the WHO budget was directed toward combating communicable diseases, with about 12 percent going toward noncommunicable diseases.[30] Worldwide, ODA by donor countries for the reduction in tobacco use was a paltry $2.3 million.[30] No one should argue

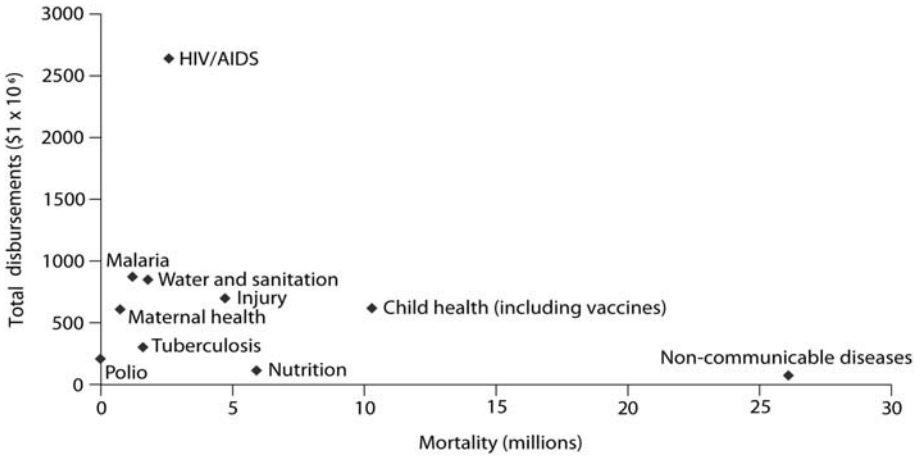

Figure 1.6 2001 worldwide mortality versus 2005 disbursements of World Bank, U.S. Government, Bill & Melinda Gates Foundation, and Global Fund to Fight HIV/AIDS, Tuberculosis and Malaria. Note: Health systems funding is not included in figure. (Adapted from Sridhar D et al., *Lancet*, 2008.)

that less funds be directed toward the prevention and treatment of infectious diseases; yet, it is now painfully clear that more funds should be applied to chronic diseases if only to mitigate the ongoing epidemic of chronic diseases in the future.

To address these misalignments between the GBD and public health spending, activists (including students) are working to make access to treatment and essential medicines reflective of the priorities outlined by the data. The striking uptake of life-saving cholesterol-lowering statin drugs, now a WHO essential medicine,[31] is a useful example of the progress that can be made. Large gains in statin utilization have been observed across Europe; for example, from 4.6 percent to just over 55 percent in the Czech Republic in the 12-year interval since 1995 and 2007, which is associated with substantial reductions in serum cholesterol levels.[32] Increased secondary and high-risk primary prevention efforts as well as increased availability of cardioprotective medicines certainly have contributed to this achievement. Currently, discussions are ongoing to implement and promote access to an inexpensive, low-dose "polypill" (containing an aspirin, diuretic, beta blocker, and a statin), which is now in Phase II clinical trials with an estimated cost of $12 per patient per year ($0.03 per day).[33] As noncommunicable disease prevalence surely will continue to rise, it is essential to provide international funding to these important prevention programs.

Yet, if there is any doubt that infectious diseases are somehow no longer important, we have failed in our presentation. These diseases continue to pose huge threats to millions around the globe. Hookworm, for instance, infects upwards of 1 billion people—many of whom who live on less than $1 per day.[34] A collection of diseases, including the neglected tropical diseases (NTDs) along with the so-called Big Three (HIV/AIDS, TB, and malaria), requires constant surveillance and elimination measures to reach the eradication milestones of smallpox and guinea worm. Drugs for hookworm, as an example, cost as little as $0.50 per year per

patient and can be adapted for inclusion in rapid high-impact packages with malaria control, making these interventions cost-effective.[34]

The challenge, then, is how best to deal with the double burden of disease. It is rapidly becoming clear that the best way forward is to invest in strengthening health systems. This is no small task. In Tanzania, there is 1 doctor for every 42,000 people, compared with 1 doctor per 500 people in the United States. Currently, more Ethiopian-trained physicians practice medicine in Chicago than in all of Ethiopia.[35] Nevertheless, the calls to actualize the Alma-Ata Declaration made 30 years ago to scale up primary care are being heard.[36] Recently, the major global donors have begun to take action. The Global Fund for HIV/AIDS, TB, and Malaria has added a health systems category intended to fund infrastructure and cross-sectoral strengthening to improve fractured systems (including sustainable financing), as well as to provide a well-trained health workforce, ensure reliable access to medical products and technologies, and create a robust information system.[37, 38] Aligning health systems with priorities in the burden, including developing a primary care essential package of medicines, vaccines, and diagnostics, is rather complex and will be highlighted by other chapters in this volume.

Intriguingly, the persisting dual burden is opening all sorts of new research questions, including whether infectious diseases and malnutrition can drive noncommunicable diseases. Currently, it is estimated that just over 20 percent of cancers are infectious in origin (for example, human papillomavirus and cervical cancer). Conversely, chronic diseases themselves may predispose one to infectious disease; a recent report shows that diabetic patients are at increased susceptibility to TB.[38] The links between chronic and infectious disease are many, from malnutrition to obesity, viruses to cancer, and infections to heart disease. The result is that poorer populations are at significantly higher risk of developing infectious diseases and also are at higher risk of suffering and dying from chronic diseases as they age.

The explosion of noncommunicable chronic diseases worldwide and the existing burden of communicable diseases pose a significant threat to the public's health. Now more than ever, the world needs continued investments in health interventions that are based on a sophisticated understanding of determinants of health and disease. Central to improving health is a better understanding of the distribution of disease both globally and locally and an improved reporting and surveillance that can direct interventions where and when they are most needed. Imagine a world in which community-trained health workers and volunteers can leverage technologies such as mobile phones, personal digital assistants, and online portals to report on disease incidence and prevalence in real time—a global mapping of disease. Although this is happening in bits and spurts, the incredible proliferation of information technology coupled with other advancements in public health just may make it possible for this century to be the one in which the greatest strides are made toward improving health in developed and developing countries alike.

REFERENCES

1. Kinsella KG. Changes in life expectancy 1900–1990. *Am J Clin Nutr.* 1992;55(6 Suppl):S1196–S1202.

2. Murray CJ, Kulkarni SC, Michaud C, et al. Eight Americas: investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Med*. 2006;3(9):e260.
3. Mathers C, Fat DM, Boerma JT, World Health Organization. *The Global Burden of Disease: 2004 Update*. Geneva, Switzerland: World Health Organization; 2008.
4. Jamison DT, Breman JG, Measham AR, et al., eds. for the Disease Control Priorities Project. 2nd ed. *Disease Control Priorities in Developing Countries*. Washington, DC: Oxford University Press/World Bank; 2006.
5. Gribble JN, Preston SH. *The Epidemiological Transition: Policy and Planning Implications for Developing Countries: Workshop Proceedings*. Washington, DC: National Academy Press; 1993.
6. DeLong JB. Cornucopia: The pace of economic growth in the twentieth century. National Bureau of Economic Research working paper w7602. Published March 2000.
7. Sen K, Bonita R. Global health status: two steps forward, one step back. *Lancet*. 2000;356(9229):577–582.
8. Bloom D, Canning D, Sevilla JP. The effect of health on economic growth: a production function approach. *World Development*. 2004;32(1):1–13.
9. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet*. 2006;367(9524):1747–1757.
10. Popkin BM. The nutrition transition and obesity in the developing world. *J Nutr*. 2001;131(3 Suppl):S871–S873.
11. Murray CJ, Lopez AD, Harvard School of Public Health, World Health Organization, World Bank. *The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability From Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020*. Cambridge, MA: Harvard University Press/Harvard School of Public Health for World Health Organization/World Bank; 1996.
12. Allotey P, Reidpath D, Kouame A, Cummins R. The DALY, context and the determinants of the severity of disease: an exploratory comparison of paraplegia in Australia and Cameroon. *Soc Sci Med*. 2003;57(5):949–958.
13. Lopez AD, Disease Control Priorities Project. *Global Burden of Disease and Risk Factors*. Washington, DC: Oxford University Press/World Bank; 2006.
14. Mathers C, Fat DM, Boerma JT, World Health Organization. *The Global Burden of Disease: 2004 Update*. Geneva, Switzerland: World Health Organization; 2008.
15. Barendregt JJ, Van Oortmarssen GJ, Vos T, Murray CJ. A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. *Popul Health Metr*. 2003;1(1):4.
16. Khan KS, Wojdyla D, Say L, Gulmezoglu AM, Van Look PF. WHO analysis of causes of maternal death: a systematic review. *Lancet*. 2006;367(9516):1066–1074.
17. Schwarcz R, Fescina R. Maternal mortality in Latin America and the Caribbean. *Lancet*. 2000;356 (Suppl):S11.
18. Marmot M. Social determinants of health inequalities. *Lancet*. 2005;365(9464):1099–1104.
19. Victora CG, Wagstaff A, Schellenberg JA, Gwatkin D, Claeson M, Habicht JP. Applying an equity lens to child health and mortality: more of the same is not enough. *Lancet*. 2003;362(9379):233–241.
20. McCord C, Freeman HP. Excess mortality in Harlem. *N Engl J Med*. 1990;322(3):173–177.

21. Rabi DM, Edwards AL, Southern DA, et al. Association of socio-economic status with diabetes prevalence and utilization of diabetes care services. *BMC Health Serv Res*. 2006;6:124.
22. Winkleby MA, Cubbin C. Influence of individual and neighborhood socioeconomic status on mortality among black, Mexican-American, and white women and men in the United States. *J Epidemiol Community Health*. 2003;57(6):444–452.
23. Cohen DA, Finch BK, Bower A, Sastry N. Collective efficacy and obesity: the potential influence of social factors on health. *Soc Sci Med*. 2006;62(3):769–778.
24. Vos T, Barker B, Begg S, Stanley L, Lopez AD. Burden of disease and injury in Aboriginal and Torres Strait Islander peoples: the indigenous health gap. *Int J Epidemiol*. 2009;38(2):470–477.
25. Singer BH, Ryff CD. Neglected tropical diseases, neglected data sources, and neglected issues. *PLoS Negl Trop Dis*. 2007;1(2):e104.
26. Anand S, Hanson K. Disability-adjusted life years: a critical review. *J Health Econ*. 1997;16(6):685–702.
27. Miguel EA, Kremer M. Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica*. 2004;72:159–217.
28. Sridhar D, Batniji R. Misfinancing global health: a case for transparency in disbursements and decision making. *Lancet*. 2008;372(9644):1185–1191.
29. Yach D, Hawkes C. *Towards a WHO Long Term Strategy for Prevention and Control of Leading Chronic Diseases*. Geneva, Switzerland: World Health Organization; 2004.
30. Stuckler D, King L, Robinson H, McKee M. WHO's budgetary allocations and burden of disease: a comparative analysis. *Lancet*. 2008;372(9649):1563–1569.
31. Kishore SP, Herbstman BJ. Adding a medicine to the WHO model list of essential medicines. *Clin Pharmacol Ther*. 2009;85(3):237–239.
32. Kotseva K, Wood D, De Backer G, De Bacquer D, Pyorala K, Keil U. Cardiovascular prevention guidelines in daily practice: a comparison of EUROASPIRE I, II, and III surveys in eight European countries. *Lancet*. 2009;373(9667):929–940.
33. Yusuf S, Pais P, Afzal R, et al. Effects of a polypill (Polycap) on risk factors in middle-aged individuals without cardiovascular disease (TIPS): a phase II, double-blind, randomised trial. *Lancet*. 2009;373(9672):1341–1351.
34. Molyneux DH, Hotez PJ, Fenwick A. "Rapid-impact interventions": how a policy of integrated control for Africa's neglected tropical diseases could benefit the poor. *PLoS Med*. 2005;2(11):e336.
35. Mullan F. Responding to the global HIV/AIDS crisis: a Peace Corps for health. *JAMA*. 2007;297(7):744–746.
36. Walley J, Lawn JE, Tinker A, et al. Primary health care: making Alma-Ata a reality. *Lancet*. 2008;372(9642):1001–1007.
37. World Health Organization. *Everybody's Business: Strengthening Health Systems to Improve Health Outcomes—WHO's Framework for Action*. Geneva, Switzerland: World Health Organization; 2007.
38. Jeon CY, Murray MB. Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies. *PLoS Med*. 2008;5(7):e152.

This page intentionally left blank

CHAPTER 2

Population Growth, Policy Options, and Public Health in the Developing World

John Bongaarts, PhD

The modern expansion of human numbers started in the late 18th century with the onset of a long-term decline in the European and Northern American death rates. This reduction in mortality was the consequence of a lower incidence of epidemics and famines and improvements in standards of living, levels of nutrition, and basic public health measures. In the two centuries from 1750 to 1950, the world's population more than tripled in size, from 0.8 to 2.5 billion, after having experienced only slow and uneven growth before 1750. Since the middle of the 20th century, a huge new spurt of growth occurred in Africa, Asia, and Latin America; once again this growth was the consequence of rapid declines in mortality. More people have been born since 1950 than were alive in that year. In 2010, the world's population reached 6.9 billion, an increase of 173 percent since 1950. Growth is expected to continue for several more decades, with the total expected to reach 9.15 billion in 2050 before stabilizing later in this century.[1] This projection implies that the world's population will have grown by more than tenfold—from 0.8 to 9.15 billion—between 1750 and 2050.

Projections for the next half-century expect a highly divergent world, with stagnation or potential decline in parts of the developed world and continued quite rapid growth in the least developed regions. Wide ranges of policies are available for governments concerned about adverse effects of these trends. In countries with low fertility (in Europe and East Asia), efforts to encourage childbearing are being considered to reverse or mitigate the adverse impact of rapid population aging and population decline. The opposite is true in much of the developing world where rapid population growth remains a key concern.

This chapter reviews population projections to 2050 for the world and major regions. It then identifies the demographic factors responsible for continued expansion of human numbers and discusses policy options for slowing population growth in the developing world. The conclusion summarizes the wide-ranging public health benefits resulting from slower growth and reduced birth rates.

FUTURE POPULATION TRENDS

The most recent United Nations (UN) medium projection expects the population of the world to continue to grow at least until 2050, adding 2.2 billion to the 2010 population of 6.9 billion. Nearly all of this future growth will occur in the Southern Hemisphere—that is, Africa, Asia (excluding Japan, Australia, and New Zealand), and Latin America—where population size is projected to increase from 5.7 to 7.9 billion between 2005 and 2050 (see table 2.1). In contrast, in the Northern Hemisphere (Europe, Northern America, Japan, and Australia/New Zealand), population size is expected to remain virtually stable, growing slowly from 1.24 to 1.28 billion between 2010 and 2050.

Projected population trends vary widely among world regions (see table 2.1). In 2010, Asia had a population of 4.17 billion, more than half of the world total, and its population is expected to grow by one-fourth to 5.23 billion by 2050. Africa, with 1.03 billion inhabitants in 2005, is likely to experience by far the most rapid expansion, doubling in size by 2050. Latin America, with 0.59 billion in 2010, is the smallest of the Southern Hemisphere regions with a growth pattern similar to Asia. Trends for the two principal regions in the Northern Hemisphere will diverge between 2010 and 2050: an increase from 0.35 to 0.45 billion in North America, but a decline from 0.73 to 0.69 billion in Europe. One consequence of these diverse regional growth rates is that the regional distribution of population will shift significantly over time. While Asia's (60.3 percent) and Latin America's (8.5 percent) proportions of the world total remain more or less unchanged, Europe's proportion declines by more than one-third (from 10.6 percent to 7.6 percent) and Africa's proportion rises (from 15 percent to 21.8 percent). Between 2010 and 2050, the Northern Hemisphere's share is expected to decline from 17.9 percent to 13.9 percent.

It may seem surprising that population growth continues at a rapid pace in Sub-Saharan Africa despite the severe AIDS epidemic that is causing large numbers of

Table 2.1
Population Estimates (1950–2005) and Projections (2005–2050), by Region

	Population (billions)			Percent Increase	
	1950	2010	2050	1950–2010	2010–2050
Africa	0.23	1.03	2.00	355	93
Sub-Sahara	0.18	0.86	1.75	371	103
Asia[a]	1.40	4.17	5.23	197	26
China	0.54	1.35	1.42	148	5
Latin America	0.17	0.59	0.73	252	24
Europe	0.55	0.73	0.69	34	−6
Northern America	0.17	0.35	0.45	105	28
Southern Hemisphere	1.72	5.67	7.87	230	39
Northern Hemisphere	0.81	1.24	1.28	52	3
World	2.53	6.91	9.15	173	32

Source: United Nations 2009.
Note:
[a] Includes Oceania.

deaths. The impact of this epidemic can be assessed by comparing the standard population projection for Sub-Saharan Africa, which includes the epidemic's effects, with a separate hypothetical projection in which AIDS mortality is excluded. The former projects the Sub-Saharan population to increase from 0.86 billion in 2010 to 1.75 billion in 2050, the latter expects population size to reach 1.87 billion in 2050. The difference between these projections with and without AIDS equals 0.12 billion (−6.5 percent) and is due to deaths from AIDS as well as to the absence of descendents from people who died from AIDS. Despite the elevated death rate from AIDS and other diseases, the population of Sub-Saharan Africa is expected to double by 2050.

WHY POPULATION GROWTH CONTINUES

The future growth that is expected for any population is attributable to four main demographic factors: high fertility, declining mortality, population momentum, and migration.[2, 3]

High Fertility

Fertility will be considered "high" if it exceeds the replacement level of two children per woman. Replacement is a critical factor in population projections because it equals the fertility level that, if maintained over time, produces zero population growth. Positive deviations from replacement lead in the long run to persistent population growth and negative deviations lead to population decline. Currently, replacement fertility equals 2.3 births per woman (bpw) in the Southern Hemisphere and 2.1 in the Northern Hemisphere (these levels exceed two because children who die before reaching the reproductive ages have to be replaced with additional births, and because the sex ratio at birth slightly exceeds one). Despite declines in many countries over recent decades, fertility in 2005–2010 averaged 4.6 bpw in Africa, 2.4 in Asia, and 2.3 in Latin America.[1] Since these levels are above replacement, high fertility remains one of the key forces contributing to further growth in the Southern Hemisphere. In contrast, fertility is now below replacement in Europe (1.5 bpw) and East Asia (1.8 bpw) and close to replacement in North America (2 bpw) (see figure 2.1).

The UN projections from 2010–2050 assume that fertility will decline in the future in countries where it is now above replacement, stabilizing at 1.85 bpw in the long run. Asia and Latin America are assumed to reach this point before 2050, but Sub-Saharan Africa is expected to remain above replacement until after 2050 (declining from current levels to 2.5 bpw in 2050). In countries where fertility is now below replacement (mostly in the Northern Hemisphere) it is assumed to rise slightly but not enough to return to replacement.

From a policy perspective, it is important to recognize that high fertility can in turn be attributed to two distinct underlying causes. The first is *unwanted childbearing*, which occurs when women have more children than their desired family size. About one-fifth of fertility in the Southern Hemisphere (excluding China) is unwanted and a larger proportion is mistimed.[4] In addition, an estimated 36 million abortions are performed each year in the developing world—many of them under

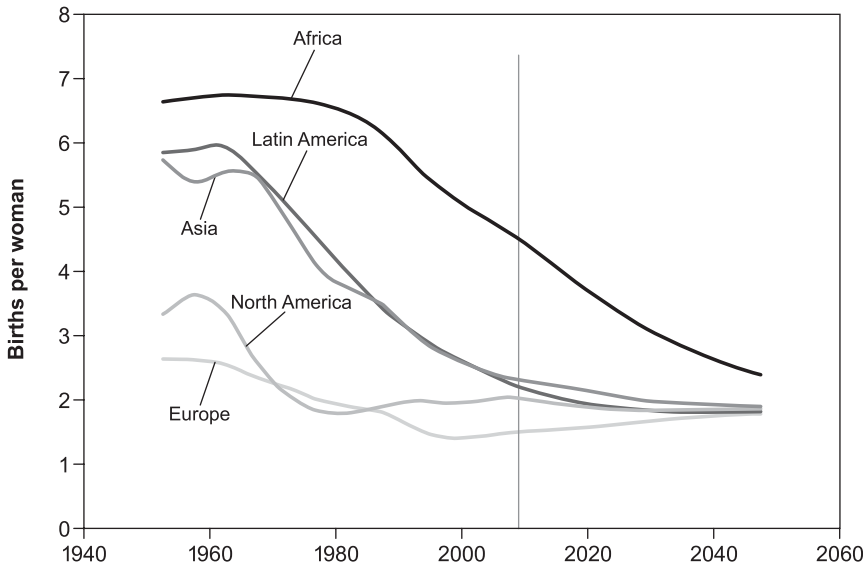

Figure 2.1 Total fertility rate, estimates 1950–2010 and projections to 2050, Sub-Saharan Africa. (Source: United Nations, 2009.)

unsafe conditions.[5] The second cause is a high *desired family size*. In most developing countries, the family size that is desired by women still exceeds two children; in some areas, such as Sub-Saharan Africa, desired family size is typically near five children.[4] As will be discussed, these two causes are associated with distinct policies.

Declining Mortality

Rapid declines in mortality have been the main cause of population growth in the past and will continue to be a factor in the future. Global life expectancy at birth in 2005–2010 stood at 68 years, more than double the preindustrial life expectancy of about 30 years.[1] Since the middle of the 20th century, the Southern Hemisphere has experienced exceptionally rapid improvements in life expectancy—from 41 years in 1950–1955 to 66 years today; these improvements are in large part due to the global spread of medical and public health technology (immunization, antibiotics) after World War II. In 2005–2010, Latin America reached life expectancy levels similar to those in the Northern Hemisphere during the 1960s, and Asia is not far behind. Africa's mortality level has been the highest, with a current life expectancy of 54 years, which primarily is a result of the severe AIDS epidemic in parts of this continent. In the Northern Hemisphere, mortality was already low in the 1950s, but life expectancy has continued to rise, reaching 77 years in 2005–2010 (see figure 2.2).

Over the next half-century, UN projections assume life expectancy to continue to rise in all regions. By 2050, Asia and Latin America are both expected to have mortality conditions similar to those in the Northern Hemisphere in 2010, but Africa will continue to lag, in part because the continent is most heavily affected by the AIDS epidemic. Achieving long-term reductions in mortality may become more difficult as countries reach ever-higher levels of life expectancy.

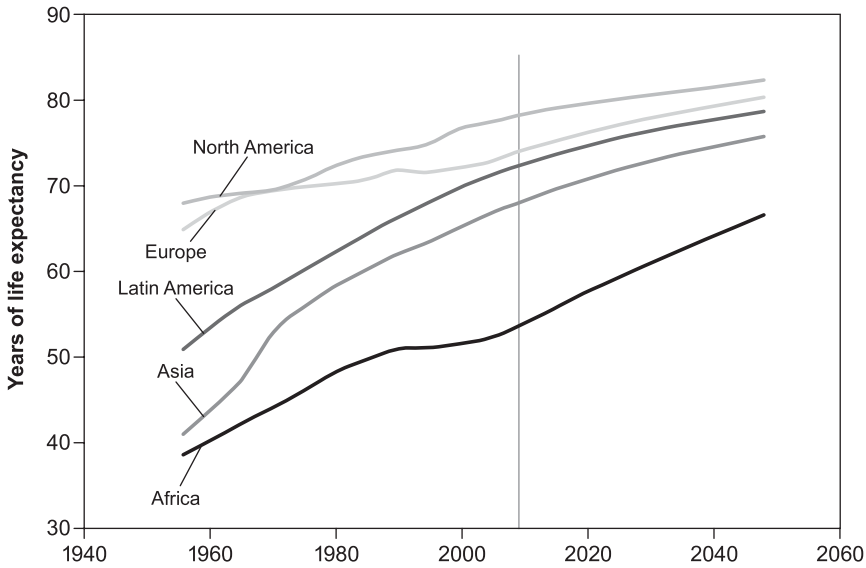

Figure 2.2 Life expectancy, estimates 1950–2010 and projections to 2050. (Source: United Nations, 2009.)

Population Momentum

Even if fertility could immediately be brought to the replacement level with constant mortality and zero migration, population growth would continue in many countries, particularly in the Southern Hemisphere. The reason for this is a young age structure, which is the result of high fertility and low mortality and rapid population growth in recent decades. With a large proportion of the population under age 30, further growth over the coming decades is assured. The relative abundance of these young people results in a birth rate that is higher than the death rate even if fertility is at replacement. This effect on future growth is called population momentum.[2]

Migration

Migration can significantly affect population growth in some countries (for example, into the United States, and out of other countries, for example, Mexico), but it often has a small demographic impact in others. At the world level, migration is necessarily zero. Most migrations between countries and between regions within countries are voluntary as individuals and families move in search of a better life. But there are also substantial flows of involuntary migrations associated with wars and civil conflicts, which lead to drastic reductions in standards of living and poor health conditions.

COMPONENTS OF FUTURE POPULATION GROWTH

The contribution of each of these four demographic factors to future population growth can be estimated with a simple series of hypothetical projections. In these

projections, the influence of the different factor is removed in successive step. Three projections are involved:

1. The *standard* projection is determined by all four factors.
2. The *replacement* projection is identical to the standard projection but fertility is set to the replacement level from 2005 onward. This projection is determined by the young age structure, declining mortality, and migration.
3. The *momentum* projection sets fertility to replacement, holds mortality constant, and has no migration, so it is only affected by the young age structure.

Figure 2.3 illustrates these projections for the developing world (excluding China). From a baseline level of 4.0 billion in 2010 the standard, replacement and momentum projections yield population sizes of 6.4, 6.1, and 5.9 billion, respectively, in 2050. Table 2.2 gives results for other major regions of the developing world.

The separate impact of each of these growth factors is calculated as the percent difference between the projections with and without the factor. For example, the effect of high fertility equals the percent difference between the 2050 population in the standard projection and the 2050 population in the replacement projection. These results are summarized in table 2.3, which presents estimates of these effects by region. In the Southern Hemisphere (excluding China) the high fertility effect equals 6 percent because the standard projection (with high fertility) is 6 percent higher than the projection with replacement fertility. The high fertility effect varies considerably by region ranging from a high of 37 percent in Sub-Saharan Africa to a low of −11 percent in China. As expected, this effect is highly correlated with the current fertility level. Note that this effect is negative in China and Latin America because fertility is assumed to decline below replacement for all or part of the projection period.

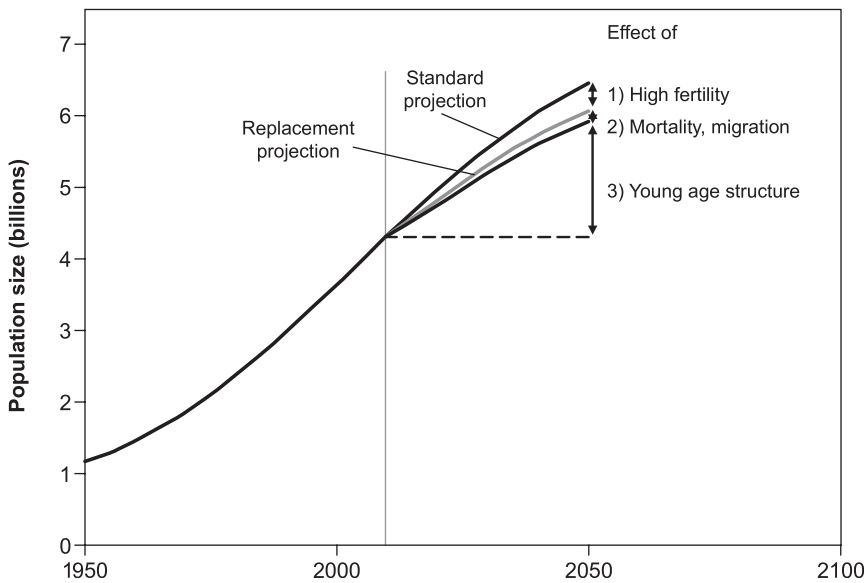

Figure 2.3 Alternative population projections for the Southern Hemisphere (excluding China) 2010–2050.

Table 2.2
Alternative Projections to 2050 for Regions of the Developing World

	Population projections to 2050 (billions)			
	Standard (High fertility, young age structure, mortality and migration)	Replacement (Young age structure, mortality and migration)	Momentum (Young age structure)	Baseline (2010)
Africa	2.0	1.5	1.5	1.0
Sub-Sahara	1.8	1.3	1.2	0.9
Asia	5.2	5.5	5.3	4.2
China	1.4	1.6	1.5	1.4
Latin America	0.7	0.8	0.8	0.6
South (excl. China)	6.4	6.1	5.9	4.3
South	7.9	7.7	7.5	5.7

Source: Based on United Nations statistics 2009.

The young age structure effect ranges from 44 percent in Sub-Saharan Africa to 14 percent in China, reflecting population momentum. For the Southern Hemisphere, population size is expected to rise by 39 percent overall between 2010 and 2050, with momentum (32 percent) accounting for most of this rise.

The effects of declining mortality (for adults only because mortality at younger ages is included in replacement fertility) and migration are small. The combined effects on population growth is around 3 percent in the different regions except in Latin America (0 percent) where the effects of migration and declining mortality offset one another.

Table 2.3
Effects of Components on Population Growth, 2010–2050

	Components of population			
	High Fertility	Young age structure	Mortality and migration	Combined (multiplicative)
Africa	31	43	3	93
Sub-Sahara	37	44	3	103
Asia	−4	27	3	26
China	−11	14	3	5
Latin America	−6	32	0	24
South (excl. China)	6	37	2	50
South	3	32	3	39

Source: Based on United Nations 2009.

POLICY RESPONSES TO RAPID POPULATION GROWTH

The difficult task of reducing poverty and bringing about sustainable development in the South will be made even harder by the expected addition of more than two billion people by the middle of the next century. Efforts to slow this population expansion should of course not include increases in mortality, and out-migration is not a realistic option for most countries. The focus therefore has to be on reducing fertility and reducing momentum. The following broad policy options can be pursued.[2]

Reduce High Fertility

High fertility has two subcomponents: unwanted and high wanted fertility. Separate policies are needed to address these issues

Reduce Unwanted Fertility and the Unmet Need for Contraception by Strengthening Family Planning Programs, Including Abortion Services

In the developing world, 137 million women who do not want to get pregnant are not using contraception.[6] This is the case despite efforts by a number governments to provide family planning services, which often are inadequate. The key cause of this unmet need for contraception is that contraception is quite costly to individuals in terms of the commodities (pills, condoms, intrauterine devices, and so on), transportation, and reimbursement of providers of contraceptives and health care services; this is the case even when the government provides subsidies. In addition, significant noneconomic costs affect use of contraception, such as health concerns, social disapproval, and spousal resistance, as well as unnecessary medical barriers (requiring a doctor instead of a nurse or other trained health care worker to provide certain contraceptives).[7] This unmet need is responsible for most of the 76 million unplanned pregnancies that occur each year. About half of these pregnancies end in abortion and the other half end in births; both contribute unnecessarily to health risks for mothers and children, to the cost of raising families, and to the adverse impact of population growth.

The existence of a high rate of unintended pregnancy and a large unmet need for contraception (first documented in the 1960s) convinced policy makers that family planning programs were necessary and would be acceptable and effective. Accordingly, in recent decades many governments in the Southern Hemisphere have implementation voluntary family planning programs. The aim of these programs is to provide information about and access to contraception to permit women and men to take control of their reproductive lives and avoid unwanted childbearing. The choice of voluntary family planning programs as the principal policy instrument to reduce fertility is based largely on the documentation of a substantial level of unwanted childbearing and the unsatisfied demand for contraception. In addition, the effectiveness of this approach was supported by experiments such as the one conducted in the Matlab district of rural Bangladesh.[8]

When this experiment began in the 1970s, Bangladesh was one of the poorest and least developed countries, and there was considerable skepticism that reproductive behavior could be changed in such a setting. Comprehensive family planning and reproductive health services were provided in the treatment area of the experiment.

The results of these improvements were immediate and pronounced, with contraceptive use rising sharply. No such change was observed in the comparison area. The differences between these two areas in contraceptive use and fertility have been maintained over time. The success of the Matlab experiment demonstrated that appropriately designed services could reduce unmet need for contraception even in traditional settings. Unfortunately, the opposition from conservative governments and institutions (in particular the Bush administration and the Vatican) has limited the investments made in these services in many countries.

Reduce the Demand for Large Families Through Investments in Human Development

Although family planning programs claim most of the attention of population policy makers and of the resources at their disposal, their potential effect is largely limited to reducing unwanted childbearing and the unmet need for contraception. Because such programs are voluntary, they cannot reduce fertility below the level wanted by couples and they cannot bring about population stabilization in countries where on average the desired number of children still exceeds two. Many individuals and couples continue to want and have large families, in part because of fears of infant and child mortality as well as the need for children to support them in family enterprises and to support them in old age. High demand for children remains a fundamental cause of population growth in many developing countries and in Sub-Saharan Africa in particular.

In many societies, sons are valued more than daughters, in part because families feel they cannot rely on daughters for their future security. As a result of this gender preference, larger numbers of births are needed to ensure the survival of sons. Son preference is common in traditional societies and, by becoming embedded in cultural norms, this preference is perpetuated as societies develop and reach low levels of fertility. In a number of contemporary societies in Asia, especially in China and India, preference for a son has led to high sex ratios at births as women resort to sex-selective abortion to ensure the births of sons.

Countries in which wanted fertility is high will need families to address these gender preferences to complete their fertility transition to replacement fertility. Such declines are usually achieved by improvements in socioeconomic conditions. Evidence indicates that desired fertility is most responsive to improvements in human development, in particular female education and child survival.[9–11] This conclusion is strongly supported by the fact that low fertility has been achieved in some poor societies such as Sri Lanka and the state of Kerala in India. Although poor, these populations have high levels of literacy and female empowerment as well as low infant mortality and ready access to methods of family planning. The immediate effect of declines in infant and child mortality is typically a boost to population growth because declines in death rates occur more rapidly than in birth rates. Mortality declines, however, are essential to bringing about lower fertility in the long run because the uncertainty surrounding the survival of children in high mortality countries has to be removed (this is one of the key rationales for parents to have many births).

Another more recent approach to stimulating social development involves microcredit schemes that can serve as potent agents of social change for poor

women. Grameen Bank conducted the first experiment with group-based lending in 1978 in Bangladesh; but similar programs are now being implemented in other developing countries.[12] Participation in microcredit schemes empowers women and raises their autonomy and decision making within the household. In Bangladesh, credit group participants showed an increased use of contraception.[13] Most governments already pursue investments in human and other development initiatives, independent of their potential role in lowering the rate of childbearing. The demographic benefits of these social policies simply strengthen the rationale for implementing them.

Address the Momentum of Population Growth

While a young age structure—the key force behind population momentum—is not amenable to modification, there is an option to offset momentum that has received relatively little attention in past policy debates. Further reductions in population growth can be achieved if the average age at which women begin childbearing rises (by delaying the first birth) and through wider spacing between births. Young women in many cultures often have little choice about whether or not to have sexual relations, when or whom to marry, and whether to defer childbearing. Short intervals between generations are often a result of the pressures on young women to marry and to bear children early as a means of finding social acceptance and long-term economic security. The early onset of fertility, and the close spacing of births, present health risks to girls and young women and limit their education and livelihood possibilities. Delaying the onset of childbearing and providing girls the opportunity to receive at least a secondary school education therefore not only can reduce population growth, but also can significantly improve personal well-being and the quality of family life, especially for women.

Governments that wish to encourage later childbearing have several options at their disposal. National legislation to raise the age at marriage has been moderately effective in a few countries, such as Tunisia and China. However, legislation has the drawback that it attempts to force rather than encourage changes in social customs that involve not only the young people but also their families. Therefore, indirect less-regulatory approaches are likely to be more effective and acceptable. A greater investment in the education of girls, particularly at the secondary level, is the most obvious example. The longer girls stay in school, the later they marry and the greater the delay in childbearing. In general, supportive measures that enhance adolescents' reproductive health, education levels, and income-generating potential will lead to more rapid human capital development and to increased productivity, and also will offset population momentum.[14]

Migration policies will not be discussed here because existing practices vary widely among countries and encouraging citizens to leave the country has little appeal to governments, except for temporary employment.

IMPLICATIONS FOR PUBLIC HEALTH

The population policy options discussed in the preceding section, if implemented, would improve public health in several ways.

Reduced Stress on the Natural Environment

The rapid ongoing growth in human numbers and in consumption per capita has raised demands on our natural environment to unprecedented levels. These demands are evident in several alarming trends: rising food and energy costs, global climate change, widespread deforestation, loss of biodiversity, shortages of fresh water, depletion of soils, and rising pollution levels. Prospects are grimmest for the poorest countries (most of them in Sub-Saharan Africa) with limited natural resources and extremely rapid population growth.[15] For example, Niger's population is projected to quadruple in size—from 13 million to 53 million—between 2005 and 2050 even though available arable land is extremely limited and threatened by desertification, and much of the current population lives on the edge of famine. Slower population growth would help relieve such stresses.

Slow Growth in Demand for Health Services

Low-income countries tend to have limited public services (health care, education, municipal), a largely untrained labor force, and weak infrastructure (roads, water supply, electricity, telecommunications, and so on). As governments struggle to overcome these problems, the situation is made more difficult by the growing populations that need to be served. In the most rapidly expanding populations, new services, new graduates, and new infrastructure have to be created at a rate of 3 or 4 percent per year simply to maintain conditions and to prevent their deterioration. A reduction in the birth rate and in population growth makes meeting these demands on the health care and education systems and on infrastructure more manageable.

More Rapid Economic Growth

Rapid population growth and high fertility are among the key causes of poverty, which in turn contributes to poor health. Conversely, rapid fertility decline creates a so-called demographic dividend, which boosts economic growth for up to a few decades by increasing the size of the labor force relative to both young and old dependents and by stimulating savings.[16]

Decline in Unwanted and Mistimed Births and Abortions

Each year 76 million unintended pregnancies in the developing world result in 142,000 pregnancy-related deaths among women and 1.4 million infant deaths.[6] Strengthening family planning programs can significantly reduce unintended pregnancies thus averting a proportion of these deaths as well as other pregnancy-related adverse health effects.

Declines in High-Risk Pregnancies

The risk of adverse health effects and maternal mortality vary with age and parity. The highest risks occur at the lowest and highest ages and among nulliparous and high parity women.[17] As desired and achieved family sizes decline so too does the proportion of pregnancies at high ages and high parities. In addition, delays in the onset of childbearing reduce the pregnancy related health risks among young adolescents.

Wider Birth Spacing

Short intervals between successive births are associated with higher mortality of infants at the beginning and end of the interval and contribute to maternal depletion.[18] In traditional societies, birth spacing is achieved by prolonged breastfeeding, which leads to periods of postpartum amenorrhea of up to a year or more. These traditional breastfeeding practices are eroding with the increasing availability of commercial infant food. Cleland et al. estimated that about 1 million deaths among children under age five could be avoided by eliminating interbirth intervals of less than two years. Increased use of contraception for spacing birth is therefore an effective way of reducing infant and child mortality.[19]

CONCLUSION

The unprecedented pace at which the world's population has grown over recent decades has had an adverse impact on social and economic development, on health care, and on the environment. Despite substantial and partially successful efforts to reduce growth in the less-developed countries, this expansion of human members is expected to continue at a rapid pace over the next decades with nearly all of this growth occurring in Africa, Asia, and Latin America. The preceding analysis identifies a range of multisectoral options for governments that consider population growth rates higher than desirable. Three key strategies are proposed: (1) strengthen family planning programs to provide women with the knowledge and means to regulate their fertility; (2) emphasize "human development," in particular education, gender equality, and child health; and (3) encourage delays in subsequent childbearing.

To be effective, population policies should include but also go beyond the provision of family planning services. Voluntary fertility reduction as a societal development goal is best achieved through mutually reinforcing investments in family planning, reproductive health, and a range of socioeconomic and education measures. Such policies operate beneficially at both the macro and micro levels; the same measures that slow population growth also improve individual health and welfare. The implementation of these population policies could have multiple benefits for public health as a result of reduced stress on the environment, slower growth in demand for health and other public services, more rapid economic growth, decline in unwanted and mistimed births and abortions, declines in high-risk pregnancies, and wider birth intervals. In the absence of further investments by governments and the international community in these policies, large further increases in the populations of developing countries will jeopardize ongoing efforts to reduce poverty and achieve sustainable development.

REFERENCES

1. United Nations, Population Division. *World Population Prospects 2008*. New York: United Nations, Population Division; 2009.
2. Bongaarts J. Population policy options in the developing world. *Science.* 1994;263(5148):771–776.
3. Bongaarts J, Bulatao R. Completing the demographic transition. *Popul Dev Rev.* 1999;25(3):515–529.

4. Westoff CF, Bankole A. *Reproductive Preferences in Developing Countries at the Turn of the Century*. Calverton, MD: ORC Macro; 2002. DHS Comparative Reports, No.2.
5. Alan Guttmacher Institute. *Sharing Responsibility: Women, Society and Abortion Worldwide; Hopes and Realities: Closing the Gap between Women's Aspirations and Their Reproductive Experiences.* New York: Alan Guttmacher Institute; 1999.
6. Susheela S, et al. *Adding it up: The Benefits of Investing in Sexual and Reproductive Health Care*. New York: Alan Guttmacher Institute/United Nations Population Fund; 2003.
7. Casterline J, Sinding SW. Unmet need for family planning in developing countries and implications for population policy. *Popul Dev Rev.* 2000;26(4):691–723.
8. Cleland J, Phillips JF, Amin S, Kamal GM. *The Determinants of Reproductive Change in Bangladesh: Success in a Challenging Environment*. Washington, DC: The World Bank; 1994.
9. Caldwell JC. Mass education as a determinant of the timing of fertility decline. *Popul Dev Rev.* 1980;6(2):225–255.
10. Jejeebhoy SJ. *Women's Education, Autonomy, and Reproductive Behavior: Experience from Developing Countries*. Oxford: Clarendon Press; 1995.
11. Sen A. *Development as Freedom*. New York: Knopf; 1999.
12. Yunus M. *Banker to the Poor: Micro-Lending and the Battle against World Poverty*. 3rd ed. New York: Perseus Books Group; 2007.
13. Steele F, Amin S, Naved RT. Savings/credit group formation and change in contraception. *Demography.* 2001;38(2):267–282.
14. Lloyd CB, ed. for National Research Council and Institute of Medicine, Division of Behavioral and Social Sciences and Education, Committee on Population and Board on Children, Youth, and Families, Panel on Transitions to Adulthood in Developing Countries. *Growing Up Global: The Changing Transitions to Adulthood in Developing Countries*. Washington, DC: The National Academies Press; 2005.
15. Alexandratos N. Countries with rapid population growth and resource constraints: issues of food, agriculture, and development. *Popul Dev Rev.* 2005;31(2):237–258.
16. Birdsall N, Kelley A, Sinding S, eds. *Population Matters: Demographic Change, Economic Growth, and Poverty in the Developing World*. Oxford: Oxford University Press; 2001.
17. World Health Organization, Family Planning and Population Division of Family Health. *Health Benefits of Family Planning.* Geneva, Switzerland: World Health Organization; 1994.
18. Conde-Agudelo A, Rosas-Bermudez A, Kafury-Goeta AC. Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. *JAMA.* 2006;295:1809–1823.
19. Cleland J, et al. Family planning: the unfinished agenda. *The Lancet.* 2006;368 (9549):1810–1827.

This page intentionally left blank

CHAPTER 3

Urbanization and Public Health: Evidence, Challenges, and Directions

Sandro Galea, MD, MPH, DrPH, and David Vlahov, PhD, RN

INTRODUCTION

Urbanization is likely the single most important demographic shift worldwide during the past century and in the new century, and it represents a sentinel change from how most of the world's population has lived for the past several thousand years.[1] Current estimates suggest that the trend toward an urbanizing world will continue well into the 21st century.[2] At the beginning of the 19th century only 5 percent of the world's population was living in urban areas. By the end of the century, about 46 percent of the world's population was living in urban areas.[3] There are approximately 50,000 urban areas in the world today and almost 400 cities containing a population of 1 million people or more.[4] Around 1940, the New York metropolitan area became the first urban area to become a megacity containing more than 10 million inhabitants. In the early 21st century, there are more than 15 megacities worldwide.[5] Overall global population growth in the next 30 years primarily will be in cities. Current projections suggest that more than half the world's population will be living in urban areas by 2007 and that nearly two-thirds of the world's population will live in urban areas within the next 30 years. By 2010, approximately 400 million people will live in urban cities worldwide.[3]

We might expect such a shift in how the majority of the world's population lives to have health implications. Indeed, researchers, both in the popular press and in the academic literature, have long been interested in cities and how they may affect the public's health. Writers from several eras in western European history considered cities as places that were detrimental to health, and in many ways, for much of history, cities were characterized by features that were linked unquestionably to poor health. Charles Dickens's novels detail and offer insights into the

This chapter was originally published as Sandro Galea and David Vlahov, "Urban Health: Evidence, Challenges, and Directions," *Annual Review of Public Health*, Vol. 26: 341–365. Reprinted, with permission, from the *Annual Review of Public Health*, Volume 26 © 2009 by Annual Reviews, www.annualreviews.org.

difficulties of city life in the 19th century.[6] As cities assumed a greater role in the life of European countries, population density, numbers of marginalized populations, pollution, and crime frequently increased, resulting, in many countries, in worse health in cities than outside of cities.[7, 8] Multiple writers, commentators, and social theorists observed the problems endemic to these growing cities and suggested that the cities themselves had a role in shaping individual well-being.[9–12]

Whereas writers in the 18th and 19th centuries overwhelmingly noted a connection between the urban context and poor health, the urban environment in many Western cities improved dramatically at the turn of the 20th century, and coincident with this sanitary awakening, the health of urban populations improved. One historical analysis showed that, although for much of the nineteenth century infant mortality rates in Imperial Germany were higher in urban area than they were in nonurban areas, infant mortality rates improved dramatically in urban areas starting in the 1870s, which preceded a comparable decline in mortality in the rest of the country.[13] This analysis suggested that improvements in the urban environment were responsible for this rapid improvement in infant health in Imperial Germany and that this pattern was typical of the pattern observed at around the same time in many European industrial societies. In the 21st century, in many countries, including the United States, aggregate health, as measured by life expectancy, all-cause mortality, and many other health indicators, is actually better in many urban areas than it is in nonurban areas.[14]

What then is urban health, and why should we concern ourselves with urban health as a specific subject of inquiry? As urban living becomes the predominant social context for most of the world's population, the ubiquity of urban living promises both to shape health directly and indirectly to affect what we typically consider risk factors or determinants of population health. Therefore, despite the truism that the urban context inherently shapes population health in cities, not all public health is urban health. We consider urban health research to be the explicit investigation of the relation between the urban context and population distribution of health and disease. Urban health, then, concerns itself with the determinants of health and diseases in urban areas and with the urban context itself as the exposure of interest. As such, defining the evidence and research direction for urban health requires that researchers and public health professionals pay attention to theories and mechanisms that may explain how the urban context may affect health and to methods that can better illustrate the relation between the urban context and health. To that end, in this review, we first discuss what we mean when considering urban areas, then we address potential mechanisms that can explain the relation between the urban context and health. We discuss particular challenges in the study of urban health and conclude with directions for potential research and practice.

CITIES AND THEIR ROLE IN THE WORLD

Cities are not static, and the very density and diversity that characterize most cities make generalizations about defining cities difficult. We will discuss the implications of these definitional challenges for the empirical study of urban health. Meanwhile, we can consider different types of cities using an example with which all readers are likely familiar.

Cities can be sprawling, diffuse, and automobile-dependent metropolitan areas. This has led to recent substantial academic discourse about urban sprawl.[15] For example, in Atlanta, Georgia, the average person travels by car more than 34 miles each day, which is more than twice as many miles as people in Philadelphia, Pennsylvania, drive.[16] Conversely, cities can be small and compact, as are many old European cities like Venice. Cities can be unique, cosmopolitan places (Paris, Casablanca), but also can look tremendously alike, as do any number of midsize North American cities (Kansas City, Denver). Cities frequently include both sophisticated and wealthy areas, featuring commercial and entertainment interests that are among the best in their country, as well as areas of extreme poverty and deprivation. For example, Rio de Janeiro has among the world's most expensive tourist resorts abutting extremely poor *favelas*; in New York City, the Upper East Side and Harlem are adjacent neighborhoods that are among the richest and poorest neighborhoods respectively in the United States. Cities generally are the centers of commerce and culture in their countries and geographic regions. Proximity to other cities frequently defines the range of opportunities available in a particular city. For example, a regional capital in a large, sparsely populated area, such as Whitehorse in Canada's Yukon Territory, is likely to have more diverse cultural offerings and a greater range of health services available than a comparable-size city close to other, far larger, urban areas. Therefore, cities can represent diverse conditions within which people live and can represent a range of human experiences. Throughout the rest of this review, we discuss how these diverse places may affect health and how the systematic study of urban health may afford opportunities to improve population health.

MECHANISMS OF DISEASE: WHY CITIES MAY SHAPE POPULATION HEALTH

How does the urban context affect health? In particular, what are the mechanisms by which cities can affect health? Before answering this question, a couple of considerations are in order. First, there is no one way in which the urban context may affect health. Although, for the sake of explication, we discuss mechanisms and health in general, different mechanisms frequently provide important potential explanations for the relations between the urban context and different diseases. As we discuss potential mechanisms, we consider health as one construct but make reference to specific theoretical distinctions and empirical examples that suggest how various factors may be important in different ways for diverse conditions. Second, as we highlight in the preceding section, cities are geographic places. Although cities are not static, and in fact cities' dynamism is one of their defining features, considering health in cities is fundamentally the study of how a particular type of place may affect health.

Explanations for these potential effects then rest primarily on how characteristics of places, in this case cities, may be important health determinants, with each characteristic having multiple implications for urban dwellers. Academic interest in urban health has waxed and waned over the past century; several authors at different points have proposed frameworks for considering the relation between city living and health, and they have identified features of the urban context that may be

particularly important for specific diseases.[17–20] Many of these frameworks build on work that discusses the social and economic determinants of better population health.[21–23] We find it useful to think of three broad categories of theories and mechanisms that may explain how city living can affect health: the physical environment, the social environment, and the availability of and access to health and social services.

The Urban Physical Environment

The urban physical environment includes the built environment: the air city dwellers breathe, the water they drink, the indoor and outdoor noise they hear, the park and areas inside and surrounding the city, and the geological and climate conditions of the site where the city is located. McNeill has suggested that primarily what distinguished the 20th century from previous ones, and cities from nonurban areas, is the degree to which humans have become the primary influence on the physical environment.[24] Although the literature on the relation between features of the physical environment and health is vast, we consider here some of the primary evidence linking key features of the physical environment to health.

The Built Environment

The human built environment can influence both physical and mental health; empirical evidence about the relation between the built environment and health conditions includes, among others, asthma and other respiratory conditions, injuries, psychological distress, and child development.[25–27] As an example, Weich and colleagues in 2002 [28] demonstrated higher levels of resident depression in areas that had less desirable built environments. In a study of New Orleans neighborhoods, Cohen [29] found that the prevalence of gonorrhea infection was higher in neighborhoods with deteriorating built environments. Different aspects of the built environment have been linked to specific health outcomes. For example, specific features of the built environment, including density of development, mixed land uses, scale of streets, aesthetic qualities of place, and connectivity of street networks, may affect physical activity.[30] In turn, low levels of physical activity are a well-established risk factor for cardiovascular disease and all-cause mortality in urban areas.[31, 32] A substantial literature addresses the relation between housing and health.[33, 34] Recent work has begun to differentiate the roles of the external and the internal built environment in shaping health.[35] Urban design may also affect health behaviors, crime, and violence rates,[36–38] suggesting close interactions among urban physical and social environments.

Urban Infrastructure, Water, and Sanitation

The urban infrastructure is a critical part of the physical environment and determines how a city provides water, disposes of garbage, and provides energy.[39] Water scarcity and water pollution are serious urban problems, particularly in less-wealthy countries. Nearly 1.5 billion people lack safe drinking water, and at least 5 million deaths per year can be attributed to waterborne diseases.[40] The relation between the urban infrastructure and health is shaped by different forces in established urban areas and in rapidly growing urban areas. In longstanding urban areas,

the decline of an aging infrastructure, coupled with frequently declining municipal resources, may challenge a city's ability to continue to provide safe water and sanitation for urban residents. Breakdowns may increase, causing health problems related to water, sewage, or disposal of solid waste.[41] In rapidly urbanizing areas, frequently in less wealthy countries, cities often are challenged to maintain an adequate fresh water supply to growing numbers of urban residents and to transport accumulating sewage and other waste. The World Health Organization (WHO) estimates that most urban populations in developing countries do not have access to proper sanitation.[42] Inadequate provision for solid waste collection frequently results in contamination of water bodies, which, coupled with the population density inherent to cities, presents a substantial risk for spreading epidemics rapidly.[43–44]

Pollution

In the first half of the 20th century, air pollution in the United States increased steadily as industrialization progressed, industries and homes used coal for power and heat, and cars proliferated. Cities had worse pollution than did nonurban areas. In the second half of the century, however, and especially in the past 25 years, many forms of pollution decreased as coal was phased out, manufacturing plants moved to the suburbs or abroad, lead was banned from gasoline, and the automobile industry was forced to build cleaner cars. However, cities still generate close to 80 percent of global carbon dioxide emissions and account for three-quarters of industrial wood use worldwide.[45] As late as the mid-1990s, investigators estimate that air pollution contributed to 30,000–60,000 deaths per year in the United States.[46] Indoor and outdoor air pollution are thought to contribute to 3 million deaths globally a year, with 90 percent of these deaths being in less wealthy countries. Worldwide, atmospheric pollution is thought to affect more than a billion people, mostly in cities.[47]

Access to Green Space

Some of the earliest studies that considered the relation between the urban context and health emphasized the role of access to parks and green space, or lack thereof, in shaping the health of urban populations. Griscom's report about housing in New York City in 1845 suggested that a lifestyle filled with "animal and vegetable exhalations" in the countryside provided "prima facie proofs" of the superiority of living in the countryside.[48] Although it remains generally recognized that public green spaces make for a more pleasant living environment, the empirical literature evaluating the relation between green space and health remains limited. Recent work has shown that living in areas with walkable green spaces, as opposed to living in areas without walkable green spaces, was associated with greater likelihood of physical activity,[49] lower cardiovascular disease risk,[50] and longevity among the elderly, independent of personal characteristics.[51–52] As more multidisciplinary work in urban health develops, more experimental and observational studies likely will assess the role of green space and urban planning in promoting health.

Urban Climate

Highways and streets can pollute water through runoff, destroy green space, influence motor vehicle use and accident rates, and contribute to the urban heat

sink—that is, absorption of heat that can increase by several degrees the temperature in cities. On warm days, urban areas can be more than 5 degrees Fahrenheit warmer than surrounding areas, an effect known as the urban health island effect.[15] This effect is primarily due to dark surfaces absorbing health and the limited ability of urban areas (with relatively few trees) to cool the air through transpiration. Global climate change may exacerbate this effect. Heat is a concern in urban areas in several ways, and ambient air temperature has been associated with a large number of hospitalizations and deaths yearly.[53] Heat exposure may result in direct health effects, including syncope or heat exhaustion, or exacerbate existing health disorders. Excess heat in urban areas also can exacerbate pollution, as cooling equipment (air conditioners) is put into heavier use to compensate for rising urban temperatures. Particular groups may be most at risk of the effects of heat in urban areas. Epidemic heat-related deaths have been particularly pronounced among socioeconomically disadvantaged and socially isolated elderly persons.[54]

Other Features of the Urban Physical Environment

Several other aspects of the urban physical environment have specific relations to human health, and a full review of all relevant features of the physical environment is beyond the scope of this chapter. However, city structures like bridges and skyscrapers may be vulnerable to natural or manmade disasters, as recent earthquakes in Japan and Iran and the September 11, 2001, terrorist attacks on New York City demonstrate, respectively. Features of the urban and social environment, such as population density and social contagion, coupled with these vulnerable urban structures, can result in substantial health consequences after disasters in urban areas.[55, 56] Other threats to health in cities include hazardous waste landfills, often located in or near urban areas, which may be associated with risks or low birth weight, birth defects, and cancers.[57] Noise exposure, a common urban problem, may contribute to hearing impairment, hypertension, and ischemic heart disease.[58]

The Urban Social Environment

The social environment has been broadly defined to include "occupational structure, labor markets, social and economic processes, wealth, social, human, and health services, power relations, government, race relations, social inequality, cultural practices, the arts, religious institutions and practices, and beliefs about place and community."[59] This definition, by its very complexity, suggests that there are multiple ways in which the urban social environment may affect health. Social disorganization, social resources, social contagion, spatial segregation, and inequality may be particularly important determinants of health in cities. Although these concepts have, in large part, arisen from sociological theory, many of them have been increasingly integrated into public health thinking that explores the relation between contextual characteristics and health.

Health and Social Services

The relation between provision of health and social services and urban living is complicated and varies between cities and countries. In wealthy countries, cities are characterized by a rich array of health and social services. Even the poorest urban

neighborhood often has dozens of social agencies, each having a distinct mission and providing different services. Many of the health successes in urban areas in the past two decades, including reductions in HIV transmission, teen pregnancy rates, tuberculosis control, and new cases of childhood lead poisoning, have depended in part on the efforts of these groups. In addition, many urban areas serve as referral centers for surrounding communities and, as such, urban areas often have a greater availability of health and social services. In general, nonurban areas have far fewer physicians and hospitals, and the travel time to health care providers is greater than in nonurban areas.[60]

Multiple mechanisms may explain how cities affect health, with different mechanisms being potentially important for different morbidities. Indeed, a big-picture perspective on the relation between the urban context and health would suggest that these relations are undoubtedly complicated and that any single analysis that isolates a feature of urban living and health is just scratching the surface. Whereas specific features of cities may affect specific diseases adversely, other features may offer protection. Interrelationships between features of the urban environment further make generalization difficult. For example, further refinements on social strain theory in urban areas include an appreciation of the fact that, in urban areas, people of different socioeconomic statuses face different stressors and have varying levels of access to resources that may help them cope with these stressors. In particular, in urban areas, formal local resources can complement or substitute for individual or family resources for transient urban populations. Therefore, the relation between urban stressors and health is likely buffered by salutary resources (for example, health care, social services) that often are more prevalent in urban compared with nonurban areas.[61] Although these resources may be available to urban residents, socioeconomic disparities in cities are linked to differential access to these resources, which suggests that persons at different ends of the socioeconomic spectrum may have different opportunities to benefit from the resources available in cities.

International Considerations

In considering the mechanisms that may explain the relations between the urban context and health, we refer to potential differences in the role of certain mechanisms cross-nationally. This point is worth emphasizing, particularly in light of the varying pace of urbanization worldwide. The pace of urbanization is projected to differ by region of the world and by initial city size. In particular, most global population growth in the coming decades will occur in less wealthy regions of the world, with the most rapid pace of growth expected to occur in Asia and Africa. Although North America and Europe are currently the most urbanized regions, the number of urban dwellers in the least urbanized region, Asia, in 2000 was already greater than the urban population in North America and Europe combined. The proportion of people living in megacities is expected to rise from 4.3 percent of the global population in 2000 to 5.2 percent in 2015.[62] The growth rate of megacities in the developing world will be much higher. For example, the anticipated growth rate for Calcutta, India, between 2000 and 2015 is 1.9 percent compared with an anticipated growth rate of 0.4 percent for New York City, United States.[21] Whereas the growth of large cities in developing countries will account for approximately

one-fifth of the increase in the world's population, small cities will account for almost half of this increase.[5] A growing number of relatively small cities throughout the world will contain most of the world's population in the 21st century, and most of the growth in cities will take place in less wealthy countries.

Thus, the relative importance of characteristics of the urban environment that may affect health may vary substantially in different cities and in different parts of the world. For example, in many rapidly growing urban areas in the developing world, lack of safe water and poor sanitation are likely to account for a greater proportion of the morbidity and mortality in a specific city than all other factors identified here. As cities become more established, an aging infrastructure can threaten health and growing inequalities, and social strains can influence both health behaviors and access to resources. In addition, the course of urbanization in different cities worldwide may have different implications for health. A newly urbanizing city is likely to be under different and probably more substantial strains than a long-established urban area. Therefore, when considering how cities may affect health, it is important for the public health researcher or practitioner to consider both place (that is, the particulars of a given city) and time (that is, the trajectory of urbanization in a particular city). No simple solutions summarize the relations between the different factors that can affect health in various countries. Rather, specific investigations and interventions would do well to bear in mind the relevant local and temporal context that may guide an appreciation of relevant and salient risk determination in a given urban area.

COMPLEXITY OF CAUSATION IN THE URBAN CONTEXT

As discussed at various points in this chapter, cities are complex communities of heterogeneous individuals, and multiple factors may be important determinants of population health in cities. For example, understanding the role that racial and ethnic heterogeneity plays in shaping the health of urban populations requires an understanding of the role of segregation in restricting access to resources in urban neighborhoods as well as the potential for greater tolerance of racial and ethnic differences in cities compared with nonurban areas. Assessing how the urban context may affect health raises challenges and introduces complexity that often is not addressed easily through the application of simple analytic methods.

In addition, cities are different from one another and may change over time. Empirical inquiry in health presupposes that identifiable factors influence health, and these factors can be identified (and potentially intervened upon). Typically, public health studies imply, for example, that we can generalize about how different foods will affect health across individuals, at least within the confines of effect modification across groups (for example, age-groups) or under different circumstances (for example, at different levels of caloric intake). However, cities are characterized by multiple factors (for example, population density, heterogeneity) that in many ways make each city unique. The complexity of cities and of city living may mean that urban characteristics important in one city may not be important in other cities, limiting the generalizations that can be drawn about how urban living influences health. Further complicating this task is the fact that cities change over time, and this change has implications for the relative contribution of different factors in

determining health in cities. For example, municipal taxation of alcohol and cigarettes may control alcohol and cigarette consumption in a particular city at one point in time.[63] However, changing social norms around smoking and alcohol use may either obviate or reinforce the influence of taxation. As such, in considering urban characteristics that affect health, it may be important to note both the prevailing context within which such characteristics operate and that the role of these characteristics may change over time.

DIRECTIONS FOR URBAN HEALTH RESEARCH AND PRACTICE

Key factors can explicitly distinguish and guide the study and practice of urban health. First, we need to consider whether specific features of the urban context are causally related to health. Appropriate specification of the research question of interest is critical. For example, understanding how living in a city as a whole may affect smoking behavior requires a different set of tools than do questions about how intraurban differences in pollution affect variability in the neighborhood prevalence of asthma. Similarly, understanding the quantitative relation between social capital in urban communities and resident well-being requires different tools than do questions about why social capital may have different implications for health in different communities or how social capital is produced or eroded in urban contexts.

Second, it is important to consider whether these features are differentially distributed between urban and nonurban areas and within urban areas (for example, between urban neighborhoods). As a corollary to this consideration, it becomes essential to consider the extent to which these features are unique to a particular city or differ between cities and, as such, to learn whether salutary features of the urban environment are adaptable in different contexts. For example, undoubtedly, much can be learned from well-studied urban areas in wealthy countries that can be applied to public health practice in less wealthy countries.

Third, identifying which characteristics of the urban context, and under which circumstances are modifiable, is an important theoretical, empirical, public health question. In many ways, the choice of an appropriate urban health framework may dictate, at least implicitly, the choice of both the question asked and the methods used in addressing the question. For example, a comprehensive framework that includes national-level policies that shape municipal financing may suggest that inquiry into and intervention on national policies may be of primary importance to urban health. In contrast, a framework that considers primarily physical characteristics of cities will address how features of the built environment at the local level can affect residents' health. Thus far, relatively little has been written about the processes through which the urban context may affect health and about further elucidation of these processes. A comprehensive appreciation of the processes that influence urban health can and should guide research and practice.

In conclusion, although this review highlighted challenges inherent to the study of urban health, this work is informed by an appreciation for the potential of urban health inquiry. Although the study of urban health embeds substantial complexity, clearly specified research questions and appropriate study designs can help focus our appreciation of the relation between specific features of the urban context and health,

both in specific cities and as generalizable to cities in national and international contexts. Recent methodologic advances, particularly the widespread acceptance of multilevel methods in public health research, have made it possible to test hypotheses about urban characteristics and their relation to specific health outcomes. Newer methods eventually may contribute to an improved understanding of the competing influences on the health of urban populations over time.[63] Such research can inform local intervention and policies across urban areas. We hope that efforts such as this review, aimed to structure our thinking about cities and health, are helpful in stimulating both empirical and theoretical developments that can lead to improved health in cities worldwide.

REFERENCES

1. McMichael AJ. Urbanization and urbanism in industrialized nations, 1850–present: implications for health. In: Schell LM, ed. *Urbanism, Health, and Human Biology in Unindustrialized Countries*. Cambridge, UK: Cambridge University Press; 1999:22–26.
2. Brockerhoff MP. An urbanizing world. *Popul Bull*. 2000;55(3):3–4.
3. Guidotti TL, et al. *Basic Environmental Health*. New York: Oxford University Press; 2001.
4. Satterthwaite D. Coping with rapid urban growth. http://www.rics.org/downloads/research_reports/urban_growth.pdf. RICS Leading Edge Series.
5. Satterthwaite D. Will most people live in cities? *BMJ*. 2000;321:1143–1145.
6. Dickens C. *The Personal History and Experience of David Copperfield the Younger*. New York: Collier; 1850.
7. Lund VK. The healthy communities movement: bridging the gap between urban planning and public health. http://www.asu.edu/caed/proceedings99/LUND/LUND.HTM.
8. Sheard S, Power H. *Body and City: Histories of Urban Public Health*. Aldershot, UK: Ashgate; 2000.
9. Durant W, Durant A. *The Story of Civilization*; vol. 10: *The Age of Rousseau*. New York: Simon and Schuster; 1967.
10. Durkheim E. *Suicide: 1897*. Glencoe, IL: Free Press; 1951.
11. Engels F. *The Condition of the Working Class in England*. New York: Lovell; 1887.
12. Marsella, AJ. Urbanization, mental health and psychological well-being: some historical perspectives and considerations. In: Harpham T, ed. *Urbanization and Mental Health in Developing Countries*. Aldershot, UK: Avebury; 1995:3–14.
13. Vogele, JP. Urban infant mortality in imperial Germany. *Soc Hist Med*. 1994;7(3):401–425.
14. Population Reference Bureau. *The Population Bulletin.* March 2004;59(1). http://www.prb.org/Template.cfm?
15. Frumkin H. Urban sprawl and public health. *Pub Health Reports*. 2002;117:201–217.
16. Texas Transportation Institute. 2002 annual urban mobility study. http://tti.tamu.edu/researcher/v38n2/annual percent5Fmobility percent5Fstudy.stm.
17. Galea S, Freudenberg N, Vlahov D. Cities and population health. *Soc. Sci. Med*. 2005;32(3):298–319.
18. Lawrence RJ. Urban health: an ecological perspective. *Rev Environ Health*. 1999;14(1):1–10.
19. Northridge ME, Sclar E. A joint urban planning and public health framework: contributions to health impact assessment. *Am J Pub Health*. 2003;93(1):118–121.

20. Vlahov D, Galea S. Urbanization, urbanicity, and health. *J Urban Health*. 2002; 79(Suppl 1):S1–S12.
21. Evans RG, Stoddart GL. Producing health, consuming health care. *Soc Sci Med*. 1990;31:1347–1363.
22. Hamilton N, Bhatti T. *Population Health Promotion: An Integrated Model of Population Health and Health Promotion*. Ottawa, Canada: Health Promotion Development Division of Health Canada; 1996.
23. Kaplan GA. What is the role of the social environment in understanding inequalities in health? *Ann N Y Acad Sci*. 1999;896:116–119.
24. McNeill JR. *Something New Under the Sun: An Environmental History of the Twentieth Century*. New York: Norton; 2000.
25. Evans RG, et al. Housing quality and mental health. *J Consult Clin Psychol*. 2000; 68(3):526–530.
26. Lynch J, et al. Income inequality, the psychosocial environment and health: comparisons of wealthy nations. *Lancet*. 2001;358:1285–1287.
27. Northridge, ME, Sclar E, Biswas P. Sorting out the connections between the built environment and health: a conceptual framework for navigating pathways and planning healthy cities. *J Urban Health*. 2003;80(4):556–568.
28. Weich S, et al. Mental health and the built environment: cross-sectional survey of individual and contextual risk factors for depression. *Br J Psychiatry*. 2002;180:428–433.
29. Cohen DA, et al. Broken windows and the risk of gonorrhea. *Am J Pub Health*. 2000; 90(2):230–236.
30. Handy SL, et al. How the built environment affects physical activity: views from urban planning. *Am J Prev Med*. 2002;23(Suppl 2):64–73.
31. Diez-Roux AV. Residential environment and cardiovascular risk. *J Urban Health*. 2003;80(4):569–589.
32. Pate RR, et al. Physical activity and public health: a recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *JAMA*. 1995;273:402–407.
33. Kingsley GT. Housing, health and the neighborhood context. *Am J Prev Med*. 2003; 24(3 Suppl):6–7.
34. Thompson H, Pettricrew M, Douglas M. Health impact assessment of housing improvement: incorporating research evidence. *J Epidemiol Community Health*. 2003;57:11–16.
35. Hembree C, et al. The built environment and overdose mortality in New York City neighborhoods. *Health Place*. 2005;11(2):147–156.
36. Berrigan D, Troiano RP. The association between urban form and physical activity in US adults. *Am J Prev Med*. 2002; 23(2 Suppl):74–79.
37. Newman O. *Defensible Space: Crime Prevention through Urban Design*. New York: Macmillan; 1986.
38. Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science*. 1997;277:918–924.
39. Melosi M. *The Sanitary City: Urban Infrastructure in America from Colonial Times to the Present*. Baltimore, MD: Johns Hopkins Press. 2000.
40. Krants D, Kifferstein B. Water pollution and society. http://www.umich.edu/gs265/society/waterpollution.htm.
41. Garrett L. *Betrayal of Trust the Collapse of Global Public Health*. New York: Oxford University Press; 2001.
42. World Health Organization. *Health and the Environment in Sustainable Development: Five Years after the Earth Summit*. Geneva, Switzerland: World Health Organization; 1997.

43. Alexander SE, Ehrlich PR. Population and the environment. In: Ernst WG, ed. *Earth Systems: Processes and Issues*. Cambridge, UK: Cambridge University Press; 2000.
44. Chanthikul S, et al. Computer simulation of leachate quality by recirculation in a sanitary landfill bioreactor. *Environ Sci Health Part A Tox Hazard Subst Environ Eng*. 2004;39(2):493–505.
45. O'Meara M. *Reinventing Cities for People and the Planet*. Washington, DC: Worldwatch Institute; 1999.
46. Samet JM, et al. Fine particulate air pollution and mortality in 20 US cities, 1987–1994. *New Engl J Med*. 2000;343:1742–1749.
47. Environmental Software and Services. Environmental impact assessment: GmbH AUSTRIA 2002. http://www.ess.co.at/AIR-EIA/LECTURES/L001.html.
48. Griscom J. *Sanitary Conditions of the Laboring Population of New York*. New York: Harper; 1845.
49. Booth ML, et al. Social-cognitive and perceived environment influences associated with physical activity in older Australians. *Prev Med*. 2000;31:15–22.
50. Lee IM, et al. Physical activity and coronary heart disease in women: is "no pain, no gain" passé? *JAMA*. 2001;285:1447–1454.
51. Takano T, Nakamura K, Watanabe M. Urban residential environments and senior citizens' longevity in megacity areas: the importance of walkable greenspaces. *J Epidemiol Community Health*. 2002;56:913–918.
52. Tanaka A, et al. Health levels influenced by urban residential conditions in a megacity—Tokyo. *Urban Stud*. 1996;33(6):879–894.
53. Basu R, Samet JM. Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. *Epidemiol Rev*. 2002;24:190–202.
54. Semenza JC, et al. Excess hospital admissions during the July 1995 heat wave in Chicago. *Am J Prev Med*. 1999;16(4):269–277.
55. Galea S, et al. Psychological sequelae of the September 11th attacks in Manhattan, New York City. *New Engl J Med*. 2002;346:982–987.
56. Schlenger WE, et al. Psychological reactions to terrorist attacks: findings from the National Study of Americans' Reactions to September 11. *JAMA*. 2002;288(5):581–588.
57. Vrijheid M. Health effects of residence near hazardous waste landfill sites: a review of the epidemiologic literature. *Environ Health Perspect*. 2000;108(Suppl 1):S101–S112.
58. Passchier-Verneer W, Passchier WF. Noise exposure and public health. *Environ Health Perspect*. 2000;108(Suppl 1):S123–S131.
59. Barnett E, Casper M. A definition of "social environment." *Am J Pub Health*. 2001;91(3):465.
60. Ormond BA, Zuckerman S, Lhila A. *Rural/Urban Differences in Health Care Are Not Uniform across States*. Washington, DC: Urban Institute; 2000.
61. Galea S, et al. Collaboration among community members, local health service providers, and researchers in an urban research center in Harlem, New York City. *Public Health Rep*. 202;116(6):530–539.
62. United Nations. World urbanization prospects: the 1999 revision. http://www.un.org/esa/population/pubarchive/urbanization/urbanization.pdf.
63. Levins R, Lopez C. Toward an ecosocial view of health. *Int J Health Serv*. 1999;29:261–293.

CHAPTER 4

Consequences of Climate Change on Human Health

Paula Brown, PhD, and Arthur DeGaetano, PhD

INTRODUCTION

The knowledge that the environment in which we live affects human health is not a recent concept, with the effect of climate on health being documented in society since the time of Hippocrates (400 BCE). Accounts of climatic change affecting food supplies, human health, and political systems are prevalent throughout history. The medieval famine, Black Death, and subsequent decay of feudalism in Europe during the 1300s are prominent historical examples of climatic influence. Concerns about recent changes in global climates and possible future trends on the health of the world's population are now emerging areas of important research. Both climatological and medical communities have increased their interest in climate-related human health issues since the early 1990s as anxiety about climatic influence on global populations grows.

CLIMATE CHANGE

The Climate System

The climate system is based on the interaction of atmosphere, hydrosphere, cryosphere, land surface, and biosphere. These five components form the basis of an energy budget that redistributes solar radiation globally, from which local climates are derived. The atmosphere is the most dynamic component of the climate system that circulates energy. Solar radiation passes through the atmosphere and is reflected or absorbed by the atmosphere, clouds, and the surface to perpetuate the radiation balance. Atmospheric composition is dominated by nitrogen and oxygen, but it is the existence of natural greenhouse gases (such as carbon dioxide, water vapor, nitrous oxide, ozone, and methane) that absorb and reradiate infrared or long-wave radiation that is critical to the climate system. These gases effectively trap heat and warm surface temperatures. Stratospheric ozone is not only a greenhouse gas when

it occurs near the earth's surface, but also forms a layer high in the stratosphere that absorbs ultraviolet radiation preventing these harmful rays from reaching the surface.

The hydrosphere (liquid water sources) provides a similar role in energy redistribution to the atmosphere; however, the hydrosphere generally operates more slowly. As 70 percent of the earth's surface is covered by water, the role of the hydrosphere in storing and transporting solar energy is critical for a stable climate. In particular, the large capacity of oceans to store both heat and carbon dioxide dampens the inherent variability of the climate. Oceanic energy redistribution by currents and circulation patterns from the equator to the poles is driven by winds and both thermal and salinity changes that drive the thermohaline circulation. The interaction between the hydrosphere and atmosphere also redistributes energy and water globally through the interaction of global circulation patterns and the hydrological cycle.

Water affects the climate system through the snow and ice present in the cryosphere. High reflectivity (albedo) of frozen surfaces causes little direct warming or absorption of energy from the sun in regions where snow and ice are present. The crysophere has a large role in regulating ocean temperatures, salinity, and sea level, as well as deep ocean circulation. The cryosphere also affects climate where seasonal snow and ice is present, especially over landmasses.

Land surfaces directly affect the energy balance due to albedo changes and also influence the hydrological cycle. Air moving over the land surface is affected by elevation, roughness, and moisture characteristics interacting with the atmosphere to change air mass characteristics, especially in large continental areas. Similarly, the biosphere interacts with the atmosphere and climate system through the radiative balance, the carbon cycle and photosynthesis. Marine and terrestrial biomes store large amounts of carbon and play an integral role in regulating atmospheric gas concentration, especially carbon dioxide.

The climate system is a complex balance of the incoming, outgoing, and stored energy sources, which vary temporally and spatially. The balance that is created between the five components of the climate system is critical to the environment in which we live. Global circulation patterns are primarily driven by the interaction between the atmosphere and surface, which transfer heat and create the air masses that characterize regional climates. For example, air masses that form over tropical oceans are warm, moist, and unstable, while those forming over ice-covered polar regions are cold, dry, and stable. Air masses are then modified by the surfaces they encounter as they move, with local surface moisture characteristics and especially orography affecting the nature of the air mass.

Climate Variability

Climatic variability is produced by interacting internal and external forcings. Internal forcings include El Niño Southern Oscillation (ENSO) variability, which causes regional circulation changes, while an example of an external forcing is solar variability that is caused by orbital cycles and is implicated in the occurrence of ice ages. These forcings create a complex and varying climate over a wide range of timescales, from interannual to millennial. On longer timescales, paleoreconstructions of past global temperatures derived from proxy records show the long-term

variability inherent in the climate system. These proxy records are based on the evidence obtained from climate indicators, such as ice cores, tree rings, boreholes, and sediments. Analyses of these paleorecords show the influence of changes in the earth's orbit, solar output, landmass, and greenhouse gases on the variability of long-term global temperatures and their contribution to glacial cycles. The resolution of paleoreconstructions is coarse, however, compared with more recent direct climatological observations.

The most prominent example of internal variability in the climate system is ENSO. Coupling between circulation patterns of the ocean and atmosphere in the tropical Pacific drive ENSO and create interannual climate variations. Two well-known manifestations of ENSO are the dichotomous El Niño and La Niña patterns. El Niño conditions occur every three to seven years, while La Niña conditions are less frequent. Although the influence of El Niño and La Niña is strongest in Pacific regions, the resultant changes in temperature and precipitation patterns have global effects on people, water resources, and food supplies. Understanding of ENSO processes now enables predictions of upcoming ENSO events, to provide forecasts of upcoming regional weather patterns and climate.

Another oscillation pattern that affects climatic variability in North America, Europe, and Asia is the North Atlantic Oscillation (NAO). Changes in atmospheric pressure anomalies in the North Atlantic alter wind patterns and storm tracks, especially in Europe. Other interannual climate oscillations, such as the Northern Annular Mode (NAM), Southern Annular Mode (SAM), and the Pacific North American (PNA) pattern, affect temperatures and precipitation regionally. Over longer timescales, an ENSO-like multidecadal oscillation in the North Pacific, called the Pacific Decadal Oscillation (PDO), modulates ENSO and affects climate variability. A similar Pacific-wide phenomenon is the Interdecadal Pacific Oscillation (IPO).

Volcanoes are an external source of climate variability and cause negative forcing or cooling. Volcanic activity affects the climate for several years and is dependant on many factors such as the size, intensity, location, and time of year of the eruption. Volcanic events, such as the Mount Pinatubo eruption in the Philippines, eject aerosols into the atmosphere that cool temperatures by reducing insolation, however these effects decrease over time.

The climate responds to changes in forcings as a complex system that operates on a number of different temporal and spatial scales. As a result of variability, these interactions can create feedbacks that influence the climate system, for example, the effect of clouds on the radiative balance. These feedbacks often do not involve a simple linear or proportional cause and effect relationship. Evidence of complex responses to variability is apparent in many examples, such as the thermohaline circulation and ENSO. Changes in climate variability also have important implications for extreme events, which can exhibit an increase or decrease in frequency and have large societal impacts.

Recent Trends

Global temperatures warmed by an average of 0.74°C (0.56°C–0.92°C) during the 100-year period from 1906 to 2005.[1] Thirteen of the 14 warmest years recorded (since 1880) occurred in the 1995–2008 period.[2] The rate of temperature change

varies, and since 1900 global temperatures show a 40-year warming trend, a 30-year period of little change, then another period of warming over the last 30 years. Spatial variability in temperatures also is evident with greater warming in the northern high latitudes, and land areas displaying larger warming rates than the oceans. Over shorter time periods, however, some regions show cooling, which is attributable to circulation changes.

The warming noted in average temperatures also has affected the incidence of extreme temperature events. Frost days and cold temperature extremes have decreased, while warm nights and hot days have increased. Warming has led to longer frost-free seasons and longer growing seasons in the mid to high latitudes. Northern Hemisphere temperatures have warmed most during winter and spring. Sea surface temperatures are also warming, which contributes to thermal expansion of the oceans and thus increasing sea level, along with increased melt water from the cryosphere.

While warmer temperatures are the predominant characteristic of recent climate change, precipitation changes show a less coherent response. Precipitation trends are closely linked to atmospheric circulation patterns and show significant increases in the mid and high latitudes of the Northern Hemisphere. Decreasing precipitation trends in the tropics after 1970 and the widespread increase in drought severity are further indications of precipitation change toward less frequent, but more intense, heavy precipitation events. Evidence also shows that intense tropical cyclones have become more frequent in the North Atlantic.

Anthropogenic Influence and Future Predictions

Over the last few decades, debate over the effect of anthropogenic influence on climate has centered on the amount of carbon dioxide in the atmosphere. Based on the most recent assessment of climate science by the Intergovernmental Panel on Climate Change (IPCC), "most of the observed increase in global average temperatures since the mid-20th century is very likely due to the observed increase in anthropogenic greenhouse gas concentrations. It is likely that there has been significant warming over the past 50 years, averaged over each continent (except Antarctica)."

Although the climate has many elements of natural variability, some of the important elements of the climate system also have been affected by anthropogenic influence. The large amounts of greenhouse gases produced globally, primarily through burning fossil fuels and agriculture, have enhanced the greenhouse effect by increasing absorption and reemission in the atmosphere. The ability of the climate system to adjust to changes in greenhouse gas levels is complex. Increases in temperature promote increases in water vapor, which itself is a greenhouse gas, resulting in a key positive feedback. Simultaneously, changes in land use affect local and regional climates as a result of changes in radiative, physical, and biological properties. Hydrological and carbon cycles are two of the most important terrestrial systems that are affected by land use change and can perturb climates. Aerosols (such as dust, smoke, and pollution) also have a regional effect on climate, as they are short-lived in the atmosphere. The negative radiative forcing created by many aerosols is a result of these particles, especially sulfates, absorbing and reflecting radiation, and contributing to increased cloudiness.

Future predictions of climate change are based on complex general circulation models (GCMs) that are designed to simulate the climate system. The most recent IPCC findings are based on projections from a number of modeling groups that estimate future climate change based on a range of scenarios.[1] These scenarios are based on different levels of greenhouse gas and aerosol emissions that may occur for the period 2000–2099 (relative to 1980–1999). Inherent uncertainties remain, however, regarding the model predictions due to the complexity of the climate system. Predictions show temperatures are expected to increase 1.8–4.0°C globally, with northern latitudes and large continental areas showing higher-than-average warming. Hot temperature extremes are very likely to increase along with a subsequent decrease in cold temperature extremes. Future precipitation patterns are much harder to predict as local and regional patterns dominate, but projections indicate that heavy precipitation events are likely to increase and exacerbate current regional precipitation patterns with dry regions becoming drier and wet areas becoming wetter. Projections also indicate extratropical storm tracks will shift poleward, which affects wind, precipitation, and temperature patterns, and also that an increase in tropical cyclone intensity is likely. Future changes in sea level are difficult to model as climate-carbon cycle feedbacks and ice flow dynamics are not well understood. However, the conservative estimates range from 0.18–0.59 meters.

THE EFFECT OF CLIMATE CHANGE ON HUMAN HEALTH

The effects of climate change on human health provide insight into current and future implications for populations. Much research into disease patterns is based on epidemiological studies, which are reliant on data availability, of which the length and quality of the data frequently are lacking.[3] Studies of climate change on infectious diseases are biased toward more developed countries that have the resources to provide data and funding for research.[4] Incidence of disease can be complicated by other explanations to determine changes, such as migration, drug resistance, population health, and environmental change.[3] Climatic changes have direct health effects from altered weather patterns, but the indirect effects on agriculture and wider population systems are important factors for the global disease burden.[5] Figure 4.1 shows many of the factors contributing to the effect of climate change on health. A warming climate produces both positive and negative implications for human health that are location specific. Sensitivity and vulnerability to the health effects of climate change is largely associated with socioeconomic conditions.

Climate change has a large effect on regional health patterns because of the susceptibility of populations as a result of interacting socioeconomic and climatic influences. The 16 national health impact assessments of climate change produced from 2001–2006, cited in the IPCC's chapter on health, reflect the increased attention that is now being paid to the earth's vulnerability to climate change. These assessments, however, have been undertaken by more developed countries where vulnerability to climate change generally is limited. Changes in temperature, precipitation, and extreme events associated with recent warming do not exhibit a consistent response to global human health. The year-to-year variability and resultant shifts in large-scale atmospheric circulation patterns caused by ENSO have been shown to affect human health.[6, 7]

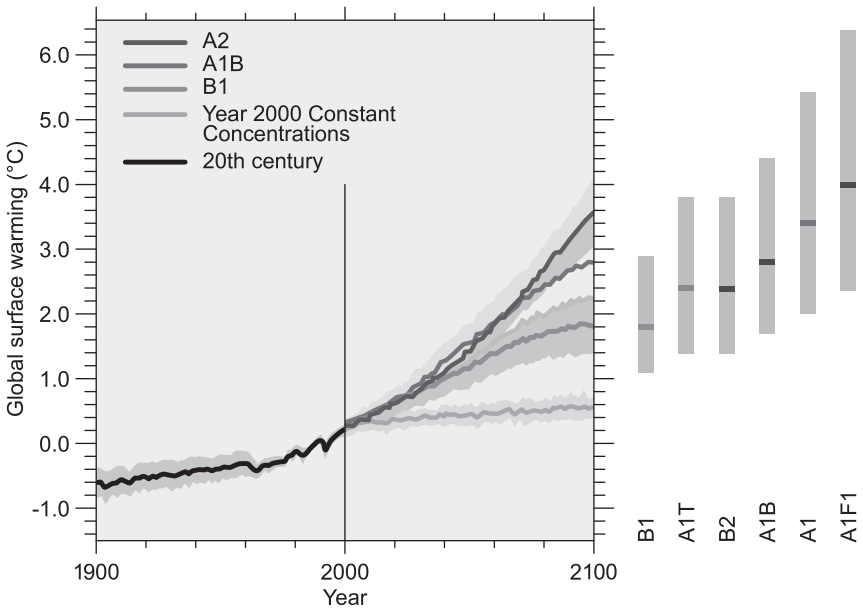

Figure 4.1 Multi-model Global Averages of Surface Warming

Many studies of infectious disease have investigated the effects of ENSO-related weather on health. ENSO conditions have a strong impact on the health of some populations associated with the shifts in large-scale atmospheric circulation that perturb temperature and rainfall patterns. El Niño events in particular are associated with increased risk of weather-related natural disasters, death, and disease.[8] Weather extremes caused by ENSO events such as flooding and droughts have large impacts on low and mid-latitudinal areas of the world.[9] The incidence of infectious diseases also increases, with malaria transmission being frequently studied due to its particular sensitivity to changes in temperature and rainfall.[8] Incidence of dengue fever, Rift Valley fever, cholera transmission, Ross River virus, and diarrheal disease are some other diseases that ENSO changes influence.[10]

Extreme Events

Human health is affected by the occurrence of extreme weather or climate events. Extreme temperature, rainfall, and wind-related events are experienced in different regions of the world every year and each event threatens human health. Socioeconomic conditions can influence the severity of extreme events on the population's health, but the effects of the most severe extreme events are indiscriminate. Recent increases in temperature have affected the frequency of extreme temperature and rainfall events.[11] The WHO identifies two different types of extreme climate events as simple extremes of climate statistical ranges or complex events, such as droughts or tornadoes. The occurrence of extreme events also is affected by circulation patterns that are influenced by ENSO conditions.

Thermal extremes can increase morbidity and mortality rates.[12] Mortality rates show a U-shaped distribution with temperature.[13] Heat waves occurring over

the last few decades include the Chicago (1995) and Europe (2003) heat waves, with 514 deaths [14] and tens of thousands of deaths [15] attributed to each event, respectively. Morbidity rates are high in heat events, with 3,300 excess emergency emissions in the Chicago heat wave.[16] Mortality and especially morbidity statistics are difficult to determine due to their nonspecific causes associated with heat extremes. Death rates due to heat waves are highest for the elderly.[12, 17] Temperature-related mortality in the United States varies by latitude with southern cities most susceptible to cold temperatures and northern cities susceptible to warmer temperatures.[13] Heat waves are more likely to occur as recent warming also brings warmer nighttime temperatures and higher humidity levels.[18] Changes in extent, breadth, intensity, and frequency of heat waves also are associated with climate change.[19] The high mortality rates associated with the 2003 European heat wave are likely to be linked to climate change and anthropogenic influence.[5] The urban environment and urban heat island effect exacerbate warm temperature events [20] and affect human exposure to pollutants.[21]

Cold and winter mortality rates are more difficult to relate to weather events than heat events.[16] Low temperatures and cold waves are a problem primarily in the Northern Hemisphere.[5] The mortality burden of cold temperatures is likely to be underestimated as its effect on respiratory and cardiovascular diseases may not be included in burden statistics.[12] Susceptibility to accidental exposure to the cold primarily occurs outdoors, with the socially deprived, outdoor workers, and elderly most at risk.[5, 22] The seasonal transmission pattern of influenza is also related to cold winter conditions.[23] Seasonal respiratory infections are the primary mechanism of cold-related mortality in temperate countries [16] where housing inadequacies play a large role.[24] Although there is some evidence of recent climate change causing a reduction of winter mortality, this trend is likely to reflect social, environmental, behavioral, and health care improvements.[25] The effect of climate change on winter deaths is limited by data deficiencies.[26]

Flooding and drought are the most frequent and deadly natural disasters.[26] Extreme and heavy precipitation events are often the result of storms as well as El Niño events. Drowning and severe injuries are the immediate effects of flooding.[16, 27] The risk of death is influenced by characteristics of the flood (speed, size, and suddenness) and population (infrastructure and economic conditions).[28] Storm surges from coastal storms have high mortality rates from drowning and can occur in densely populated low-lying areas that also have a high health burden.[5] Floods create a large amount of damage to dwellings, infrastructure, and water supply systems. Lower income countries have limited sanitation infrastructure and experience an increase in the severity of disease burden compared with high-income countries, except for extreme flooding events.[5] Populations are then exposed to infectious diseases and contaminated water supplies. Pollutants following major disasters are an increasing problem.

Disease increases related to fecal-oral disease, vector-borne disease (VBD), and rodent-borne disease are commonly associated with floods.[28] Instances of diarrheal disease (such as cholera) increase and people experience deterioration of chronic diseases, which exacerbate the health effects of flood disasters.[29] In 2007, flooding in China caused the deaths of 535 people and affected 105 million others.[29] Disaster statistics, however, often do not include ongoing health impacts

from unsafe and unhealthy conditions.[5] Mental health problems such as Post Traumatic Stress Disorder (PTSD), anxiety, and depression are caused by floods.[28] Floods also create health issues associated with gender differences.[5] Women experience increases in domestic violence and PTSD.[30, 31]

In contrast to the immediate impact of floods, droughts are a persistent climatological condition. ENSO cycles affect the incidence of drought in many regions. Recent climate change has resulted in the increased length and intensity of droughts since 1970.[11] Droughts primarily affect health indirectly through their effect on food production.[16] The decrease in viable agricultural land leads to nutritional problems, famine, and migration. Water scarcity also leads to the use of compromised water sources causing the spread of water-washed diseases, such as trachoma and scabies.[5] Low rainfall is linked to increased incidence of diarrhea.[32] There is some evidence that dry, dusty conditions created by drought and winds contribute to the seasonal patterns in the Meningitis Belt in Africa.[33–36] The prevalence of HIV/AIDS is noted to amplify the effect of drought on nutrition in Southern Africa.[37]

Infectious Disease

Infectious diseases and their transmission are strongly influenced by their ecological environments and local climatic conditions. Transmission of diseases can occur by direct contact, through indirect contact (for example, plants, animals, water), or from a pathogenic vector (for example, mosquitoes, ticks, rats).[9] Exposure to infectious diseases has altered because of changes in temperature, humidity, rainfall, and sea-level rise.[38] Recently there has been an increase in infectious diseases, and their geographic distribution.[9] Some evidence of changes in the distribution of mosquito, tick, and bird vectors has been attributed to climate change.[5]

Among the numerous VBDs, some of the most frequently studied are those transmitted by mosquitoes. Ambient temperatures affect the life cycle and abundance of the mosquito. Flooding and drought can have complicated effects on mosquitoes and subsequently infectious diseases. Disease outbreaks can occur after flooding because of the increases in stagnant water, but breeding sites also can be washed away.[39] Similarly, drought conditions generally limit mosquito habitats, but subsequent rainfall events can promote mosquito activity.[3] Mosquitoes can transmit diseases, such as malaria, dengue, yellow fever, and Japanese encephalitis, but their sensitivity to weather conditions can inhibit or enhance their efficacy as a vector. Malaria is spread by mosquitoes, which are inhibited from transmitting the parasite in cooler temperatures.[9] Incidence of malaria is sensitive to changes in climate and has been implicated in the deaths of nearly 1 million people (predominantly children) annually.[26] Epidemics can be created by small changes in environmental conditions.[10] Malaria rates exhibited a fivefold increase after a six-week flood in Mozambique [18] and strong ENSO conditions are related to outbreaks of malaria in South Asia and South America.[10] Decadal decreases in rainfall also are linked to reduced malaria transmission rates.[5]

Despite the many locations where changes in weather patterns affect malaria risk, the effect of climate change on this risk is more complicated.[5] The resurgence of highland malaria in Africa may be linked to climate change and warming, but the conflicting results from numerous studies mean this link is inconclusive.[5, 38] A study from the East African highlands showed increases in malaria cases

correspond to long-term temperature increases.[40] Studies relating to temperature or rainfall changes to malaria transmission in Madagascar,[41] Ethiopia,[42] and East Africa [43] indicate that short-tem climate variability rather than long-term climate change is responsible for this relationship.[5] Population growth and land use changes also contribute to an increase in malaria incidence in East Africa.[44] Deforestation may have increased the risk of malaria in the Africa and the Amazon.[45, 46] These incidences are complicated by a lack of historical observations, complex disease dynamics, and other nonclimatic factors.[5] Harvell et al. [47] note that a recent increase in malaria transmission is likely to be attributable to antimalarial drug resistance, failed vector control programs, and climatic change.

Dengue fever is another VBD, but unlike the malaria parasite, four viruses transmitted by mosquitoes cause dengue.[38] Dengue is a tropical urban disease that is rapidly spreading,[26] and it is the most important vector-borne viral disease in the world.[5] Increases in the disease over the last 40 years are primarily a result of urban growth and mosquito breeding sites.[26] The seasonality of increasing dengue transmission during warmer, more humid months and periods of higher rainfall totals show the influence of climatic variability.[16] Variability in dengue incidence has been attributed to temperature and rainfall changes associated with ENSO conditions.[48, 49] The complexities of linking climate change to dengue are similar to those discussed previously for malaria.

Other major mosquito-borne diseases that may be sensitive to climate change include yellow fever, West Nile fever, chikungunya fever, and filariasis.[26, 50] These diseases show similar sensitivity to climate change as malaria and dengue, because of its influence on the mosquito life cycle.[5] More specifically West Nile fever and chikungunya fever have been associated with outbreaks of disease following El Niño rainfall events and La Niña drought events, respectively.[51, 52] Warm winters followed by spring droughts are thought to have a role in the amplification of the West Nile virus.[53]

Tick-borne diseases are the most common VBD to affect temperate regions.[38] Lyme disease and tick-borne encephalitis (TBE) are two well-monitored and studied diseases carried by ticks. Lyme disease is spreading northwards in North America and Europe as winters warm.[53] Similarly, TBE has shown a shift to the north and to higher altitudes in Europe.[54] This northward expansion linked to recent warming has also been noted in the incidence of schistosomiasis [55] and leishmaniasis [56] transmitted by snail and sand fly vectors, respectively.

Pandemics of the plague are prevalent throughout history and recent recurrences in the southwestern United States are linked to late winter–early spring precipitation events and are strongly influenced by ENSO and the PDO.[57, 58] Furthermore the global reemergence of the plague is linked with warmer wetter conditions regionally, which increase flea and rodent host populations, and subsequently plague transmission rates.[58] Confalonieri et al. [5] also note that periods of heavy rainfall and flooding can result in increased human contact with disease-carrying rodents with diseases such as leptospirosis.

Other Disease-Climate Relationships

Air quality and pollutants are affected by the weather and climate, and they can cause negative health effects.[59] Asthma rates in the United States have quadrupled

since 1980, with rising carbon dioxide levels, dust, and poor air quality contributing to this increase.[53] Meteorological conditions affect the formation and dispersal of pollutants. Ground-level ozone is a pollutant that is formed by photochemical reactions in hot, sunny conditions.[5] Respiratory and cardiovascular diseases are created and exacerbated by air pollution.[26] Ground-level ozone levels are generally increasing, although regional variability exists.[60, 61] Higher ozone levels are associated with increased hospital admissions and mortality rates.[62]

There is stronger evidence of negative health impacts from fine particulate matter (PM) than for ozone.[5] Chemical components of PM from vehicle emissions, diesel, and wood burning showed the highest contribution to risk of hospitalization in the United States.[63] Air pollutants from forest fire smoke can cause adverse eye and respiratory health effects, separate from direct burns and injuries.[12] The effect of dust storms is still unclear, although links have been made between the incidence of some diseases and dust events. In the southwest United States dust storms increase coccidioidomycosis fungal disease. Pollen, fungal spores, and other aeroallergens are generally associated with the seasons; however, recent climate change has altered seasonal patterns of some vegetation types as well as the amount of allergens in the atmosphere affecting asthma and allergy rates.[19]

Although not directly related to climate changes, depleted levels of stratospheric ozone increase the incidence of skin cancer, ocular disorders, and immune diseases, and has implications for both excess and deficiencies of sunlight that add to the global disease burden.[64, 65] These health problems are a long-term issue due to the slow rate of recovery of the ozone layer. Carbon dioxide–induced climate change may further slow this recovery.

Sea-level rise associated with climate change is another source of concern for human health, especially with a large proportion of global populations living in low-lying areas. A number of small island states and low-lying areas are at risk.[66] Rising sea levels are implicated in increases in coastal flooding, loss of cropland, salinization of freshwater resources, the spread of infectious diseases, and population displacement.[66, 67] Saltwater intrusion into groundwater aquifers has contributed to sea-level change,[68] which affects water quality. The effect of Hurricane Katrina on New Orleans exhibits the risk that low-lying densely populated urban areas face.[5] Warming of coastal waters contributes to increasing incidence of diseases and infections from algal blooms and biotoxin poisonings.[69, 70] However, there is no definitive evidence of climate change on coastal disease patterns.[15] Increasing temperatures are known to cause a linear increase in the incidence of food poisoning from salmonella [71] and to a lesser extent for campylobacter bacteria.[72]

FUTURE IMPACTS AND VULNERABILITIES

The implications of climate change on human health in the future are complex. Population health is influenced by a number of factors, including social, environmental, and economic conditions.[73] Projected impacts of climate change on specific diseases generally are limited to those with strong epidemiological evidence. Assumptions are also made based on current understanding of the interaction between climate and diseases. The IPCC [1] projections show that temperatures will continue to warm based on the most likely emission scenarios. Therefore, trends in

human health associated with recent warming are likely to continue. However, other external influences may affect future patterns human health. Vulnerability to climate change will depend on responses to prevention, adaptation, adaptive capacity, mitigation, and future advances in disease control. Socioeconomic conditions are an important factor in determining health impacts. Although the risk of climate change on human health is global, it is disproportionately skewed toward the poor.[74]

Some health benefits of climate change are projected, such as decreases in cold-related mortality and pollutant-related mortality where local abatement strategies are enforced.[5] Temperature or rainfall-related changes are also projected to reduce the disease distribution where thresholds are exceeded for vectors or parasites.[5] However, projections show climate change will have predominantly negative consequences for human health.[19] Modest changes in the burden of climate-sensitive health outcomes are projected to occur over the next few decades, followed by larger increases in the second half of the 21st century.[5]

A global burden of disease study by the WHO estimated the amount of premature morbidity and mortality caused by climate change. The effect of climate change on malnutrition, diarrhea, malaria, and floods by 2030 showed disease benefits would be outweighed by increased risk.[75] Small proportional changes in diarrhea and malnutrition have a large effect on the disease burden,[76] especially for poorer nations who currently experience a large burden of disease.[75] Coastal floods also will have a large proportional effect on mortality rates, but the low burden of disease means the net effect will be relatively small.[5] However a number of uncertainties surround these estimates, including those associated with climate model projections, longitudinal climate-disease relationships, and population or socioeconomic dynamics.[76]

Models can include assumptions, including population scenarios, acclimatization, and adaptation to climate change. Bed nets and indoor sprays are effective disease-control methods [77] and changes in their usage will affect future disease rates. Because of the uncertainties associated with projecting future disease dynamics, there is more confidence in projected changes of the geographic range of vectors than for disease incidence.[5] Changes to malaria mosquito viable regions under future climate scenarios show regional expansion and contraction both spatially and seasonally.[78] Predictions show the mosquito-borne dengue transmission season will lengthen, and the latitudinal and altitudinal range of the disease will increase.[14, 79] Dengue transmission is projected to increase by 2 billion by the 2080s because of climate change, if no other determinants changed.[14] Two tick-borne diseases, TBE and Lyme disease, are also predicted to shift geographically in range, although an increase in incidence has been predicted only for the latter.[80, 81] An understanding of how climate affects infectious disease is insufficient to quantify predictions with confidence.[23]

Projected warming trends have a more direct influence on future heat and cold related mortality. Cold-related morbidity and mortality will be reduced by warming, especially in the higher northern latitudes.[26] The population at risk of heat-related mortality is highly likely to increase in frequency, intensity, and duration.[12] Those most at risk are likely to be young, old, ill, or poor.[82] Research on the effect of thermal stress on mortality is focused on developed nations [5] whose populations are better able to adapt to such conditions than developing countries. Some

acclimatization will occur due to warming temperatures and may mitigate some of the effects of extreme heat events on human health.

Air quality scenarios for urban areas have been modeled, with increases in backgrounds levels of ground-level ozone predicted,[83, 84] contributing to future cardiovascular and pulmonary illness.[82, 85] Scenarios are dependant on future emissions and weather patterns to create the local conditions that cause ozone events.[86] Scenario-based predictions for future particulate matter are uncertain as studies show conflicting results,[84] although increases in particulate matter from more frequent wildfires (associated with climate change) may become increasingly important.[87] Some scenarios show increases in risk, but the potential for the spread of disease may not reflect actual risk because of the effectiveness of disease control activities, public health systems, or adaptive capacity.[5, 75, 78]

Models make assumptions about greenhouse gas emissions and warming. Limitations are placed on the models as the relationship between climate change and human health is complex and often interrelated with other environmental determinants. Therefore the lack of certainty about the future impacts of climate change on human health means that this is an active area of current research.

ADAPTATION, IMPLICATIONS, AND PUBLIC HEALTH POLICY

Much uncertainty surrounds predictions of climate change on human health. Figure 4.2 shows the pathways by which climate change affects human health, including local modulating influences. Mitigation, adaptation, and health policy advances will have a large role in determining the effects of climate change on health. The WHO has been making efforts to enhance public health services related to climate change, passing a resolution on climate change and health in May 2008. Control of disease vectors and sanitation is essential to protect human health.[26] Better monitoring of population health data is required to provide evidence of the effects of climate change on health, especially in lower income regions that are the most vulnerable populations. Improved economic conditions in many countries will reduce their vulnerability and increase capacity to adapt.[88] Few modeling studies predicting disease incidence account for adaptive capacity.[89] Enhancements made to model performance and more comprehensive data sets will allow for a better understanding of the relationship between climate and disease, and therefore their future impacts.

Population dynamics such as migration are a complicating variable for predicting future disease patterns. The spread of disease is enhanced by population mobility and transportation systems.[90] Both climate change and extreme events can cause migration, where they may encounter or introduce disease.[19] Temporary migration patterns are likely to shift with climate change.[19] Climate change is expected to intensify additional contributors to infectious disease emergence, including global trade and transportation, land use, and human migration.[91] Education, awareness, adaptive capacity, infrastructure, and technology advances will help to reduce future disease burdens.

Forecasting and predictions will be important for future climate-related health events. ENSO predictability allows El Niño and La Niña events, which often are associated with disasters, to be forecast and allow preventative measures to be taken. Prediction and early detection systems enable a reduction in morbidity and mortality associated with climate events through control and mitigation measures. Developments

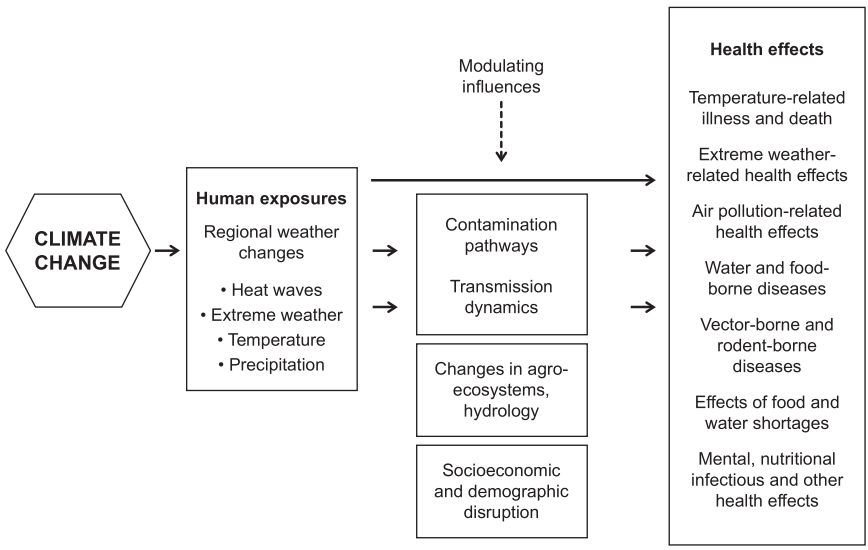

Figure 4.2 Pathways by Which Climate Change Affects Human Health

in both climate and disease modeling will allow more confidence to be placed in projections for the future.

REFERENCES

1. Intergovernmental Panel on Climate Control. Summary for policymakers. In: Solomon S, Qin D, Manning M, et al., eds. *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press; 2007.
2. National Climatic Data Center. *Climate of 2008: Annual Report*. Asheville, NC: National Oceanic and Atmospheric Administration; 2009. http://www.ncdc.noaa.gov/oa/climate/research/2008/ann/global.html.
3. Haines A. Climate change and health: strengthening the evidence base for policy. *Am J Prev Med*. 2008;35:411–413. doi:10.1016/j.amepre.2008.08.003.
4. Jones KE, Patel NG, Levy MA, et al. Global trends in emerging infectious diseases. *Nature*. 2008;451:990–993.
5. Confalonieri U, Menne B, Akhtar R, et al. Human health. In: Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson CE, eds. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, UK: Cambridge University Press; 2007:391–431.
6. McMichael AJ. Global climate change and health: an old story writ large. In: McMichael AJ, Campbell-Lendrum D, Corvalan C, et al., eds. *Climate Change and Human Health: Risks And Responses*. Geneva, Switzerland: World Health Organization; 2003:1–17.
7. Patz JA, McGeehin MA, Bernard SM, et al. The potential health impacts of climate variability and change for the United States: executive summary of the report of

the health sector of the US National Assessment. *Environ Health Perspect*. 2000;108:367–376.
8. World Health Organization. *El Niño and Its Health Impact*. Geneva, Switzerland: World Health Organization; 2000. http://www.who.int/mediacentre/factsheets/fs192/en/index.html. Fact Sheet, No.192.
9. McMichael AJ, Woodruff R. Climate change and infectious disease. In: Mayer KH, Pizer H, eds. *Social Ecology of Infectious Disease*. New York: Academic Press; 2008:378–407.
10. Kovats RS, Bouma MJ, Hajat S, Worrall E, Haines A. El Niño and health. *Lancet*. 2003;362:1481–1489.
11. Trenberth KE, Jones PD, Ambenje P, et al. Observations: surface and atmospheric climate change. In: Solomon S, Qin D, Manning M, et al., eds. *Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK/New York: Cambridge University Press; 2007:235–335.
12. Ebi KL, Balbus J, Kinney PL, et al. Effects of global change on human health. In: Gamble JL, Ebi KL, Sussman FG, Wilbanks TJ, eds. *Analyses of the Effects of Global Change on Human Health and Welfare and Human Systems. A Report by the US Climate Change Science Program and the Subcommittee on Global Change Research*. Washington, DC: US Environmental Protection Agency; 2008:39–87.
13. Curriero FC, Heiner KS, Samet JM, Zeger SL, Strug L, Patz JA. Temperature and mortality in 11 cities of the Eastern United States. *Am J Epidemiol*. 2002;155:80–87.
14. Whitman S, Good G, Donoghue ER, Benbow N, Shou W, Mou S. Mortality in Chicago attributed to the July 1995 heat wave. *Am J Public Health*. 1997;87:1515–1518.
15. Kovats RS, Haines A. Global climate change and health: recent findings and future steps. *Can Med Assoc J*. 2005;172:501–502.
16. Hales S, Edwards SJ, Kovats RS. Impacts on health of climate extremes. In: McMichael AJ, Campbell-Lendrum D, Corvalan C, et al., eds. *Climate Change and Human Health: Risks and Responses*. Geneva, Switzerland: World Health Organization/World Meteorological Organization/United Nations Environment Programme; 2003:79–102.
17. Hemon D, Jougla E. La canicule du mois d'aout 2003 en France [The heatwave in France in August 2003]. *Rev Epidémiol Santé Publique*. 2004;52:3–5.
18. Epstein PR. Climate change and human health. *N Engl J Med*. 2005;353:1433–1436.
19. Institute of Medicine. *Global Climate Change and Extreme Weather Events: Understanding the Contributions to Infectious Disease Emergence*. Washington, DC: The National Academies Press; 2008.
20. United States Environmental Protection Agency. *Excessive Heat Events Guidebook*. Washington, DC: Environmental Protection Agency; 2006. http://www.epa.gov/heatisland/about/heatguidebook.html.
21. European Environment Agency. *Air Pollution by Ozone in Europe in Summer 2003: Overview of Exceedances of EC Ozone Threshold Values during the Summer Season April–August 2003 and Comparisons with Previous Years*. Copenhagen, Denmark: European Economic Association; 2003. http://reports.eea.europa.eu/topic_report_2003_3/en. Topic Report, No.3/2003.
22. Ranhoff AH. Accidental hypothermia in the elderly. *Int J Circumpolar Health*. 2000;59:255–259.
23. Burke DS. Impacts of global climate change on infectious diseases. In: Institute of Medicine, ed. *Global Climate Change and Extreme Weather Events: Understanding the*

Contributions to Infectious Disease Emergence. Washington, DC: National Academies Press; 2008:90–99.
24. Healy JD. Excess winter mortality in Europe: a cross country analysis identifying key risk factors. *J Epidemiol Community Health*. 2003;57:784–789. doi:10.1136/jech.57.10.784.
25. Carson C, Hajat S, Armstrong B, Wilkinson P. Declining vulnerability to temperature-related mortality in London over the 20th century. *Am J Epidemiol*. 2006;164:77–84.
26. World Health Organization. World health day 2008: protecting health from climate change. http://www.who.int/world-healthday/toolkit/report_web.pdf.
27. Hajat S, Ebi KL, Kovats S, Menne B, Edwards S, Hames A. The human health consequences of flooding in Europe and the implications for public: a review of the evidence. *Applied Environ Sci Public Health*. 2003;1:13–21.
28. Ahern M, Kovats RS, Wilkinson P, Few R, Matthies F. Global health impacts of floods: epidemiologic evidence. *Epidemiol Rev*. 2005;27:36–46.
29. Scheuren JM, le Polain de Waroux O, Below R, Guha-Sapir D, Ponserre S. Annual disaster review 2007. http://www.emdat.be/Documents/Publications/Annual%20Disaster%20Statistical%20Review%202007.pdf. Published 2008.
30. Wilson J, Philips B, Neal D. Domestic violence after disaster. In: Enearson E, Morrow BH, eds. *The Gendered Terrain of Disaster: Through Women's Eyes*. Miami, FL: International Hurricane Center, Florida International University; 1998:115–122.
31. Galea S, Nandi A, Vlahov D. The epidemiology of post-traumatic stress disorders after disasters. *Epidemiol Rev*. 2005;27:78–91.
32. Lloyd SJ, Kovats RS, Armstrong BG. Global diarrhea morbidity, weather and climate. *Clim Res*. 2007;34:119–127.
33. Molesworth AM, Cuevas LE, Connor SJ, Morse AP, Thomson MC. Environmental risk and meningitis epidemics in Africa. *Emerg Infect Dis*. 2003;9:1287–1293.
34. Yaka P, Sultan B, Broutin H, Janicot S, Philippon S, Fourquet N. Relationships between climate and year-to-year variability in meningitis outbreaks: a case study in Burkina Faso and Niger. *Int J Health Geogr*. 2008;7:34.
35. Teyssou R, Rouzic EML. Meningitis epidemics in Africa: a brief overview. *Vaccine*. 2007;25:A3–A7.
36. Sultan B, Labadi K, Guegan JF, Janicot S. Climate drives the meningitis epidemics onset in West Africa. *PLoS Med*. 2005;2:e6. doi: 10.1371/journal.pmed.0020006.
37. Mason JB, Bailes A, Mason KE, et al. AIDS, drought, and child malnutrition in southern Africa. *Public Health Nutr*. 2005;8:551–563.
38. McMichael AJ, Githeko A, Akhtar R, et al. Human population health. In: McCarthy JJ, Canziani OF, Leary NA, Dokken DJ, White KS, eds. *Climate Change 2001: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press; 2001:453–485.
39. Haines A. Climate change, extreme events and human health. In: Institute of Medicine, ed. *Global Climate Change and Extreme Weather Events: Understanding the Contributions to Infectious Disease Emergence*. Washington, DC: National Academies Press; 2008:57–74.
40. Pascual M, Ahumada JA, Chaves LF, Rodó X, Bouma M. Malaria resurgence in the East African highlands: temperature trends revisited. *Proc Natl Acad Sci U S A*. 2006;103:5829–5834.
41. Bouma MJ. Methodological problems and amendments to demonstrate effects of temperature on the epidemiology of malaria: a new perspective on the highland epidemics in Madagascar, 1972–1989. *Trans R Soc Trop Med Hyg*. 2003;97:133–139.

42. Abeku TA, De Vlas SJ, Borsboom GJ, et al. Effects of meteorological factors on epidemic malaria in Ethiopia: a statistical modeling approach based on theoretical reasoning. *Parasitology*. 2004;128:585–593.
43. Zhou G, Minakawa N, Githeko AK, Yan G. Climate variability and malaria epidemics in the highlands of East Africa. *Trends Parasitol*. 2005;21:54–56.
44. Patz JA, Olsen SH. Malaria risk and temperature: influences from global climate change and local land use practices *Proc Natl Acad Sci U S A*. 2006;103:5635–5636.
45. Guerra CA, Snow RW, Hay SI. A global assessment of closed forests, deforestation and malaria risk. *Ann Trop Med Parasitol*. 2006;100:189–204.
46. Vittor AY, Gilman RH, Tielsch J, et al. The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of falciparum malaria in the Peruvian Amazon. *Am J Trop Med Hyg*. 2006;74:3–11.
47. Harvell CD, Mitchell CE, Ward JR, et al. Climate warming and disease risks for terrestrial and marine biota. *Science*. 2002;296:2158–2162.
48. Gagnon AS, Bush ABG, Smoyer-Tomic KE. Dengue epidemics and the El Niño southern oscillation. *Clim Res*. 2001;19:35–43.
49. Cazelles B, Chavez M, McMichael AJ, Hales S. Nonstationary influence of El Niño on the synchronous dengue epidemics in Thailand. *PLoS Med*. 2005;2:e106. doi:10.1371/journal.pmed.0020106.
50. Haines A, Kovats RS, Campbell-Lendrum D, Corvalan C. Climate change and human health: impacts, vulnerability, and mitigation. *Lancet*. 2006;367:2101–2109.
51. Linthicum K J, Anyamba A, Tucker CJ, Kelley PW, Myers MF, Peters CJ. Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science*. 1999;285:397–400.
52. Chretien JP, Anyamba A, Bedno SA, et al. Drought-associated chikungunya emergence along coastal East Africa. *Am J Trop Med Hyg*. 2007;76:405–407.
53. Center for Health and the Global Environment. *Climate Change Futures: Health, Ecological and Economic Dimensions*. Cambridge, MA: Harvard Medical School; 2005.
54. Süss J, Klaus C, Gerstengarbe FW, Werner PC. What makes ticks tick? climate change, ticks, and tick-borne diseases. *J Trav Med*. 2008;15:39–45.
55. Yang GJ, Vounatsou P, Zhou XN, Tanner M, Utzinger J. A potential impact of climate change and water resource development on the transmission of *Schistosoma japonicum* in China. *Parassitologia*. 2005;47:127–134.
56. Lindgren E, Naucke T. Leishmaniasis: influences of climate and climate change epidemiology, ecology and adaptation measures. In: Menne B, Ebi K, eds. *Climate Change and Adaptation Strategies for Human Health*. Darmstadt, Germany: Steinkopff; 2006:131–156.
57. Enscore RE, Biggerstaff BJ, Brown TL, et al. Modeling relationships between climate and the frequency of human plague cases in the southwestern United States, 1960–1997. *Am J Trop Med Hyg*. 2002;66:186–196.
58. Stenseth NC. Plague and climate. In: Institute of Medicine, ed. *Global Climate Change and Extreme Weather Events: Understanding the Contributions to Infectious Disease Emergence*. Washington, DC: National Academies Press; 2008:128–143.
59. Street R, Maarouf A, Jones-Otazo H. Extreme weather and climate events–implications for public health. In: Ebi KL, Smith J, Burton I, eds. *Integration of Public Health with Adaptation to Climate Change: Lessons Learned and New Directions*. Leiden, the Netherlands: Taylor and Francis; 2005:161–190.
60. Chan E. Regional ground-level ozone trends in the context of meteorological influences across Canada and the Eastern United States from 1997 to 2006. *J Geophys Res*. 2009;114:D05301. doi:10.1029/2008JD010090.

61. Jaffe D, Ray J. Increase in surface ozone at rural sites in the Western US. *Atmos Environ*. 2007;41:5452–5463.
62. Bell ML, Dominici F. Effect modification by community characteristics on the short-term effects of ozone exposure and mortality in 98 US communities. *Am J Epidemiol*. 2008;167:986–997.
63. Peng RD, Bell ML, Geyh AS, et al. Emergency admissions for cardiovascular and respiratory diseases and the chemical composition of fine particle air pollution. *Environ Health Perspect*. 2009;117:957–963.
64. McMichael AJ, Lucas R, Ponsonby AL, Edwards S. Stratospheric ozone depletion, ultraviolet radiation and health. In: McMichael AJ, Campbell-Lendrum D, Corvalan C, et al., eds. *Climate Change and Human Health: Risks and Responses*. Geneva, Switzerland: World Health Organization/World Meteorological Organization/United Nations Environment Programme; 2003:159–180.
65. Lucas RM, McMichael AJ, Armstrong BK, Smith WT. Estimating the global disease burden due to ultraviolet radiation exposure. *Int J Epidemiol*. 2008;37:654–667.
66. Nicholls RJ, Wong PP, Burkett VR, et al. Coastal systems and low-lying areas. In: Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson CE, eds. *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press; 2007:315–356.
67. Campbell-Lendrum D, Pruss-Ustun A, Corvalan C. How much disease could climate change cause? In: McMichael AJ, Campbell-Lendrum D, Corvalan C, et al., eds. *Climate Change and Human Health: Risks and Responses*. Geneva, Switzerland: World Health Organization/World Meteorological Organization/United Nations Environment Programme; 2003:133–158.
68. Werner AD, Simmons CT. Impact of sea-level rise on sea water intrusion in coastal aquifers. *Ground Water*. 2009;47:197–204.
69. Hunter PR. Climate change and waterborne and vectorborne disease. *J Appl Microbiol*. 2003;94:37–46.
70. Tuyet DT, Thiem VD, Von Seidlein L, et al. Clinical, epidemiological, and socioeconomic analysis of an outbreak of *Vibrio parahaemolyticus* in Khanh Hoa Province, Vietnam. *J Infect Dis*. 2002;186:1615–1620.
71. Kovats RS, Edwards S, Hajat S, Armstrong B, Ebi KL, Menne B. The effect of temperature on food poisoning: time series analysis in 10 European countries. *Epidemiol Infect*. 2004;132:443–453.
72. Tam C, Rodrigues L, O'Brien S, Hajat S. Temperature dependence of reported *Campylobacter* infection in England, 1989–1999. *Epidemiol Infect*. 2006;134:119–125.
73. Pitcher H, Ebi K, Brenkert A. Population health model for integrated assessment models. *Clim Change*. 2008;88:35–57. doi: 10.1007/s10584-007-9286-8.
74. Patz JA, Gibbs HK, Foley JA, Rogers JV, Smith KR. Climate change and global health: quantifying a growing ethical crisis. *Ecohealth*. 2007;4:397–405.
75. McMichael AJ. Climate change. In: Ezzati M, Lopez A, Rodgers A, Murray C, eds. *Comparative Quantification of Health Risks: Global and Regional Burden of Disease Due to Selected Major Risk Factors*; vol. 2. Geneva, Switzerland: World Health Organization; 2004:1543–1649.
76. Campbell-Lendrum D, Woodruff R. Comparative risk assessment of the burden of disease from climate change. *Environ Health Perspect*. 2006;114:1935–1941.
77. Breman JG, Mills A, Snow RW, et al. Conquering malaria. In: Jamison DT, Breman JG, Measham AR, et al., eds. *Disease Control Priorities in Developing Countries*. Oxford: Oxford University Press; 2006:413–431.

78. van Lieshout M, Kovats RS, Livermore MTJ, Martens P. Climate change and malaria: analysis of the SRES Climate and socio-economic scenarios. *Glob Environ Change*. 2004;14:87–99.
79. Jetten TH, Focks DA. Potential changes in the distribution of dengue transmission under climate warming. *Am J Trop Med Hyg*. 1997;57:285–97.
80. Randolph SE. The shifting landscape of tick-borne zoonoses: tick-borne encephalitis and Lyme borreliosis in Europe. *Philos Trans R Soc Lond Biol Sci*. 2001;356:1045–1056.
81. Ogden NH, Maarouf A, Barker IK, et al. Climate change and the potential for range expansion of the Lyme disease vector *Ixodes scapularis* in Canada. *Int J Parasitol*. 2006;36:63–70.
82. National Institutes of Health. Health effects of climate change. www.nih.gov/about/researchresultsforthepublic/climate_change.pdf. Published 2009.
83. Prather M, Gauss M, Berntsen T, et al. Fresh air in the 21st century? *Geophys Res Lett*. 2003;30:1100. doi:10.1029/2002GL016285.
84. Kinney PL. Climate change, air quality, and human health. *Am J Prev Med*. 2008;35459–467.
85. Bell ML, Goldberg R, Hogrefe C, et al. Climate change, ambient ozone, and health in 50 US cities. *Clim Change*. 2007;8261–76. doi:10.1007/s10584-006-9166-7.
86. Mickley LJ, Jacob DJ, Field BD, Rind D. Effects of future climate change on regional air pollution episodes in the United States. *Geophys Res Lett*. 2004;30:L24103. doi:10.1029/2004GL021216.
87. Jacob DJ, Darrell A. Winner effect of climate change on air quality. *Atmos Environ*. 2009;43:51–63.
88. Ebi KL. Healthy people 2100: modeling population health impacts of climate change. *Clim Change*. 2008;88:5–19. doi:10.1007/s10584-006-9233-0.
89. Haines A. Climate change and health: strengthening the evidence base for policy. *Am J Prev Med*. 2008;35:411–413. doi:10.1016/j.amepre.2008.08.003.
90. Macpherson DW, Gushulak BD, Macdonald L. Health and foreign policy: influences of migration and population mobility. *Bull World Health Organ*. 2007;85:200–206.
91. Institute of Medicine. *Microbial Threats to Health: Emergence, Detection, and Response*. Washington, DC: The National Academies Press; 2003.

CHAPTER 5

Global Nutrition in the 21st Century: Opportunities and Challenges for the Developed and Developing Worlds

Allan J. Davison, MSc, PhD

INTRODUCTION: WHERE WE ARE AND HOW WE GOT HERE

Food is a basic human need. With roughly 1 billion humans suffering from overnutrition and a similar number unable to find enough food to subsist, no one seriously disagrees about the urgency of addressing world hunger. Yet, as we enter the second decade of the 21st century, hunger and malnutrition continue to be leading risks to health and well-being worldwide. Hunger and malnutrition adversely affect physical and mental development to the point that one in six individuals do not get a sufficient amount of food to be healthy enough to lead a normally active life. Many factors contribute to this tragic situation, including natural disasters (most recently in Haiti), conflict (much of Africa), poverty (a global reality), poor agricultural infrastructure, and financial crises (the global recession of 2008–2009).

Those who work to mitigate world-scale hunger and malnutrition generally agree on dimensions of hunger, its causes, and indeed on how to measure and address it.[1] The cost of alleviating hunger amounts to a levy on the rich of less than 1 percent of their earnings; however, implementing remedies often takes on a political and economic dimension that is as difficult to resolve as hunger itself. This chapter focuses on the social and economic issues of global nutrition; evaluates what is being done to alleviate hunger and malnutrition, especially in resource-poor countries where the need is greatest; and frames a perspective on what lies ahead. As the chapter unfolds, the reader might experience a mix of optimism, indignation, and impatience.

Global malnutrition is not new, but it was not until 1978 when a commission was formed to identify the challenges pivotal to the poor, namely the following: (1) the magnitude of, and trends in, undernutrition[1]; (2) the relationship of malnutrition

1. The Food and Agricultural Organization of the United Nations (FAO) World Health Organization (WHO) definition of undernutrition is habitual calorie intake below the calorie expenditure required to earn an income or otherwise gain access to resources sufficient to maintain a body weight consistent with health and the activities of daily life. This definition has problems, but it has utility in population studies.

to infectious diseases; and (3) orienting food policy and agriculture to problems of the neediest.[2] The content of the final report rocked the developed world. The statistics were, and remain, unimaginable. In 1970, a third of humankind, more than 1.4 billion people, went to bed hungry. Children were especially affected; undernutrition is estimated to have contributed to more than 9 million child deaths annually, two-thirds of which were in the first year of life.[2] As we enter the second decade of the 21st century, this situation has not improved substantially in many parts of the world.

Abruptly, the rich nations were convinced that "someone" must "do something." A decade of international ferment led to an unprecedented international consensus—a road map framed as eight specific goals called the Millennium Development Goals (MDGs), which consist of 21 quantifiable targets and 60 timelined indicators.[3] (See table 5.1 for the MDGs related to hunger and malnutrition.) At the turn of the 21st century, 189 nations signified their commitment to address the global crisis of poverty by signing the MDG agreement. What ensued is a cascade of initiatives that have had varying degrees of success—some spectacular, most steady, and a few catastrophic. A more in-depth discussion of the MDG is presented later in this chapter. First, to understand these events, we need to understand the contributing factors to hunger and malnutrition, who the hungry are, and the causes and correlates of extreme poverty.[3]

MALNUTRITION, HEALTH, AND THE POVERTY TRAP

Nutrition is the most immediately modifiable factor that affects individual and public health. A core concern in the science of nutrition is how to deliver the essentials of a healthy diet to a hungry world. Today, poverty is the major determinant of malnutrition and hunger. Consequently, there is little debate that improving nutrition is contingent on economic development.

The poverty trap can begin with failure of any of the three links—malnutrition, ill health, or economic marginalization. Once started, it continues, even across successive generations, unless there is outside intervention. When it occurs in a country, the cycle of poverty is called the "development trap." A major purpose of development aid is to get countries out of the vicious cycle and onto the ladder of economic development and sustainable agriculture. Ironically, most of the hungry are farmers and subsistence farmers working on small holdings, usually in rural areas. Their families' survival depends on what they can scrape from an uncooperative earth. What they eat this month is the yield from last month's planting, or what they can forage within a walk of several hours. A day's work often leaves the family hungry. In inaccessible areas, there is no paid work as there is no one to employ them. Even if they

2. The question of causes of hunger was not raised at this stage, but we will return to causes hereunder.
3. Extreme poverty refers to the condition of the poorest of the poor—not having the means to afford basic human needs such as clean water, nutrition, health care, clothing, and shelter. This is also referred to as absolute poverty or destitution. Relative poverty is having fewer resources or less income than others within a society or country, or compared with worldwide averages. Regression analysis shows that for a population the threshold income for extreme poverty is equivalent to US$1.25 at 2005 international prices. The common reference value "$1.25 PPP" is corrected for purchasing power parity and inflation.

have anything left over to sell, there are few purchasers of the product. If the family has "spare time," it is invested in fetching water or fuel (often animal dung) for a fire. Construction or repair of rudimentary shelters is preempted by foraging. In a settled community, the burial area with many small graves attests to the fact that one month's delay in the rainy season can be deadly, especially for the young.

Understanding the needs of hungry people need not be difficult. Fieldworkers hear the plea: please, no more anthropologists, no more needs assessments. Look around. If you can't *see* what we need, *ask* us. We will tell you. It is *not* money. Often one of the last things the poorest need is money. It has no intrinsic value to them; there is nowhere to spend it. A trading post may be 30 miles away. Their needs are much more fundamental than money. Their needs are for clean water to drink, water to use on the crops, quality seeds, fertilizers, low-technology tools, mosquito nets, perinatal clinics, and dispensaries.

HOW NATIONS BECOME POOR

Virtually all countries once had a sufficiency of food. Hunger came later, almost entirely for reasons unrelated to those affected. In fact, states fail for a multitude of reasons almost all external: armed conflict, being landlocked, uncertain rainfall and drought, being the recipient of international boycotts (for example, Zimbabwe), a scarcity of resources (for example, Lesotho), and an appropriation of resources without commensurate compensation (too numerous to list).[4, 5] Nobel Laureate Muhammad Yunus likens the process to taking a seed from a tall tree and planting it in a small pot where it becomes a bonsai. It is futile to examine the seed to learn why it became poor. The causes of poverty are multifactorial after all.[6]

Most "poor" countries are not poor at all when one considers their natural resources, which in many cases could feed most if not all the inhabitants were they afforded the means to realize their initial endowment.[5] As it is, the copper of Chile or Katanga, or the oil of Nigeria disappears with no sustained trickle-down effect to the poorest. The Gold Coast of Africa now has no gold, and its people are none the richer. However, the matter does not rest there; the discovery of rich new resources in a poor country is usually followed by *deepening* poverty for the populace.[7] The conclusion that a nation can become poorer by finding resources may elicit skepticism in the wealthy nations, but it has considerable currency in nations recently emerged from colonial status. The hungry of Africa remember the enthusiasm the settlers had for their vanadium and chrome, and the settlers remember the glitter in the eyes of the British when they annexed the diamond fields of Kimberly and then declared war on the nascent Transvaal republic when gold was discovered.

The process of foreign land acquisition has increased dramatically over the recent decades.[8] In the past five years, for example, developing countries[4] have seen the purchase of 5 million hectares of farmable land by foreign interests. The

4. The term *developing country* is inconsistently used, and so we will use the imperfect World Bank definition, which considers all low- and middle-income countries as "developing." In its most recent classification, countries with gross national incomes (GNI) per capita below US$11,905 in 2008 were considered in this category.

Congo, Ethiopia, Mali, Ghana, Indonesia, and other nations have sold potential farmland to agribusiness in the United States, Norway, Egypt, India, South Africa, Saudi Arabia, and South Korea.[8] Multinational entities have been a partner to this situation, presumably in the search for profit, with carbon credits and agrifuels providing an added incentive.[9] Whatever the motive, the consequences to the poor are the same: the loss of farmable land. Market forces are not just supply and demand. On a fair bargaining table, a government selling land could exact concessions of comparable value.[5]

OUTMODED BELIEFS ABOUT HUNGER AND POVERTY

It is no longer true that the poor nations are getting poorer. It is not true that most of the poor live in African countries with failing economies. Differences in average income among countries are diminishing, and most lower- and middle-income countries are already on the ladder of development. Currently, 80 percent of the poor do not live in Africa or in the poorest countries. Most of the poor live in economies that are growing impressively with increasing numbers of wealthy people—for example, India and China.[5] The exceptions are about 30 fragile or failing states, half of them south of the Sahara.

In contrast, within most countries, including the developed world, the income gap between the rich and the poor is increasing. Poverty persists even in the richest countries of the world. In the United States, for example, 50 percent of children younger than age 18 rely on government subsidies (for example, food stamps), and among ethnic minorities, the proportion can be as high as 80 percent.[10] In China, which met the first MDG 14 years ahead of the 2015 target date, stark inequities remain. While widespread hunger is no longer a problem in China, average income in the poorest provinces is about one-tenth that in the richer. In Canada, a wealthy nation by all standards, 70 percent of Inuit children live in food insecure households compared with 5.6 percent of non-Inuit households where children are food insecure.[11] In reality, there are no longer poor nations in which almost *everyone* is emaciated. While progress has been made, so much more needs to be done. Ironically, in many nations where hunger and malnutrition are prevalent, we now see the growing problems of overnutrition, which is becoming a major public health problem. In these nations, such as India, obesity and diabetes increasingly contribute to the burden of heart disease. Of the 250 million people with diabetes, 80 percent live in low- and middle-income countries, where 12–20 percent of adults in countries in Asia, the Middle East, and the Caribbean are affected. By 2025, 80 percent of the new cases of diabetes are estimated to be in developing countries.[12] In these countries, diabetes is diagnosed at a later stage where it is less treatable, and access to treatment is not always available. Consequently, the burden of diabetes, and the cost to society, affects developing countries much more severely. The dual burden of the cost of chronic diseases and diseases associated with poverty coexist.

5. Such concessions, for example, could include large-scale irrigation projects, leasing as opposed to sale of the land, or a first claim on the food produced.

GLOBAL AID

Almost no nation, especially developing nations, has achieved economic success without substantial outside financial help. For example, the richest nations in the world today include those in the European Union (EU) in part because of the Marshall Plan designed to rebuild Europe after World War II. There are, however, those who uncompromisingly object to aid of all kinds. Overwhelmingly they cite two reasons: (1) all the aid given to date "has had no effect," or (2) aid "ends up in numbered bank accounts of corrupt dictators." Both assertions are wildly inconsistent with reality. Yet, in most parts of the world, poverty is decreasing, and aid is at least partly responsible. Unfortunately, with the transfer of large sums of money and aid comes the temptation for graft and greed. While bribery is repellent, it is no more common among the poor than among the rich.[13] For every corrupt person who accepts a bribe, there is a corrupt person who offered one. Africans, for example, applaud the edict that aid must reach those for whom it is intended. When a dishonest official is bribed, the people suffer twice. First, a valuable resource flows out of their country at a fraction of its true worth. Then the proceeds of the bribe go out of the country via a foreign bank account. Twenty-four nations, mostly in Africa, have signed the "Extractive Industry Transparency Initiative," which contains rules to eliminate covert deals in this important area of trade.[14]

Aid can carry risks for both the donor and recipient. Indeed, certain kinds of aid are not merely valueless, they are harmful—for example, food aid was originally conceived in the 1950s as a way to dispose of U.S. food surpluses earmarked for burning. A potential solution is to give priority to purchasing food from local farmers, thus helping to preserve local capacity.

Despite the global need, statistical data show that agricultural aid declined drastically from 1980 to 2000. For example, Canada, New Zealand, and the United States each have decreased agricultural aid as a fraction of their total aid package by 20 percent to 4 percent.[15] In absolute amounts, EU aid totaled more than $59 billion, while the United States (with roughly the same population) gave less than half that amount.[16] Moreover, a major fraction of aid from the member states of the North Atlantic Treaty Organization (NATO) goes to nations that have strategic importance to NATO rather than to the neediest nations. Aid, when given, can come with unacceptable strings attached. A common requirement in World Bank loans is that recipients must open their markets to product, including foods from the developed world. Most developed nations subsidize farmers so that their foods can flow, below production cost, into the recipient country, beggaring their farmers.[17] Meanwhile, the rich countries, almost universally, limit their markets to food products from the developing world. The most visible critic of this kind of aid is Dambisa Moyo, Zambian economist and author of *Dead Aid*. Moyo argues against aid to Africa and is a proponent for more innovative ways for Africa to finance development, including trade with China, accessing the capital markets, and microfinance.[18] She asks why Africa is wasting its time in the World Trade Organization, which will never remove trade barriers. Instead Africa should foster its relationship with countries that are interested in actually buying African produce. If Western Hemisphere markets do not want to trade with Africa to protect their own markets, Africa should focus on China and India.

PROGRESS TOWARD THE MDGs AT THE HALFWAY MILESTONE

The 15-year timescale of the MDGs was set by the willingness of the 20 wealthiest nations to provide financial resources to the less developed nations. Five European nations have consistently provided the pledged 0.7 percent or more of their gross domestic product (GDP) (Sweden 0.98 percent; Luxembourg 0.92 percent; Norway 0.88 percent; Denmark 0.82 percent; Netherlands 0.80 percent). Nations outside the Development Assistance Committee also have provided assistance—for example, Kuwait gives 0.82 percent of its GDP and Saudi Arabia gives 0.4 percent. Other countries, despite their pledge, are contributing far less—for example, Australia, Canada, Japan, and the United States are paying one-third to one-sixth of what they pledged.[17]

At the halfway point, only one MDG goal (primary education) is likely to be met by 2015. The first "primary" goal, the mitigation of hunger, will not be met. Certainly, the global financial recession of 2008–2009 provided further setback to reaching this goal, and, not surprisingly, hunger and poverty increased during this period. Specific measures of hunger and poverty are tracked by numerous organizations, including the World Bank, Food and Agriculture Organization of the United Nations (FAO), the U.S. Department of Agriculture (USDA), and others, and the indication is that far too many around the world go to bed hungry or are considered malnourished.[19] Although organizations assess hunger using differing criteria, the most useful measure is the World Bank's US$1.25 per day purchasing power parity (PPP).[6] The percentage of people living below US$1.25 PPP per day is based on World Bank data.[20] Between 1980 and 2005, the percentage of people in poverty decreased at an overall average of 1.25 percent per year. Inherent in this less than robust figure is the fact that substantial differences exist among geographic areas. This modest overall improvement reflects progress in the large population economies in East and South Asia. In Latin America, improvements are small. Although Africa as a whole is improving, the Sahel[7] and other war-torn nations inside Africa are rapidly deteriorating.[21] Afghanistan remains one of the poorest countries in the world. Haiti, too, is an example of a failing state. Long the poorest nation in its region, its economy has deteriorated steadily for 50 years, and without help, it seems destined to become one of the poorest in the world. The misery following the devastating earthquake of 2010 has captured the world's attention and brought promises of *sustainable* aid that this time perhaps will be kept. Thus, quite a lot of work remains to be done to boost nations and people out of poverty.

Without doubt, the 2008–2009 world recession has had a negative impact on efforts to reduce global hunger. Relying on current data and projections, the current recession will have an impact similar to that in 1986 and 1996. GDP (inflation adjusted) dropped by 2.5 percent from 2008 to 2009 in Sub-Saharan Africa (excluding

6. The figure of US$1.25 PPP represents original World bank US$1, adjusted for inflation to 2005, corrected country by country for purchasing power parity, and further adjusted within a country for urban and rural lifestyles reflecting the sources and costs of food, shelter, and other necessities of life. World Bank changed this yardstick in 2008 and has not updated their earlier tables since then, consequently the data do not yet reflect the economic problems of 2008.
7. Defined here as Burkina Faso, Chad, The Gambia, Guinea Bissau, Mali, Mauritania, Niger, Senegal, and Sudan.

South Africa). This aggregate for Sub-Sahara likely reflects worldwide events, and substantiate FAO historical data that show that recessions typically lead to a one- or two-year downturn with four to six years required to overtake pre-recession levels.[22] Although the setback hopefully will be temporary, it will not be without cost. Many lives undoubtedly will be lost. The increasing population of some developing countries certainly steepens the path to achieving the MDGs, but this was factored into the timeline. Furthermore, the world's agricultural capacity undoubtedly has limits; unless the world moves to contain population growth, limits to food and other resources are likely to be reached this century.

BEYOND THE MDGs

Among several collections of goals for eliminating global nutrition disparities, the MDGs are considered by many to be the most compelling. They acknowledge the inseparability of health, nutrition, and development, and emphasize poverty as an underlying mechanism impeding progress. Many view the MDGs as the best available path for the reduction and even the elimination of hunger, there are also many other innovative projects that are synergistic with the MDGs. Three will be discussed here: (1) the Millennium Village Projects (MVPs); (2) the Grameen enterprises and "microcredit"; and (3) an unprecedented outpouring of philanthropy from many foundations and non-profit organizations.

THE MILLENNIUM VILLAGE PROJECT

The MVP is a collaborative project of the Earth Institute at Columbia University, the United Nations Development Programme, the Millennium Promise, the FAO, the MDG Center for East and Central Africa, UNCDF, United Nations Children's Fund, and others. A scientific council composed of the UN Millennium Project and the Earth Institute, both of which are headed by Jeffrey Sachs, provides oversight of the MVP work. Millennium Promise, co-founded by Sachs and philanthropist Ray Chambers, is a supporting NGO. In September 2006, financier and philanthropist George Soros pledged $50 million to Millennium Promise to fund 33 Millennium Villages.

The MVP's objective is to end extreme poverty as well as address the problems of poverty, health, gender inequality, and disease. To achieve these objectives, the MVP advocates an integrated approach to rural development, which combines a cost-sharing and planning partnership with local and national governments, and rural, African communities. The focus is on capacity building and community empowerment. The MVP tries to ensure that communities living in extreme poverty have a real, sustainable opportunity to lift themselves out of the poverty trap. Inherent in this goal is the importance of improving access to clean water, focusing on sanitation, promoting education, improving food production, providing basic health care, and ensuring environmental sustainability. In a sense, the MVP is an expression of the MDG global strategy.[23]

The MVP was promoted by the economist Jeffrey Sachs. He came to see the limitations of free enterprise aid in dealing with the problems of the poorest of the poor. To avoid the unreliability of government sources, funding for each village is

committed in advance from the United Nations and private sources, often including a university, and one or more nongovernmental organizations (NGOs). Each Millennium Village is selected for its suitability for a five-year partnership in a development team. In a village, the MDGs are implemented in a five-year, five-step sequence, with specific amounts of funding year by year. Minimum standards for each village are specified, including, for example, a clinic-cum-dispensary-cum-first-aid-station in the village or a clinic that is reachable in a one-day roundtrip walk.

By the end of 2008, the MVP had been operating for almost four years. During this time, more than 400,000 individuals in 14 sites in 10 countries had been served. By the end of 2009, 80 Millennium Villages were created, each with 5,000 residents. More than half of the facilities have three nurses and a clinical officer (nurse practitioner or physician) on staff. A nurse or midwife provides four prenatal visits to expectant mothers. The clinics, in turn, have collaborating district hospitals within reach of ambulances (four-wheel, motorcycle, or pedal-powered). By the end of 2008, 51 clinics were operating, serving 300,000 outpatients. One could make an argument that the rapid success of this venture, including the level of support that they have received from donors, and the recipient countries compensates for the disappointing pace of the government-supported MDGs.

The long-term vision is for each Millennium Village to seed itself, spawning offshoots in the surrounding villages. For example, the government of Mali presented a plan to scale up the MVP approach to 166 of its most vulnerable communities covering approximately 2 million people. Based on its experience, the government of Nigeria is proposing a national MDG scale-up effort in 111 local government administrations, reaching approximately 20 million people. In Uganda, the Ministry of Health has expressed strong interest in scaling up across the country the midwife program launched in the Millennium Village site of Ruhiira.

MICROCREDIT AND THE GRAMEEN FAMILY OF SOCIAL ENTERPRISES

Microcredit and the Grameen family programs have significantly helped alleviate poverty and hunger in resource-poor areas. This program was established by a $79 gesture made by Muhammad Yunus to a small village in Bangladesh. This small microloan mushroomed into a multibillion dollar enterprise spanning 37 countries. Its success was highlighted by the organization being awarded the 2006 Nobel Peace Prize, which was accepted by Yunus on behalf of the seven million coowners (borrowers) of the Grameen Bank, nine of whom came with him to Oslo to receive the prize. The trust Yunus inspires has opened door after door, branching into a labyrinth of associations. Worldwide, many new ventures have emulated the theme of borrowing from the rich to lend to the poor. Serving a variety of projects, they have in common seed-loans with a business plan that become self-sustaining.[24]

The Grameen Community Development Bank is the microloan segment of the family, and it is now dwarfed by an increasing number of Grameen Family Social Businesses. The Grameen initiatives include some 14 mammoth enterprises crafted specifically to benefit the poor: risk capital for small to medium-size business; the largest telephone company in Bangladesh; software and hardware networking; village aquaculture and dairy; renewable energy in remote regions; education loans

for literacy and technology; providing nutritious food at near cost; poverty alleviation for the working poor; and the Grameen Trust, which exists to fund social enterprises worldwide, in 37 countries to date.[25] Almost half of these initiatives are for-profit enterprises that invest their profits in growth and provide their services at cost. Each will, in the end, be owned by the beneficiaries.

The following is a case study to illustrate how the initiative works. In Bangladesh villages, the Grameen Danone Corporation sells a yogurt called Shakti Doi "power yogurt" priced at a competitive $0.05 per 100 milliliters. It sells well. Being sweeter and richer than most local yogurts, it appeals to children while it provides protein and micronutrient requirements. Typically a "Grameen lady," funded initially by a Grameen loan, works a four-hour shift daily to sell 100 or more cups of yogurt to wholesale or retail customers. Local farmers with perhaps seven cows each provide milk to the factory. Several hundred farming and distribution jobs are created locally by each factory. The enterprise started in 2006 when Danone, a giant dairy company, approached Yunus to see how it could help. Within an hour, they had reached a $1 million agreement to manufacture and distribute fortified dairy products in Bangladesh. Profits would not be returned to those providing the funding. Instead they would be reinvested to bring health through food to the people of Bangladesh. In 2007, Grameen Danone launched a mutual fund to raise $135 million (paying an interest rate of 3 to 4 percent) to finance a second facility to generate 3,000 tons of dairy products annually. The business plan calls for nine more factories and other social businesses to fight malnutrition and poverty. The funders eventually will recoup their initial capital, but they will not receive any additional monies. The stock exchange on December 17, 2009, approved the second Grameen Bank Mutual Fund, the largest mutual fund the Dhaka stock exchange has ever approved.

Some critics argue that the poor are hard-pressed enough in seeking salaried employment and that it is asking too much to set them on a path to entrepreneurship for which they may have neither the risk tolerance nor the aptitude. Others point out that the poor are already entrepreneurs for whom meeting the day's needs is a challenge. In a softer stance for the extreme poor, Grameen has features applicable to very poor communities, including loans to beggars with no payback date. Some critics hold that it should be possible to invest money for profit, but at a reduced profit with the balance going to charity. Overall, many believe that microcredit is an idea whose time has come. It has had a visible impact on the economy of Bangladesh. As of 2008, 4 million poor have been helped and 1 million microloans have been generated with almost no defaults. (See the Grameen Foundation Report.[25])

AN UNPRECEDENTED OUTPOURING OF PHILANTHROPY

In terms of overseas development aid, individual contributions usually are dwarfed by the sums that governments provide. In recent years, however, a cascade of extremely large donations have come in from the superrich, "the Kings of Philanthropy." *Forbes* magazine identified 14 of the world's 793 billionaires who have each given away more than $1 billion.[26] Some are young and accompany the donation with a passionate idealism to invest in a cause. Some of the goals may seem idiosyncratic, but it is churlish to complain about such overflowing generosity. The *Forbes* survey covers only individuals foundations, not families or institutions, and

does not include many other worthwhile charity foundations such as the Ford Foundation, the Rockefeller Fund, and the Clinton and the Carter Foundations, to name a few. (The Foundation Center lists the U.S. sources of funding by asset size.[27]) Among the philanthropies that relate to global health, nutrition, and development, the Bill and Melinda Gates Foundation is arguably not only the largest, but also the most influential. As such, it deserves special attention. To those engaged with global health, whether academics seeking to mitigate inequities or those experiencing inequities, the priorities of the Gates Foundation are important. A major priority is laid out in a webcast: "Living Proof: Why We Are Impatient Optimists" and its accompanying Web site.[28]

In sum, there are many examples of synergism between the MDGs, the Millennium Villages, and the microcredit initiatives. The Gates Foundation provides general funding for MVPs. It also supports microfinancing with $38 million in grants to make financial tools more available in the developing world.

EFFORTS AND INITIATIVES TO ALLEVIATE HUNGER AND MALNUTRITION

Addressing Malnutrition

While the above discussion focused on the interplay of hunger, malnutrition, and economics, it is important not to lose sight of what malnutrition is and its negative effects on health and well-being. The World Health Organization cites malnutrition as the gravest single threat to the world's public health, and improving nutrition is widely regarded as the most effective form of aid. Millions of children around the world, primarily in the world's poorest countries, die from *protein-energy undernutrition* (previously called *protein-energy malnutrition*), which is an energy deficit due to chronic deficiency of all macronutrients. Types include Kwashiorkor (protein malnutrition predominant) and marasmus (deficiency in both calorie and protein nutrition). One million individuals a year die from *micronutrient deficiencies*, especially deficiencies in vitamin A, folic acid, iodine, iron, and zinc. Zinc, in particular, holds promise primarily because it has been shown to blunt the potentially devastating effects of diarrhea, a disease that kills more children than either malaria or AIDS.[29]

Many micronutrient treatment initiatives are available, and they do not require much money to make a difference.[30] Oral rehydration solutes (ORS) have been shown to be a huge success in stemming the number of deaths from protein-energy undernutrition due to diarrheal dehydration. Simple and inexpensive, ORS is astonishingly effective, saving an estimated 500,000 lives each year.[31] ORS come in pouches that simply require the addition of water (or a vegetable soup) to dissolve the salts. Where ORS is unavailable, 1 teaspoon of salt plus 8 teaspoons sugar dissolved in a liter of water can substitute. *Ready-to-use foods* (RTUF) are now used by many agencies for famine relief or are provided to the mother of a child in fragile nutritional status.[32] *Novel foods,* nutritionally adequate bread and cereals and fortified cookies, have been developed and provide essential nutrients so important for survival. Measures include providing deficient micronutrients through fortified sachet powders, such as peanut butter, or directly through supplements. Such products

include numerous vitamins and minerals required for optimal physical health and development.[33]

Micronutrient provision is one of the most cost-effective ways to improve global health. The Micronutrient Initiative and the Global Alliance for Improved Nutrition (GAIN) was created in 2002 at a Special Session of the UN General Assembly. The objective is to create a world without malnutrition by reducing malnutrition through food fortification and other sustainable strategies. GAIN encourages public-private partnerships and provides financial and technical support to deliver healthier foods and supplements to populations in need. Presently, innovative partnership projects in 26 countries reach more than 200 million people; more than half of these individuals are women and children. GAIN's Nutrition Program aims to reach 1 billion individuals.[34]

Advances in Appropriate Technology for Agriculture

Food shortages are caused mostly by the lack of technology needed for the higher yields found in modern agriculture, such as nitrogen fertilizers, pesticides, and irrigation. Many, if not most, farmers cannot afford fertilizer at market prices, leading to low agricultural production and wages and high, unaffordable food prices. Overfarming the land, too, has contributed to low yields. Although this topic is beyond the scope of this chapter, the message that needs to be conveyed is that soil impoverishment around the world has long been a problem to be resolved. Today we should learn from the techniques used in the past, such as *contour plowing*, which decreases rain runoff and mitigates soil loss to erosion; *terracing*, which brings previously unavailable land into productive use; and *crop rotation*, which helps impoverished soils to recover.

An innovation in more recent times has been the use of informally trained "barefoot" agronomists who learn about these tools on educational farms and then travel from region to region teaching others how to teach these techniques. South Africa, China, India, Bolivia, and other countries have used such advisors, to good effect, and these farming strategies are now widely practiced.[35] At another level, the African Centre for Crop Development at University of KwaZulu-Natal and Rockefeller University is a training program for doctoral students in Africa. Students from all over Africa train in selection and improvement of seed quality. They will return to rural areas to implement grassroots programs for seed improvement. The goal is to contribute to food sufficiency in Africa and eventually to a continental surplus.[36]

Safe Drinking Water

Of Earth's 6 billion people, 1 billion lack safe drinking water. Clean water and sanitation can do much to reduce disease. Providing safe drinking water and sanitation is part of the MDGs, but progress is slow. In India alone, contaminated water kills 300,000 children annually. Simple and relatively inexpensive solutions exist—for example, boiling, chemical treatment, or filtration. "Safe-water drinking straws," too, can be distributed to ensure that people will drink clean water. The straws are able to filter pathogenic bacteria from two liters of water a day over a two-year life. While much remains to be done to provide clean water to those in need, initiatives are under way in many parts of the world.

Water is a crucial commodity, especially in drought-prone regions. A number of technological solutions to extend water resources are available:

- *Subsurface drip irrigation* may seem like a low-cost substitute for watering cans, but it allows a given amount of water to be used many times more efficiently. The strategy has spread rapidly. One model is simply a drum on a pedestal with a one-quarter-inch-diameter pipe that is long enough to siphon water (often mixed with manure) to plants. The outlet end is pushed about two to five inches into the soil among the roots and a roughly metered amount of water is applied to where it diffuses and evaporates more slowly than the surface water. More complex (and more expensive) systems allow unattended drip irrigation over many acres.
- *Deficit irrigation* [37] involves supplying crops with far less than the amount of water needed to promote optimal growth. In regions where rainfall is intermittent but heavy when it comes, much is lost in runoff and evaporation.
- *Rainwater collection pits* are dug, perhaps 6 inches wide and 6 inches deep and 20 to 30 feet long, with rows of stones to direct runoff into them. These can be surfaced with foam sheets or pellets to deter evaporation, or plastic lining to deter seepage. When empty, the bottom remains damp and crops are planted there.
- *Low technology $25 treadle pumps* can irrigate a half acre, increasing crop yield by $100 per year compared with watering cans or bucket brigades. For the initial startup, four factories were set up to make treadle pumps, produce spare parts, and repair them; two years later there were 75 factories.
- "*Playground pumps*" are a novelty that has attracted multimillion dollar funding from U.S. sources. Several hundred have been installed in African locations. They consist of a merry-go-round that turns an underground pump that drives water into a high water tank. The PlayPump can lift 1,400 liters of water per day to an elevated 2,500-liter tank.[38]

Agrobiodiversity

Agrobiodiversity is a subset of general biodiversity, including all forms of life directly relevant to agriculture, such as crops and livestock, as well as many other organisms such as soil fauna, weeds, pests, and predators. It encompasses the variety and variability of animals, plants, and micro-organisms that are necessary for sustaining key functions of the agroecosystem, including its structure and processes for, and in support of, food production and food security.[39] An overview of the key roles of agrobiodiversity is beyond the scope of this chapter. Suffice it to say that potentially exciting initiatives are being developed, including drought-resistant seeds and hybrid rice, as examples.

Biofortification

The promise of selective breeding and biotechnology to yield crops with enhanced nutrient density remains elusive. Testing of novel foods is costly and onerous. High-lysine wheat has for decades been in the development phase. The problem was that high-lysine maize with a nutritional value reportedly approaching that of skim milk proved unpalatable. A more palatable variety was approved for food use in Japan, South Korea, Canada, Australia, and New Zealand; however, because of an EU presumption regarding risks associated with genetically modified foods,

production has been abandoned. A few promising initiatives are in progress, including carotene-enriched (orange fleshed) sweet potato that improves the vitamin A status of children.[40]

The 2009 World Food Prize was awarded to Dr. Gebisa Ejeta, an acclaimed plant breeder and geneticist who was born and grew up in rural Ethiopia.[41] Dr. Ejeta was awarded this honor for his major contributions in the production of sorghum, one of the world's five principal grains. He developed a high-yield sorghum that was drought resistant and weed resistant. This product has improved the food supply of hundreds of millions of people in Sub-Saharan Africa alone. This new and exciting area of research should yield further advances that could do much to alleviate hunger and malnutrition around the world.

Thanks to sponsorship from the MDG programs and Gates Foundation, the use of fortified foods is continuing to spread in the developing world. In Egypt, wheat flour is supplemented with folic acid, and iron is widely used in a staple flatbread. In South Africa, a similar program was initiated to supplement maize meal and wheat flour with folic acid. This initiative decreased folic acid–related birth defects by 30 percent. In China, soy sauce supplemented with iron has helped decrease the incidence of anemia. Table salt supplemented with iodine remains one of the most widespread and cost-effective means of saving lives. The list goes on. Supplements, fortifications, and the like have been shown to make substantial differences in the nutritional level of so many around the world.

EDUCATION INITIATIVES

Teaching materials are increasingly available online. The Global Health Educational Consortium, for example, provides no-cost modules on global health topics, including nutrition for developing countries.[42] In the developing world, Web-based education is flourishing as a result of technological advances and because it is cost-effective. For example, the University of South Africa now has more than 200,000 distance-learning students and it has satellites in four countries. In the area of nutrition education, programs utilize the computer as an intelligent assistant in the learning process. In 2002, a consortium of universities in the United States produced a set of computer modules for teaching nutrition to medical students.[43] The modules cover the full range of topics for medical students, including perinatal nutrition, pregnancy, and lactation. When adapted to needs of developing countries, the course will be invaluable especially if there is a shortage of skilled personnel. The FAO food security program e-learning courses and resources for trainers is also available online (see http://www.foodsecinfoaction.org/dl/dlcourselist_en.asp). Most certainly the scope and number of Web-based tools and modules will increase as this means of sharing information and teaching expands.

Green Energy

Green energy is the term used to describe sources of energy that are considered to be environmentally friendly and nonpolluting, such as geothermal, wind, solar, and hydro. The emphasis on green energy could yield cheap portable energy sources and wireless transmission of power. Low-cost energy has enormous implications for

Table 5.1
Eight Millennium Development Goals Related to Nutrition

Goal	Relationship of Nutrition to the Goal
1. Halve the percentage of extremely poor and those suffering from hunger	Poverty is the main determinant of hunger. In turn, malnutrition irreversibly compromises physical and cognitive development and thus transmits poverty and hunger to succeeding generations.
2. Achieve universal primary education	Malnutrition diminishes the chance that a child will go to school, stay in school, or perform well in school.
3. Promote gender equality, empower women	Women's malnutrition impairs the whole family's health and nutrition, in pregnancy, perinatal health, child-rearing, and contribution to family income.
4. Reduce by two-thirds the under-5 mortality rate	Delivery of a live healthy child is dependent, above all, on a well-nourished mother. Protein and folic acid are critical here.
5. Reduce by two-thirds risk factors for maternal health	Malnutrition accentuates all major risk factors for maternal mortality. NB protein, iron, iodine, vitamin A, and calcium intake deficiencies must be addressed.
6. Combat serious infectious diseases	Malnutrition aggravates infections, weakening defense systems. Loss of immune competence increases transmission/mortality in HIV, malaria, and tuberculosis.
7. Global partnership for development	Agricultural and economic development leads to improved nutritional status.
8. Mitigate impact of global warming	Climatic deterioration directly affects food production.

Source: Adapted from *Millennium Development Goals: Progress towards the Health-Related Millennium Development Goals.* www.who.int/mediacentre/factsheets/fs290/en/index.html.

nutrition, especially in areas with a short growing season. In the developed world, there is already a striking proliferation of greenhouses. With inexpensive energy, greenhouses become feasible even in the developing world and colder climates. Year-round inexpensive local food then becomes feasible.

CONCLUSION

Hunger is a global issue, and one that is not restricted to the poorest nations. A recent U.S. Department of Agriculture study showed that a record number of households had trouble getting sufficient food at one time or another in 2008. Even

before the recession began, more than two-thirds of families with children were having difficulty in providing adequate nutrition for their children. The number of people in households that lacked consistent access to adequate nutrition for their children rose to 49 million in 2008, or 13 million more than in the previous year and the most since the federal government began keeping statistics.[44]

Despite these depressing statistics, the response at the start of this century to address global hunger and malnutrition was an unprecedented collective commitment to action. There are compelling reasons for optimism, impatience, and indignation. History clearly has shown that reducing poverty, alleviating hunger and eliminating malnutrition is not simple and will take time. There is impatience that the pace is so slow. There have been repeated defaults and delays. Milestones have been missed, rescheduled, and missed again, while those who defaulted almost seem to be deliberately dragging their feet. There is indignation that millions will die undeserved, unnecessary deaths because the world has not given the issue of hunger the priority it deserves. UN Secretary General Ban Ki-moon's declared at the end of the 2009 Rome Food Summit: "Today, more than one billion people are hungry. Six million children die of hunger every year—17,000 every day. Urgent action is critical. In 2050, the world will need to feed 2 billion more mouths—9.1 billion in all." The hungry are aware that there is no worldwide shortage of food. The cost of an adequate diet for all is for less than one nation spends on warfare in a month. The Secretary General explained:

"Our job is not just to feed the hungry but to empower the hungry to feed themselves . . . The small-holder farmers are the heart and soul of food security and poverty reduction. A major goal is to provide them with seeds and land and offer them access to better markets and fairer trade."[45]

REFERENCES

1. Svedberg P. *Poverty and Undernutrition: Theory, Measurement, and Policy*. Oxford: Oxford University Press; 2002.
2. Chronic undernutrition among children: An indicator of poverty. FAO. Jan. 2004. www.fao.org. Accessed January 27, 2010.
3. Inter-agency Group for Child Mortality Estimation. Levels and trends of child mortality in 2006. http://www.childinfo.org/files/infant_child_mortality_2006.pdf. Accessed January 27, 2010.
4. United Nations Secretary-General/UN Development Group 2006. Millennium project. http://www.unmillenniumproject.org/goals/index.htm. Accessed January 24, 2010.
5. Collier P. *The Bottom Billion. Why the Poorest Countries Are Failing and What Can Be Done about it*. Oxford: Oxford University Press; 2007.
6. Crockford PM. Why is the third world the third world? Slide show and manual from Global Health Initiative University of Alberta. http://globalhealthedu.org/resources/Pages/WhyistheThirdWorldtheThirdWorld.aspx. Accessed January 15, 2010.
7. Muhammad Y. *Creating a World without Poverty: Social Business and the Future of Capitalism*. New York: Public Affairs Books; 2009.
8. Cotula L, et al. Land grab or development opportunity? Agricultural investment and international land deals in Africa. http://www.iied.org/pubs/display.php?o=12561IIED. Accessed January 27, 2010.

9. Buying farmland abroad: outsourcing's third wave. *The Economist*. May 23, 2009:61. http://www.economist.com/world/international/displaystory.cfm?story_id=13692889. Accessed September 15, 2009.
10. Rank MR, Hirschl TA. Estimating the risk of food stamp use and impoverishment during childhood. *Arch Pediatr Adolesc Med*. 2009;163(11):994–999.
11. Egeland GM, et al. Food insecurity among Inuit preschoolers: Nunavut Inuit child health survey, 2007–2008. *Can Med Assoc J*. 2010;182(3). http://www.cmaj.ca/papbyrecent.dtl. Accessed January 26, 2010.
12. World Diabetes Foundation. Diabetes in the developing world. http://www.worlddiabetesfoundation.org/media(7902,1033)/MEDIA_Diabetes_dev.09.pdf. Accessed January 28, 2010.
13. Transparency International. Annual global corruption perceptions index 2009. http://www.transparency.org/policy_research/surveys_indices/cpi/2009. Accessed January 27, 2010.
14. Extractive Industry Transparency Index International. Membership list 2010. http://eitransparency.org/. Accessed January 27, 2010.
15. Trinity College, Dublin. Trends in agricultural aid. http://www.tcd.ie/iiis/policycoherence/index.php/iiis/development_cooperation/trends_in_agricultural_aid. Accessed January 27, 2010.
16. Global Issues. US and foreign aid assistance: 2008 data. http://www.globalissues.org/article/35/us-and-foreign-aid-assistance. Accessed January 28, 2010.
17. International Development Statistics. IDS online databases on aid and other resource flows: 2008 data. http://www.oecd.org/dac/stats/idsonline. Accessed January 27, 2010.
18. Moyo D. *Dead Aid: Why Aid Is Not Working and How There Is a Better Way for Africa*. New York: Penguin Books; 2009.
19. Food and Agriculture Organization of the United Nations. Undernourishment around the world. In: FAO, ed. *The State of Food Insecurity in the World 2009*. Rome, Italy: Publications Division of FAO; 2009:Chapter 1. http://www.fao.org/docrep/012/i0876e/i0876e00.htm. Accessed January 5, 2010.
20. The World Bank. 2009 world development indicators world view. http://siteresources.worldbank.org/DATASTATISTICS/Resources/wdi09introch1.pdf. Accessed December 5, 2010.
21. The World Bank. Global and regional poverty trends. In: *Global Economic Prospects 2009 World Bank Data on Poverty and Inequality*. http://search.worldbank.org/all?qterm=globalpercent20economicpercent20prospectspercent202009. Accessed January 15, 2010.
22. US Department of Agriculture. Economics research service data sets. http://www.ers.usda.gov/Data/Macroeconomics/. Accessed January 28, 2010.
23. The Earth Institute, Columbia University. The Millennium Villages Project annual report January 1–December 31, 2008. http://www.millenniumvillages.org/docs/MVP_Annual_Report_2008.pdf. Accessed January 14, 2010.
24. Grameen Foundation. Combining the power of microfinance and technology to defeat global poverty. http://www.grameenfoundation.org/. Accessed January 27, 2010.
25. Grameen Foundation. Our impact. http://www.grameenfoundation.org/our-impact. Accessed January 27, 2010.
26. Whelan D, et al. Billion dollar donors. *Forbes Magazine*. August 24, 2009. www.forbes.com. Accessed Jaunary 28, 2010.
27. The Foundation Center. http://www.foundationcenter.org. Accessed January 27, 2010.

28. Bill and Melinda Gates Foundation. Progress towards nutrition: winning the fight against malnutrition. http://www.gatesfoundation.org/livingproofproject/Documents/progress-towards-nutrition.pdf. Accessed January 28, 2010.
29. Walt, V. Can one pill tame the illness no one wants to talk about? *Time Magazine*. August 17, 2009. www.time.com/time/magazine/article/0,9171,1914655,00.html.
30. Micronutrient Initiative. MI around the world. http://www.micronutrient.org/English/view.asp?x=554. Accessed January 28, 2010.
31. Drucker G. ORT 10 years after: 500,000 lives saved each year. *Health News*. 1988;9(6):2.
32. Manary M, et al. Home based therapy for severe malnutrition with ready-to-use food. *Arch Dis Child*. 2004;89(6):557–561.
33. Anderson T. Firms target nutrition for the poor. BBC News. http://news.bbc.co.uk/2/hi/business/8114750.stm. Accessed January 27, 2010.
34. Global Alliance for Improved Nutrition. http://www.gainhealth.org/. Accessed January 28, 2010.
35. Pender J. Agricultural technology choices for poor farmers in less-favored areas of South and East Asia. Washington, DC: International Food Policy Research Institute, Environmental and Production Technology Division; 2008. http://www.ifad.org/operations/projects/regions/pi/paper/5.pdf. International Fund for Agricultural Development, Asia and the Pacific Division Discussion paper, No.5. Accessed January 28, 2010.
36. African Centre for Crop Development. Global partnership initiative for plant breeding capacity building. http://km.fao.org/gipb_old/index.php?option=com_content&task=view&id=271&Itemid=339. Accessed January 27, 2010.
37. Yenesew M, Tilahun K. Yield and water-use efficiency of deficit-irrigated maize in a semi-arid region of Ethiopia. *Afr J Food Agric Nutr Dev*. 2009;9(8):1635–1651.
38. Watson S. How play-pump works. http://science.howstuffworks.com/earth/green-technology/sustainable/community/playpump.htm. Accessed January 27, 2010.
39. Food and Agriculture Organization of the United Nations. http://www.fao.org/docrop.htm. Accessed January 28, 2010.
40. World Health Organization. WHO vitamin and mineral nutrition information system. http://www.who.int/vmnis/iodine/status/en/index.html.
41. The World Food Prize. http://www.worldfoodprize.org. Accessed January 28, 2010.
42. Global Health Education Consortium. http://www.globalhealtheducation.org. Accessed January 28, 2010.
43. Lindell KC, et al. The evolution of nutrition in medicine, a computer curriculum. *Am J Clin Nutr*. 2006;83:956–962.
44. US Department of Agriculture. Household food security in the United States, 2008. Washington, DC: USDA Economic Research Service; 2009. Economic Research Report No.(ERR-83): 66.
45. Ban K., Secretary-General, United Nations. Speech closing the Rome Food Summit November 19, 2009. U.N. chief: hunger kills 17,000 kids daily. CNN. http://edition.cnn.com/2009/world/europe/11/17/italy.food.summit. Accessed January 29, 2010.

This page intentionally left blank

CHAPTER 6

Essential Medicines for Everybody in the World: The World Health Organization Essential Medicines Program

Marcus M. Reidenberg, MD

No matter how one defines "health," prevention and treatment of disease is an essential prerequisite for achieving health and well-being. Implicit in this is that the right medicine be available at the time and place of need. In 1977, the World Health Organization (WHO) initiated an essential medicines and pharmaceutical policies program to advise member states on improving the availability of selected medicines for their populations. The intent was to ensure that people everywhere in the world have access to essential medicines; that the medicines would be safe, effective, and of assured quality; and that the medicines would be prescribed and used rationally. Of issue to most nations, resource-poor countries in particular, was how best to select from a very large number of prescription drugs on the market and how to afford to purchase the medicines for their respective national health services. The focus of this chapter is to provide the reader with a brief historical overview of the WHO Essential Medicines Program.

For much of the 20th century, many prescribed drugs were actually ineffective for their indicated purpose, and no regulations were in place to mandate proof of the safety of drugs before they could be sold on the open market. In the United States, it was not until the Kefauver-Harris Act of 1962 (Drug Efficacy Amendment to the 1938 Food, Drug, and Cosmetic Act) that evidence of efficacy as well as safety was required for marketing approval for a prescribed drug. The Kefauver-Harris Act introduced a requirement for drug manufacturers to provide proof of the effectiveness and safety of their drugs before Food and Drug Administration (FDA) approval. The act further required drug advertising to disclose accurate information about side effects, and stopped the practice of cheap generic drugs being marketed as expensive drugs under new trade names as new "breakthrough" medications. It introduced a "proof-of-efficacy" requirement that was not present before the act. The act was passed in response to the thalidomide tragedy in which thousands of children were born with birth defects as a result of their mothers taking thalidomide for morning sickness during pregnancy.

While this act was a milestone in efforts to ensure that effective and safe medications were being marketed to the public, a comprehensive review was initiated several years later to identify the extent of ineffective drugs being marketed. Findings showed that approximately one-third of the 3,000 drugs being prescribed in the United States were not effective.[1] Advertising and promotion contributed to the widespread prescribing of these drugs. If this situation could exist in the United States even with supposed safeguards in place, one could easily imagine the state of the pharmaceutical market in other countries, especially those with limited resources for medical care, a poorly educated population, and little or nonexistent means to inform and educate physicians about drug efficacy. Indeed, this was the state of much of the world before the WHO took an active role in this issue.

In 1975, the director general of the WHO presented a report to its governing body describing the drug-prescribing problems facing so many nations, especially those with limited resources. At that time, the concept of a national drug policy was unknown to most countries and the rational use of medicines was extremely limited. Using the experience of some Scandinavian countries, where programs supplying basic or essential drugs had been implemented, the director general sought to introduce a worldwide program whose purpose would be to provide a structure to make the most necessary drugs available to populations whose basic health needs were not being met. He proposed that the WHO develop a program to advise member states about the selection and use of essential drugs to help address the problem of limited access to medicines. His proposal, the Essential Medicines List, was created, adopted, and implemented in 1977.

Even before the Essential Medicines List was created, the first step was to define what an "essential drug" was. The agreed-on definition was that an "essential drug" was a drug needed to satisfy the health needs of the majority of the population. The concept of an essential drugs program included not only the definition of an essential drug, but also the idea that essential drugs should be available at all times, in adequate amounts, and in the appropriate dosage forms. An international expert committee was appointed by WHO to describe how a national essential drugs program could be implemented.[2] The committee's criteria for determining whether a drug was essential included the following:

1. The drug should have adequate evidence of efficacy and safety based on clinical studies in a variety of medical settings;
2. The drug should be available in a form that is made properly with adequate bioavailability and stability under anticipated storage conditions; and
3. The drug should have its indicated purpose well established.

When several drugs used for the same indication meet these criteria, the drug that appears "best" would be selected. Cost considerations would be taken into account. The first expert committee then used these criteria to develop a Model List of Essential Drugs, illustrating how these ideas could be implemented in practice. This list contained 186 different medicines.[3] The most recent list includes more than 300 drug products for approximately 250 different chemical entities.[4] The expert committee expressed its view that lists of essential drugs should be drawn up locally because of differences between countries in disease prevalence and other factors.

The committee also stated that the selection of drugs is a continuing process, taking into account changing disease patterns as well as new medicines. The WHO Action Program on Essential Drugs was established as the focal point for investigations into the methods of implementation of the essential drug concept to meet unsatisfied basic health needs of populations that did not have essential medicines.

The advantages of having a limited list of high-priority drugs include (1) the ability to keep ineffective or toxic compounds out of common use; (2) the ability to choose the most cost-effective drug among therapeutically equivalent products; (3) more efficient management of a limited pharmacy inventory; (4) the need to make drug information available only for a defined, limited number of drugs; and (5) facilitating the detection of counterfeit drugs entering the supply system when there are fewer drug products to monitor.[5] Additional benefits have been suggested (for example, fewer products may lead to less waste or larger orders of these products may get bulk discounts).[6]

The disadvantages of having a limited list of essential drugs are (1) delayed inclusion of new medicines until the group selecting the essential drugs meets to consider additions to the list; (2) limitation of doctors' ability to prescribe drugs not on the specific essential medicines list; and (3) a potential impact on new drug research and development (R&D) programs if the sponsor has concerns that a new medicine, when developed, will not have a market because of its exclusion from the list.

When resources are constrained, the benefits outweigh the disadvantages, especially because the first two disadvantages can be easily addressed:

1. Delayed inclusion can be prevented through frequent updating of the list.
2. An essential drugs list need not be overly restrictive in limiting doctors' ability to prescribe. Different patients may respond differently to drugs that are very similar and have similar average responses in clinical trials. Nonsteroidal anti-inflammatory drugs are an example of this phenomenon. Several similar drugs can be on a single essential drugs list to enable individualization of therapy for these types of drugs when resources permit. The ability to prescribe unlisted but available drugs can be addressed administratively by the organization administering the program.
3. The issue of potentially suppressing new drug research and development requires substantial thought and new approaches to funding R&D of medicines. One such approach is delinking R&D funding from the sales revenue of the resultant product. One idea is to award prizes for achievement of R&D milestones in exchange for putting the intellectual property developed by the R&D into the public domain.[7] Another idea is the development of patient pools to facilitate the development of pediatric formulations and fixed dose combination products of drugs to treat HIV/AIDS.[8] Much more innovation is required for ways to fund R&D for needed medicines when the expected market is too small to support the R&D costs.

The essential drugs concept of purchasing a limited list of essential drugs for a health service and making them generally available has been accepted by 156 countries and most, if not all, donor organizations. Although public evaluations of how well this concept is actually implemented by countries are few, the consensus is that the benefits of a good essential medicines program exceed the harms. Of the evaluations that do exist, evidence suggests that implementation is occurring.[3, 9, 10]

One issue with respect to implementation is whether the Model List should be used as an actual purchasing list, as is the case in some organizations. The original intent of the Model List was to be an example of how the concept of a limited formulary could be implemented. It was expected that each health service would develop its own specific list of drugs for purchase based on local needs. From the beginning, many drugs were listed as examples of a therapeutic class from which other drugs in the class could be selected for actual purchase. Current examples of these include lidocaine for local anesthesia, prednisolone for the systemic steroid, and hydrochlorothiazide for a diuretic. It was expected that if local conditions or cost differentials made a different therapeutically equivalent drug a better choice for a specific formulary, that drug, rather than the one named in the Model List, would be purchased for that formulary.

United Nations and donor organizations such as United Nations Children's Fund (UNICEF), *Médecins Sans Frontières* (Doctors Without Borders), and religious groups have accepted the concept of essential drugs. Some of these organizations even restrict their purchases to drugs included on the Model List. The rationale is that by selecting drugs from the list, it helps these organizations to select the most important drugs to donate. It also helps these organizations handle commercial pressure that they sometimes face to add drugs for donation that are not necessarily essential.

A recent effect of having a medicine on the Model List is that the determination that a medicine is essential influences the international discussion about the cost of the drug and its affordability and availability to all who need it. But it also means that if a drug is not on the Model List, it will not be purchased by anyone using the Model List as a purchasing list. As new drugs are brought to market, it is necessary to update the Model List to make appropriate additions and subtractions. An expert committee meets every two years to revise the Model List. The potential for up to a two-year lag in adding to or deleting from the Model List is recognized as a limitation that needs to be addressed.

Members of each expert committee are highly qualified health care professionals who are selected for geographic diversity as well as for specialties relevant to the discussion of medicine policy and the selection of essential medicines. Committee members and advisors include physicians, pharmacologists, pharmacists, nurses, and academics, such as epidemiologists and biostatisticians. The listing of the committee membership can be found in the WHO technical reports written by the expert committee at http://www.who.int/medicines/publications/essentialmeds_committeereports/en/index.html. The last four expert committees included 7–10 members and 5–8 temporary advisors in addition to WHO staff. The temporary advisors brought perspectives to the meeting that were not necessarily represented by the members. A clinical pharmacologist chaired each of the last four meetings. Decisions were made by consensus based on the members' interpretations of the evidence in the application, the written reviews and comments about the application posted on the WHO Web site, and additional evidence presented at the meeting.

As this WHO program has evolved, various questions of a broader nature than whether to add a specific drug to the Model List have been discussed at the expert committee meetings. Over a period of several meetings, the question of how to address the problem of antimicrobial resistance was raised and discussed. A solution proposed at the time was to urge restrictions on the use of selected antibiotics

so that they would only be used to treat patients with infections resistant to the standard antibiotics.

The expert committee has recognized that diseases such as malaria, tuberculosis, and HIV should be treated with a combination of antimicrobial drugs given simultaneously. Patients adhere to their treatment regimens better when they take fewer tablets less frequently. The appropriate combination of medications in a single oral dosage form would be a therapeutic advance. Such a combination tablet would ensure that antimicrobial drugs would not be used singly for these diseases, a practice leading to antimicrobial resistance. Because fixed-dose combination drug products often do not exist for the recommended combination of drugs, new product development is required. Often, innovation is needed to make a proper stable dosage form in which the rate and extent of absorption of each of the drugs in the combination must be the same as when these drugs are given separately. Thus, the need for appropriate fixed-dose combination-drug products expressed by the WHO can stimulate new product development.

Another current issue is related to the definition of essential drugs, a limited number of well-chosen drugs to satisfy the needs of the majority of the population. This leaves out drugs needed to treat rare diseases. The 2006 expert committee, for example, discussed whether or not drugs for rare diseases are "essential." A suggestion was made for a separate Model List of orphan medicines.[11] Implementing this suggestion requires an answer to the question: how many drugs should be on the Orphan Drug List? More than 6,000 rare diseases and related conditions are listed on the National Institutes of Health Web site.[12] It is unclear how having an extensive list of orphan drugs would aid a WHO member state wanting advice on what to purchase for its health service.

Excluding drugs for orphan diseases from the Model List raises the issue of whether distributive justice, the proper distribution of benefits and burdens, is being carried out. Is it right for one patient to benefit from a health service and for another to be ignored only because of the prevalence of their disease? A better way to determine which medicines to purchase is to use cost-effectiveness analysis to evaluate a drug for a rare disease and compare the result of its analysis with that of similar analyses of drugs for diseases already on the Model List. If the cost-effectiveness of the drug for the rare disease is similar to that of drugs on the Model List, it should be considered "essential" and added to the Model List.[13]

The need to address price openly in determining whether a drug is essential or not became apparent with the development of effective and expensive drugs for HIV infection. Previously, price had been considered in an informal and covert way. Recommending purchase of a drug that a country could not afford did not seem to make sense. Until 2002, there was a perception among recipient countries that the Model List would only contain cheap generic medicines that everyone could afford. Consideration of the anti-HIV drugs led to the view that if a drug was medically essential, a high price should not exclude it from the list. Declaring that these drugs for HIV were essential helped change international opinion and led to substantial price reductions for these drugs for the public sector of low- and middle-income countries. Thus, the Model List serves a function of identifying medicines that are essential but beyond the means of many countries. The formal identification that these drugs are essential may help countries obtain the drugs at affordable prices.

Another consideration in preparing a limited list of drugs is cost compared with effectiveness. The Essential Medicines Program helps countries set priorities for what drugs to purchase with limited resources. Cost-effectiveness analysis has been part of the process of determining which drug in a class of drugs should be the example used for the Model List. With the anti-HIV drugs, price was explicitly excluded as a factor for inclusion. Many new drugs now coming to market, especially in the antineoplastic category, are quite costly and have variable degrees of efficacy. As such, it has become imperative to consider cost and effectiveness in determining which, if any, of these drugs should be on a limited list of drugs available in resource-constrained countries. Incorporating rigorous and valid cost-effectiveness analysis into decision making is a major task for the future. Rigorous consideration of cost-effectiveness requires the development of appropriate methodology for use by low- and middle-income countries and the development of professionals capable of using this methodology properly to contribute to decision making.

The lack of pediatric formulations on the Model List for drugs that are appropriate for children also needed to be addressed. A joint WHO-UNICEF meeting was held in August 2006 and the recommendation was that the WHO add to the Model List any missing essential medicines for childhood diseases and pediatric formulations of medicines appropriate for children.[14] Several specific issues were raised at that meeting, including that only some drugs needed by children had pediatric formulations available in countries with rigorously regulated pharmaceutical industries. Registration in these countries enabled the quality of the drug product to be ensured, and these products could be recommended for purchase. Other medicines were registered for use in children, but no formulations were identified in any reference country. Still other medicines on the Model List are appropriate for children, but not registered for children, and no formulations exist in any reference country.

Addressing all of these issues was a large task and provided the impetus for the WHO expert committee to form a subcommittee to design a model list of medicines essential for children. This was published in the expert committee report for 2009.[15] Furthermore, a five-year WHO program, Make Medicines Child-Size, was started in 2007 to implement the recommendation of the subcommittee.[16] Success in getting registration and formulations of drugs needed for children not currently marketed will certainly help children in all countries, not just poor children in countries with limited resources.

Any comprehensive medicine program requires information about drugs in addition to the drugs themselves. A manual containing information about the drugs on the Model List, the WHO Model Formulary, already exists. The WHO is in the process of developing new guidelines and revising old treatment guidelines to be fully in accord with the idea of using evidence-based recommendations. The WHO has additional publications designed to advise member states about ways to improve therapeutics. These are available through the Essential Medicines Library at http://apps.who.int/emlib/. The full structure and activities of the WHO Department of Medicines Policy and Standards is available at http://www.who.int/medicines/about/psm_contact/en/index.html. Users can get information about the various programs and publications of the Medicines Department from its homepage (http://www.who.int/medicines/en/).

Additions and deletions of medicines to the Model List are made at the biennial meeting of an expert committee. These changes are based on applications

submitted to WHO by any interested party. The applications must give an evidence-based justification for inclusion or deletion of the medicine. Applications are posted on the WHO Medicines Department Web site for comment from the public and are reviewed by selected members of the expert committee and WHO program staff. At the meeting, each application is discussed and a decision about whether to add the drug to the Model List is made by the committee. A limitation of this process is that the committee only considers drug products for which applications have been made. Overcoming this limitation requires that those interested in the issue should take a proactive stance in reviewing the current Model List on the WHO Medicines Department Web site and submit applications for additions and deletions to the Model List, which would be considered at the next expert committee meeting.

Hogerzeil [17] of WHO has written about how the concept of essential medicines, although developed more than 30 years ago to help resource-poor countries, can now benefit all countries. He, along with others, identified international treaties or covenants that identify a legal right to the provision of essential drugs to citizens of any country that has ratified that treaty or covenant. One that is specifically related to provision of essential medicines is the International Covenant on Economic, Social, and Cultural Rights (ICESCR) of 1966, which has been accepted by 156 countries. The official interpretation of article 12 states that the right to medical services stated in that covenant includes the provision of essential medicines as defined in the WHO Model List of Essential Drugs. Legal cases have been identified in which courts have recognized the right to essential medicines, with the court directing the government to fulfill its obligation to its citizens under its treaty or covenant obligations.[18] This opens a new pathway to making essential drugs "available at all times in adequate amounts and in the appropriate dosage forms."

In summary, 30 years after the WHO initiated the Essential Medicines Model List, four out of every five countries have adopted a national essential medicines list. More than 100 countries have a national drug policy in place or under development. Furthermore, a network of 83 countries provides global monitoring for adverse drug reactions and potential safety problems. Regarding pricing, 30 years ago, there was virtually no publicly available price information and few countries actively encouraged generic substitutions. In the 21st century, at least 33 countries provide such information.

The essential medicines idea was developed to help limited-resource countries make choices to use their medical resources for the greatest good for the greatest number. It has evolved into a concept, as relevant in 2010 as it was in 1977, to help all countries and societies provide the best medical care possible to all of its citizens.

REFERENCES

1. Center for Drug Evaluation and Research of the U.S. Food and Drug Administration. A brief history of the center for drug evaluation and research. http://fda.gov/cder/about/history/Page37.htm.
2. World Health Organization, Expert Committee. *The Selection and Use of Essential Medicines*. Geneva, Switzerland: World Health Organization; 1977. WHO technical report Series, No.615.

3. Laing R, Waning B, Gray A, Ford N, 't Hoen E. 25 years of the WHO essential medicines lists: progress and challenges. *Lancet.* 2003;17;361(9370):1723–1729.
4. 16th WHO model list of essential medicines. http://www.who.int/selection_medicines/committees/expert/17/WEB_unedited_16th_LIST.pdf. Published March 2009. Accessed August 12, 2009.
5. Reidenberg MM, Bradley AC. Counterfeit and substandard drugs. *Clin Pharmacol Ther.* 2001;69:89–193.
6. Reidenberg MM. Essential drugs and the WHO model list: addressing new issues. *Clin Pharmacol Ther.* 1996;59:251–257.
7. Travis J. Research funding: prizes eyed to spur medical innovation. *Science.* 2008; 319(5864):713.
8. Doctors Without Borders. Taking the plunge: how a "patient pool" could help solve the access to medicines crisis. http://doctorswithoutborders.org/news/article.cfm?id=2870&cat=field-news. Published July 31, 2008. Accessed August 12, 2009.
9. Levy M, Reidenberg, MM. What has been the impact of the concept of essential drugs? *Clin Pharmacol Ther*. 2003;73:275–278.
10. Mendis S, et al. The availability and affordability of selected essential medicines for chronic diseases in six low- and middle-income countries. *Bull World Health Organ*. 2007;85(4):279–288.
11. Stolk P, Willemen MJC, Leufkens HGM. "Rare essentials": drugs for rare diseases as essential medicines. *Bull World Health Organ*. 2006;84:745–751.
12. National Institutes of Health, Office of Rare Diseases. Information page. http://rarediseases.info.nih.gov.
13. Reidenberg MM. Are drugs for rare diseases "essential"? *Bull World Health Organ*. 2006;84:686.
14. World Health Organization/UNICEF. Essential medicines for children expert consultation. Meeting Report from joint WHO/UNICEF consultation on essential medicines for children; August 9–10, 2006; Geneva, Switzerland. www.who.int/medicines/publications/UNICEFConsultation.pdf.
15. World Health Organization, Expert Committee. Unedited draft report of the 17th expert committee on the selection and use of essential medicines. http://www.who.int/selection_medicines/committees/expert/17/WEBuneditedTRS_2009.pdf. Accessed July 24, 2009.
16. World Health Organization. WHO model list of essential medicines for children. http://www.who.int/childmedicines/en/index.html. Accessed July 24, 2009.
17. Hogerzeil HV. The concept of essential medicines: lessons for rich countries. *Br Med J*. 2004;329:1169–1172.
18. Hogerzeil HV, Samson M, Casanovas JV, Rahmani-Ocora L. Is access to essential medicines as part of the fulfillment of the right to health enforceable through the courts? *Lancet*. 2006;368:305–311.

SECTION 2

PUBLIC HEALTH PREVENTION AND SAFETY

This page intentionally left blank

CHAPTER 7

Getting Home Safe and Sound: Occupational Safety and Health Administration at 38

Michael Silverstein, MD, MPH

> *Society entrusts regulatory and enforcement agencies with awesome powers. They can impose economic penalties, place liens upon or seize property, limit business practices, suspend professional licenses, destroy livelihoods. . . . How regulatory and enforcement agencies use these powers fundamentally affects the nature and quality of life in a democracy.*
>
> —Malcolm Sparrow[1]

The Occupational Safety and Health Act of 1970 (OSHAct) [2] was adopted with high expectation during an era of significant social reform. Its statement of national purpose and provision of regulatory authority marked a departure from years of government indifference to workplace injury, illness, and death. And yet, 38 years after Congress established the Occupational Safety and Health Administration (OSHA) to implement the OSHAct and to "assure so far as possible every working man and woman in the Nation safe and healthful working conditions," a worker still becomes injured or sick from a dangerous job every 2.5 seconds [3] and a worker dies from a workplace injury or illness every eight minutes.[4]

DESPITE PROGRESS, STILL WORKING IN HARM'S WAY

Although it is difficult in the 21st century to find anyone who believes that the promise of the OSHAct has been met, workplaces generally have become safer since 1970. OSHA rules such as those for cotton dust, inorganic lead, and blood-borne pathogens have resulted in reduced exposures and illnesses.[5] Injury rates have decreased in workplaces in which there have been inspections and enforcement. [6, 7, 8, 9] OSHA has given unions tools that enhance their ability to get employers'

This chapter was originally published as Silverstein M, "Getting Home Safe and Sound: Occupational Safety and Health Administration at 38." *Am J Public Health*. 2008; 98(3): 416–423. Reprinted with permission from *The American Journal of Public Health,* Volume 98 ©2008 by *The American Journal of Public Health*.

attention. It has given safety and health professionals in the business community more credibility and leverage with upper management.

These improvements mask serious problems remarkably resistant to change. Although the national rates for workplace injuries and deaths dropped from 11 per 100 workers in 1973 to 4.6 per 100 in 2005, resulting at least in part from activities set in motion by the OSHAct,[10] the rate of more severe cases (with restricted or lost workdays) has stayed almost flat, declining only from 3.4 to 2.4 per 100.[11] Using state trauma registry data on severe occupational trauma, Friedman and Forst found that "in contrast to reports from national surveillance data sets, we do not observe a significant decline in occupational injuries between 1995 and 2003."[12]

Of the top-four causes of workplace fatalities, only one (homicides) has decreased in number in the past 15 years; the other three (falls, highway incidents, and being struck by objects) have increased.[13] Most injuries and illnesses are predictable and preventable, with a disproportionate burden on such high-risk workers as construction laborers, nurses' aids, and farmworkers. For example, the number of fatal work injuries among foreign-born Hispanic workers has nearly tripled in less than 15 years,[13] partly because the numbers of immigrants at work has been increasing and also because these workers are concentrated in dangerous jobs.

Twenty-first-century workers confront four types of workplace risks, each of which presents unmet challenges. First, many dangers, such as falls from roofs or amputations from unguarded machinery, were widely recognized in 1970 and should have been virtually eliminated years ago. For some, simple means of correction have been understood for thousands of years. Herodotus (c. 484–425 BCE) described the prevention of fatal trench collapses in the Persian Wars in 450 BCE,[14] yet every week American workers die in trenches inadequately sloped, shored, or shielded.

Second, workers face hazards present in 1970 but only fully appreciated more recently, such as the forceful exertions, repetitive movements, and awkward postures that cause the musculoskeletal disorders that account for more than 30 percent of all workers' compensation claims. Other examples include workplace violence and biohazards.

Third, some hazards have entered the workplace since 1970; these include food flavoring containing diacetyl, which cause bronchiolitis oblierans ("popcorn lung"),[15] and modern metalworking fluids that cause hypersensitivity pneumonitis.[16, 17]

Fourth, the changing political, economic, and legal landscape of work is creating potential new dangers. Global businesses, lean manufacturing, outsourcing, reduced pension security, an aging workforce, declining unionization, and changing immigration patterns have altered the experience of work. The breakdown of long-term employer-employee relationships, the increase in "independent" contracting, and the disappearance of traditional career paths have eroded job stability, leading to stresses whose consequences are not yet fully understood.[18, 19]

OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION'S FADING PROMISE

Although workplace dangers have evolved, the OSHAct has remained virtually unchanged for 38 years. Congress based it on conventional principles of state power in a democracy, blending long-established tools of regulation, enforcement, and technical assistance. Congress established a general duty for employers to provide work "free from recognized hazards that are causing or are likely to cause death or serious

physical harm."[20] It authorized OSHA to set additional specific rules and empowered the agency to secure compliance by exercising its police power, supplemented by assistance for voluntary compliance. Workers are provided the right to file complaints and accompany inspectors. States may assume jurisdiction if their programs are at least as effective as OSHA's.

The OSHA model presumes that market incentives to reduce hazards are so imperfect that the threat of citation and penalty is necessary for worker protection,[21] and that without federal oversight, state governments cannot ensure equal protection. This approach reduces inequities by making it difficult for unscrupulous employers to attain competitive advantages by saving on the costs of injury prevention.

Notwithstanding the fundamental soundness of design, the system has structural and political weaknesses that were exposed almost immediately by critics across the ideological spectrum.[22] The business community was angered by OSHA's first inspectors issuing thousands of citations for nonserious violations. Labor representatives were angered by OSHA's failure to identify and cite serious violations. The situation worsened when OSHA turned hundreds of obsolete and sometimes trivial consensus standards into enforceable rules, exposing the agency to public ridicule. Then, in a confidential memo that later became public, the agency's first assistant secretary, George Guenther, pledged "four more years of properly managed OSHA for use in the campaign" to reelect Richard Nixon, including "no highly controversial standards."[21]

Following this start, OSHA has never been free enough from political turmoil to address the limitations of the OSHAct itself. Twenty-five years after the act's passage, Kniesner and Leeth, writing for the politically conservative Cato Institute, described OSHA's shortcomings in language that could have been written by OSHA's more liberal supporters:

> OSHA inspectors frequently overlook dangerous working conditions, and even when they find serious health and safety violations, inspectors often cannot compel firms to eliminate the hazards discovered. To encourage timely compliance, administrators often slash assessed penalties, further reducing the already minor economic incentives for firms to observe health and safety standards. Firms realize that it is unlikely that they will be inspected . . . and if they are inspected, firms can avoid paying severe fines by simply agreeing to abide by OSHA's regulations in the future. . . . OSHA as originally conceived in 1970 is withering away.[23]

Even when OSHA has been especially visionary and creative in pushing the margins of the OSHAct beyond routine regulation and enforcement, these incentive efforts have failed to overcome legal and political barriers. OSHA's generic cancer policy was pulled back by the Reagan administration after the Supreme Court's decision to overturn the rule OSHA had established to regulate workplace exposure to benzene.[24] The update of hundreds of permissible exposure limits was invalidated by the 11th Circuit Court's requirement by substance-by-substance analysis.[25] OSHA's ergonomics standard was repealed under the Congressional Review Act,[26] and the Cooperative Compliance Program, which provided incentives for safety and health management programs in exchange for reduced inspection priority, was overturned in court.[27] Other examples of imaginative and promising OSHA programs whose impact and use have remained limited include instance-by-instance citations for egregious violations, corporate-wide settlement agreements, and economic protection for workers removed from lead-exposed jobs for medical reasons.

OSHA has become essentially irrelevant to most employers (although still vilified by many lobbyists). Infrequent OSHA inspections and small OSHA penalties provide little incentive for employers to pay attention to and comply with the agency's rules. Acts of willful disregard of OSHA rules leading to workplace deaths regularly escape meaningful punishment. Employees are discouraged from filing complaints because the OSHAct provides insufficient protections from discrimination. Nothing has better illustrated OSHA's impotence than the failure to use its authority to protect emergency response and recovery workers after the September 11, 2001, attack on the World Trade Center. Many of these workers now show evidence of preventable lung disease, caused at least in part by OSHA's decision not to enforce its rules requiring the use of respirators.[28, 29] The aftermath of the attack revealed other stubborn problems: public employees not covered by the OSHAct, overlapping and ambiguous government jurisdictions, thousands of immigrant and temporary employees without clearly defined employer-employee relationships, and a lack of rules for basic safety and health management programs, medical surveillance, or exposure assessment.

PROBLEMS WITH THE PARADIGM

Four barriers to OSHA's success are so inherently linked to the statutory design of regulation, inspection, and education that even the most skilled and determined administration cannot expect to overcome them.

First, OSHA enforcement, the agency's most demonstrably effective tool, falls short. Although evidence indicates that injuries decline in workplaces that have been inspected by OSHA,[6, 7, 8, 9, 30] only 1 percent of workplaces are inspected yearly and little evidence suggests that inspections have a ripple effect on similar, uninspected workplaces,[31, 32, 33, 34] although one study did find an industry-wide reduction in injuries from falls following a new fall protection rule.[35] This is hardly surprising given OSHA's infrequent inspections and consultation visits and the small penalties it issues, with those for serious violations (involving substantial probability of death or serious harm) averaging under $900 in 2005.[36]

Even if the penalties had stimulated businesses to comply with OSHA rules, the impact on injury rates would have been limited because of the poor match between these rules and the causes of workplace fatalities and injuries.[37] In the construction industry, 31 percent of injuries in 1994 involved workers being struck by objects, but only 4 percent of OSHA citations that year were for violations related to this hazard. In 2005, assaults and violent acts were responsible for 14 percent of workplace deaths, but there are no OSHA rules addressing the causes of workplace violence. In addition, 30 percent of all workers compensation claims are from work-related musculoskeletal disorders, yet only Californias' has OSHA standards for ergonomics. Increased compliance with existing rules can be only partially effective in reducing hazards and preventing injuries.

Second, other than anecdotes, little evidence suggests that OSHA's consultation and other voluntary programs, accounting for 28 percent of its 2003 budget, have had a measurable impact on hazards, injuries, or illnesses.[6] The U.S. General Accounting Office recently evaluated OSHA's four major voluntary compliance programs and

concluded that, although participants believe they are effective, "the agency does not yet have adequate data to assess their individual and relative effectiveness."[38]

Third, OSHA sends its small group of inspectors to inspect one workplace at a time, a never-ending pursuit as inefficient and unsatisfying as trying to season food one grain of salt at a time. Even with vastly increased resources, OSHA would never come close to the field visibility and penetration necessary for a major impact. With only 2,000 federal and state inspectors for some 8 million workplaces, only about 100,000 inspections are conducted each year. Each workplace can expect an inspection once every 88 years (compared with federal requirements that each school cafeteria have two food safety inspections per year [39] and that each underground coal mine have four inspections per year [40]).

Fourth, the OSHAct was written when employees were more likely than they are today to hold a long-term job with a single, stable employer and to be represented by a union. The 21st-century employment landscape—characterized by transnational businesses, global outsourcing, contingent employment, "independent" contracting, and a shadow economy of underground business relationships—makes it more difficult to find the kinds of workplaces for which the OSHAct was written.

PUSHING THE BOUNDARIES

Many ideas for administrative and statutory OSHA improvements have been proposed.[22, 41, 42] Most, such as higher penalties or bigger budgets, have sought incremental progress in implementing the current model. Given today's hazards and OSHA's design, however, it is not likely that simply working harder and better to do more of the same will achieve significant results. Although fine-tuning the OSHAct is necessary, a more profound shift in strategy is also needed, one that might feature four elements: leveraged impact, safety and health systems, independent employee rights, and equal protection. Some proposals for even more extreme redesign of the system are worth noting but are not further discussed here, including required redesign of technology,[43] making business licenses contingent on a reduction in injury rates,[44] and implementing injury taxes.[45]

LEVERAGE AND ACCOUNTABILITY

Under the current honor system, employers are required to comply with the rules but do not have to demonstrate compliance unless they are inspected. An alternate paradigm would leverage the small government inspectorate into a vastly expanded field presence by requiring every workplace to obtain an annual certification that it was inspected and in compliance with OSHA rules or operating under an abatement plan. The annual inspections would be done by private professionals licensed by and operating under rules established by OSHA. OSHA would conduct random field audits and continue to respond to complaints, fatalities, and catastrophes. Business owners would sign a declaration of compliance and be accountable for acts of negligence, fraud, and collusion.

Variations of this model already exist: construction and maritime crane certification under some state OSHA programs, pressure vessel inspections by insurance

companies, hospital accreditation by the private Joint Commission on Accreditation of Health Care Organizations, clinical laboratory inspections by private accrediting organizations overseen by the Department of Health and Human Services, the designation of private persons to inspect aircrafts on behalf of the Federal Aviation Administration, and certification of facilities by private nonprofit accreditation bodies under the Mammography Quality Standard Act.[46]

Proposals having some of these features are made periodically, with some business support and strong labor opposition,[47] For example, Sen. Mike Enzi's (R-Wyo.) Occupational Safety Partnership Act, proposed in 2005 but not enacted, would have allowed employers to engage private consultants for voluntary safety audits, with exemptions from most OSHA penalties.

The leveraged inspection program I suggest would differ significantly from the Enzi model in that inspections and certification would be mandatory. OSHA would perform audit and quality control functions, and business owners would be held accountable for performance. Sufficient checks and balances would avoid the dangers of uneven enforcement and of regulators becoming unduly influenced by the regulated community, which can happen when government delegates functions to the private sector.[48]

Even if the concept of a leveraged enforcement system is judged to be sound, substantially more discussion of the strengths and weaknesses of alternate program designs and operational details is needed. For example, would results of certification inspections be publicly available or discoverable in legal proceedings? Who would pay the inspectors and how could their independence be protected? How would OSHA maintain sufficient control over the inspection procedures?

SAFETY AND HEALTH SYSTEMS

Annual certification would have little value if it just documented compliance with rules for a limited number of specific hazards. It would be more meaningful in the context of a generic requirement for safety and health systems that find and fix recognized hazards. Voluntary models that include management commitment, employee participation, training, exposure assessment, hazard control, and medical surveillance already exist.[49] Several state OSHA or workers' compensation laws require safety and health systems for at least some employers.[50] Since 1989, OSHA has believed in "a strong correlation between the application of sound management practices in the operation of safety and health programs and a low incidence of occupational injuries and illnesses."[51] Smitha et al. have reported that such safety programs are associated with reduced workplace injury rates.[52]

INDEPENDENT EMPLOYEE RIGHTS

The legislative history of the OSHAct demonstrates the importance Congress placed on employee participation.[53] Accordingly, employees have rights to file complaints, accompany inspectors, be informed about hazards, and be protected from discrimination. These provisions essentially provide workers with opportunities to help OSHA carry out its duties rather than providing means to play an independent role in their protection. As a result, especially in nonunion workplaces, the

dynamic of safety and health is bipartite, between employer and government, with workers playing a marginal role.

A more balanced tripartite system would include stronger rights to refuse unsafe work and mandatory labor-management safety and health committees (with or without authority to shut down imminently dangerous operations). Smitha et al. have reported an association between mandatory safety committees and reduced injury rates.[52] One challenge in establishing safety committees would be in ensuring that they did not become de facto management-dominated bodies violating Section 8(a)(2) of the National Labor Relations Act.[54]

Another step to strengthen employee involvement would be giving workers the right to initiate legal action when employers endanger them or discriminate against them. There is precedent for such rights. The Federal Mine Safety and Health Act provides workers the right to file antidiscrimination actions on their own behalf with the Mine Safety and Health Review Commission or a federal court of appeals.[55] More than a dozen federal environmental laws, including the Clean Air Act, Safe Drinking Water Act, and Emergency Planning and Community Right to Know Act, provide for citizen suits.[56]

EQUAL PROTECTION

Most public employers are exempted from all requirements of the OSHAct, except for state and local government agencies in the 26 states that operate OSHA-approved state plans and the U.S. Postal Service. Federal agency heads are directed by executive order, but not by law, to comply with OSHA rules. This gap needs to be closed.

POLICY AND POLITICS

Good ideas often are held hostage to political realities, and these have made significant change to the OSHAct almost unthinkable. The substantive changes to the OSHA system suggested in the previous section will be impossible unless several political barriers are addressed.

Framing

Workplace safety has been marginalized as a matter of public concern. For example, soldiers are killed "in the line of duty," while construction workers die in "freak accidents." There are many reasons for our national disrespect and disregard for workplace injury and illness, including that workplace deaths usually happen one at a time behind factory walls or construction gates, as well as the persistent myth that workers who die must have been careless.[57] Whatever the reasons, this public blind spot must be addressed. Workplace health and safety need to be linked to the broad, resonant themes of the American experience and myth-opportunity, self-expression, family, fairness, justice, and human rights. Our vocabulary and symbols must convey that an injury at work breaks the American promise of a fair reward for a hard day's work, and that workplace injury is a violation of a revered trust.

A few examples show how workplace safety and health can be linked with a broader health and human rights agenda. The fight against workplace lead poisoning in the 1970s was coupled with efforts to protect children from lead paint and to

eliminate community harm from leaded gasoline. In the 1980s, the campaigns for the right of workers and communities to know about hazardous exposures were effectively combined. As of this writing, state legislatures seem particularly interested in protecting health care workers from the hazards of heavy manual lifting when worker safety is connected with patient safety and described as "safe patient handling."

One barrier to public recognition of workplace injuries has been significant underreporting. Twenty years after the National Academy of Sciences reported that the U.S. system for counting workplace injuries and illnesses was grossly inadequate and recommended 24 improvements, the only significant advance has been the Census of Fatal Occupational Injuries, which provides an improved count of deaths from workplace injuries (but not illnesses).[58] The Bureau of Labor Statistics' annual survey still fails to count as many as two-thirds of all work-related injuries and illnesses.[59] Rosenman et al. have noted the need for a multisourced surveillance system for collecting data on nonfatal injuries and occupational diseases that is comparable to the system for traumatic fatalities and is not dependent on employer reporting.[3]

Connections

The connections between workplace, environmental, and community health become meaningful through practical work around issues of shared importance. Previous efforts (for example, the Urban Environment Conference in the 1970s and the OSHA/Environmental Network in the 1980s) did not achieve lasting success.[60] Recent promising developments include the New Jersey Work Environment Council campaign for safe design technology for chemical plants to protect workers, communities, and the environment,[61] and the emergence of worker centers, which have discovered that "safety violations march hand in hand with racial and ethnic discrimination."[62]

Infrastructure

Durable success requires an institutional infrastructure that provides the intellectual framework and operational capacity for change. Infrastructure would include community-based advocacy groups, unions, professional and trade associations, labor-management organizations, think tanks, legal support, and training resources. It would provide mechanisms bringing safety and health professionals from the business community together with union and community activists to encourage the exploration of common interests and objectives.

The most notable earlier infrastructure effort joined professionals with union and community representatives in committees on occupational safety and health (COSHs) in the 1970s and 1980s. The COSH agendas included training and technical support to worker groups along with political organizing. These groups functioned within a supportive network, including university-based labor centers,[63] National Institute for Occupational Safety and Health (NIOSH) education resource centers, the Society for Occupational and Environmental Health,[64] the Association of Occupational and Environmental Clinics,[65] the American Public Health Association's Occupational Health and Safety Section,[66] and the safety and health

department of the American Federation of Labor and Congress of Industrial Organizations (AFL-CIO). The power of this network was evident in the 1980 defeat of Sen. Richard Schweiker's (R-Pa.) OSHA Improvement Act, which would have exempted many employers from inspections, and the successful campaign for local right-to-know laws, setting the stage for OSHA's 1983 Hazard Communication Rule, which requires employers to provide employees with useful information about the identify and the hazards of workplace chemicals.

Grants from OSHA's New Directions Program, launched in 1978, supported the growth of infrastructure. Originally designed for long-term institutional competency building, it was subsequently severely narrowed by OSHA. The current targeted training grant program is more narrowly aimed at technical training: only 25 percent of its 67 grants provide institutional support. Twenty-two COSHs still exist,[67] but most are marginal operations, many having shifted direction with grant support for asbestos and hazardous waste removal. A new program is needed that is based on the original New Directions model, substantially larger than the current program, and at least partially supported with private funds so it is less vulnerable to shifting political winds.

The Scientific and Technical Base

Preventing injury and illness often means taking action, in the face of uncertainty, that is based on public health principles and the best available evidence. This requires programs that teach the science of public health in the context of public health policy, law, and politics. NIOSH launched its education and research centers in 1977 with this linkage in mind. Sixteen of these centers currently train physicians, nurses, industrial hygienists, and safety professionals; however, for years, they have been troubled by flat, unstable budgets, and their programs have become more narrowly technical. There is a need to develop a new core of knowledgeable, articulate, and committed scientists who accept the challenge of building bridges between public health science and public health policy.

A ROAD MAP FOR CHANGE

The OSHAct of 1970 was landmark legislation, straddling the border between labor law and public health. It expressed a stunning set of principles, notably that every working man and woman is entitled to safe and healthful working conditions, and that employers are responsible for work being free from all recognized hazards. The full realization of Congress's vision will be beyond reach without significant new steps. In summary, these include the following:

1. Leverage and accountability: a requirement that business owners certify that their workplaces have passed an annual inspection for OSHA compliance, together with oversight of third-party inspectors and provisions to prevent fraud and conflicts of interest.
2. Safety and health systems: a requirement that employers have safety and health management systems that effectively find and fix recognized hazards.
3. Independent employee rights: establishment of a right for individual workers to take legal action for relief from workplace hazards and authority for labor-management safety committees to shut down jobs with imminent danger.

4. Equal protection: extension of OSHAct protections to all workers, including those in all federal, state, and local government workplaces.
5. Framing: increased public understanding that workplace safety is part of America's promise of fairness, justice, and human rights.
6. Infrastructure: support for a stable national network of safety and health institutions.

REFERENCES

1. Sparrow M. *The Regulatory Craft*. Washington, DC: Brookings Institution Press; 2000:2.
2. Occupational Safety and Health Act. Pub L 91-596. USC 651 et seq. 1970.
3. Rosenman K, et al. How much work-related injury and illness is missed by the current national surveillance system? *J Occup Environ Med*. 2006;48:357–365.
4. Leigh JP, et al. Occupational injury and illness in the United States: estimates of costs, morbidity, and mortality. *Arch Intern Med*. 1997;157:1557–1568.
5. Office of Technology Assessment. *Gauging Control Technology and Regulatory Impacts in Occupational Safety and Health—An Appraisal of OSHA's Analytic Approach*. Washington, DC: US Congress, Office of Technology Assessment; 1995. OTA-ENV Publication No.635.
6. Baggs J, Silbverstein B, Foley M. Workplace health and safety regulation: impact of enforcement and consultation on workers' compensation claims rates in Washington State. *Am J Indust Med*. 2003;43:483–494.
7. Gray WB, Scholz JT. Does regulatory enforcement work? a panel analysis of OSHA enforcement. *Law Soc Rev*. 1993;27:177–213.
8. Nelson NA, et al. Falls in construction: injury rates for OSHA-inspected employers before and after citation for violating the Washington State Fall Protection Standard. *Am J Ind Med*. 1997;31:296–302.
9. Fan ZJ, et al. *The Effect of DOSH Enforcement Inspections and Consultation Visits on the Compensable Claims Rates in Washington State: 2004–2005*. Olympia, WA: Washington State Department of Labor and Industries; 2006. SHARP Technical Report No.70-03-2006.
10. Weil D. If OSHA is so bad, why is compliance so good? *Rand J of Econ*. 1996;27:618–640. Conway H, Svenson J. Occupational injury and illness rates, 1992–1996: why they fell. *Mon Labor Rev*. 1998:36–58.
11. US Department of Labor, Bureau of Labor Statistics. Overview of BLS statistics on worker safety and health. http//www.bls.gov/bls/safety.htm. Accessed April 2, 1007.
12. Friedman L, Forst L. Occupational injury surveillance of traumatic injuries in Illinois, using the Illinois Trauma Registry: 1995–2003. *J Occup Environ Med*. 2007;49:401–410.
13. US Bureau of Labor Statistics. Census of fatal occupational injuries. http://www.bls.gov/iif/oshwc/cfoi/cfch004.pdf. Accessed April 17, 2007.
14. Herodotus. Book 7: Polymnia. In: Rawlinson G, trans. *The Persian Wars*. New York: Modern Library; 1942.
15. Kenwal R, et al. Evaluation of flavorings-related lung disease risk at six microwave popcorn plants. *J Occup Environ Med*. 2006;48:149–157.
16. Kreiss K, Cox-Ganser J. Metalworking fluid-associated hypersensitivity pneumonitis: a workshop summary. *Am J Indust Med*. 1997;32:423–432.
17. Bracker A, et al. An outbreak of hypersensitivity pneumonitis at a metalworking plant: a longitudinal assessment of interventional effectiveness. *Appl Occup Environ Hyg*. 2003;18:96–108.

18. National Institute for Occupational Safety and Health. *The Changing Organization of Work and the Safety and Health of Working People: Knowledge Gaps and Research Directions*. Cincinnati, OH: National Institute for Occupational Safety and Health; 2002. DHHS (NIOSH) publication 2002–116.
19. Landsbergis P. The changing organization of work and the safety and health of working people: a commentary. *J Occup Environ Med*. 2003;45:61–72.
20. Occupational Safety and Health Act. Sec 5(a)(1). 29 USC 654.
21. Ashford N. *Crisis in the Workplace: Occupational Disease and Injury*. Cambridge, MA: MIT Press; 1976.
22. Page JA, O'Brien MW. *Bitter Wages*. New York: Grossman Publishers; 1972.
23. Kniesner T, Leeth J. Abolishing OSHA. *Regulation*. 1995;4:46–56.
24. Industrial Union Department v American Petroleum Institute. 448 US 607, 8 OSH Cases 1586; 1980.
25. ALF-CIO v OSHA. 965 F2d 962, 15 OSH Cases 1729; 11th Cir 1992.
26. The Congressional Review Act. Pub L 104–121.
27. Chamber of Commerce v OSHA. 174 F3d 206; DC Cir 1999.
28. Banauch G, et al. Pulmonary function after exposure to the World Trade Center collapse in the New York City Fire Department. *Am J Respir Crit Care Med*. 2006;174:312–319.
29. Herbert R, et al. The World Trade Center disaster and the health of workers: five-year assessment of a unique medical screening program. *Environ Health Persp*. 2006;114:1853–1858.
30. Cooke W, Gautschi F. OSHA plant safety programs and injury reduction. *Indust Relat*. 1981;20:245–257.
31. Viscusi WK. The impact of occupational safety and health regulation. *Bell J Econ*. 1979;10:117–140.
32. Viscusi WK. The impact of occupational safety and health regulation: 1973–1983. *Rand J Econ*. 1986;17:567–580.
33. Bartel A, Thomas L. Direct and indirect effects of OSHA regulation. *J LawEcon*. 1985;28:1–26.
34. Ruser J, Smith R. Re-estimating OSHA's effects: have the data changed? *J Hum Resour*. 1990;26:212–235.
35. Lipscomb H, Leiming L, Dement J. Work-related falls among union carpenters in Washington State before and after the Vertical Fall Arrest Standard. *Am J Indust Med*. 2003;44:157–165.
36. Shapiro S, Rabinowitz R. Punishment versus cooperation in regulatory enforcement: a case study of OSHA. *Admin Law Rev American University*. 1997;49:713–762.
37. Weil D. Assessing OSHA performance: new evidence from the construction industry. *J Policy Anal Manage*. 2001;20:651–674.
    Mendeloff J. The role of OSHA violations in serious workplace accidents. *J Occup Med*. 1984;26:353–360.
38. Occupational Safety and Health Administration. *Workplace Safety and Health-OSHA's Voluntary Compliance Strategies Show Promising Results, but Should be Fully Evaluated before They Are Expanded*. Washington, DC: US General Accounting Office; 2004. GAO Publication No.04-378.
39. US Food and Drug Administration. 2005 food code. http://www.cfsan.fda.gov/-dms/fc05-toc.html. Accessed April 2, 2007.

40. The Federal Mine Safety and Health Act of 1977. Sec 103 Pub L 91–173.
41. Wegman D, Robbins A, eds. A new OSHA: the tasks for the first 100 days. *J Pub Health Policy*. 1988;9:319–345.
42. McGarity T, Shapiro S. *Workers at Risk: The Failed Promises of the Occupational Safety and Health Administration*. Westport, CT: Praeger; 1993.
43. Ashford N. Regulating occupational health and safety: the real issues. *Challenge*. 1976;39–42.
44. O'Neill P, Secretary of the Treasury. presentation at Georgetown University Workplace Safety Summit; March 30, 2001; Washington, DC.
45. Smith R. The feasibility of an "injury tax" approach to occupational safety. *Law Contemp Probl*. 1974;38:730–744.
46. Mammography Quality Standards Act. 42 USC 6A(11)(F)(3)263; 1994.
47. Levenstein C, Eisen EA. Certified environmental audits: a proposal. *J Pub Health Policy*. 1987;3:303–308.
48. Administrative Conference of the United States. Recommendation 1CFR 305.94–1: The Use of Audited Self-Regulation as a Regulatory Technique. ABA Administrative Procedure Database. Published June 16, 1994.
49. Occupational Safety and Health Administration. Voluntary safety and health program management guidelines. *Fed Regist*. 1989;54:3904–3916.
50. Injury and Illness Prevention Program. *California Code of Regulations, Title 8, sec 3203*.
51. Occupational Safety and Health Administration. Voluntary safety and health program management guidelines. op. cit. p. 3094.
52. Smitha M, et al. Effect of state workplace safety laws on occupational injury rates. *J Occup Environ Med*. 2001;43:1001–1010.
53. 91st Cong., 2nd Sess, 11–12. *US Senate Report No 1282*. Washington, DC: US Senate; 1970.
54. National Labor Relations Act. Section 158(a)(2) and 8(a)(2). 29 USC 7, 151–169.
55. Federal Mine Safety and Health Act of 1977. sec 815(3). 30 USC 22.
56. May J, ed. Environmental citizen suits at thirty something: a celebration and summit Part I and II. *Widner Law Rev*. 2003;10:1–2.
57. Barab J. Acts of god, action of man. In: Mogensen V, ed. *Worker Safety under Siege*. Armonk, NY: M.E. Sharpe; 2006:3–16.
58. National Research Council. *Counting Injuries and Illnesses in the Workplace. Proposals for a Better System*. Washington, DC: National Academies Press; 1987.
59. Leigh P, Marcin J, Miller T. An estimate of the US Government's undercount of nonfatal occupational injuries. *J Occup Environ Med*. 2004;46:10–18.
60. Samuels S. A memoir on risk assessment and environmental policy. *Ann N Y Acad Sci*. 1997;837:418–432.
61. New Jersey Work Environment Council. http://www.njwec.org. Accessed April 2, 2007.
62. Fine J. *Worker Centers: Organizing Communities at the Edge of the Dream*. Ithaca, NY: ILR Press; 2006:160.
63. Labor Occupational Health Program at University of California-Berkeley. Welcome to LOHP. http://www.socrates.berkeley.edu/-lohp. Accessed April 2, 2007.
64. Society for Occupational and Environmental Health. SOEH: an international society. http://www.soeh.org. Accessed April 2, 2007.

65. Association of Occupational and Environmental Clinics. http://www.aoec.org. Accessed April 2, 2007.
66. American Public Health Association. Occupational health and safety section. http://dept.washington.edu/oshalert/index.html. Accessed April 2, 2007.
67. The National Council for Occupational Safety and Health (formerly the National COSH Network). http://www.coshnetwork.org. Accessed April 2, 2007.

This page intentionally left blank

CHAPTER 8

Planning for and Responding to Public Health Emergencies: Modeling the Health Outcomes and Logistical Requirements of a Large-Scale Bioterrorism Attack

Nathaniel Hupert, MD, MPH, Daniel Wattson, MD, and Wei Xiong, PhD

Concern about the threat of large-scale bioterrorism grew after the 2001 U.S. anthrax attack, which killed five people and paralyzed the nation for weeks. Since then, the U.S. federal government has spent more than $15 billion on bioterrorism preparedness—including surveillance, laboratory and hospital surge capacity, and countermeasure stockpiles.[1] Under the Cities Readiness Initiative (CRI) of the U.S. Centers for Disease Control and Prevention (CDC), local public health agencies across the country have expanded their capability to dispense these countermeasures to targeted populations within 48 hours of a bioterrorism event. Neither the health outcomes nor logistics associated with this goal are well characterized. Using the example of a large-scale aerosol anthrax exposure over an urban locale, we explore key determinants of health outcomes and health system surge capacity using several modeling techniques (state transition, queuing network). We suggest that such models can provide valuable insights for forecasting the logistical and staffing needs of large-scale prophylaxis campaigns for a range of intentional and natural disease outbreaks, such as the 2009 influenza A (H1N1) pandemic.

INTRODUCTION

Biological weapons have been used throughout history to gain tactical and psychological advantage. The siege of the Crimean city of Caffa in the year of 1346 by invading Mongol hordes who catapulted plague-infested cadavers into the walled city is often cited as the first and most important recorded incident of biological warfare. Two more recent bioterrorist events in the U.S. involved contamination of the food supply. In 1984, a large community outbreak (751 cases) of *Salmonella typhimurium* was caused by intentional contamination of restaurant salad bars by members of the Rajneesh religious cult in The Dalles, Oregon.[2] In 1996, an outbreak of *Shigella dysenteriae* type 2 among 12 laboratory workers at a large Texas medical center was linked to muffins and donuts anonymously left in the break room.[3] Most recently, the U.S. anthrax attacks of October 2001, which was propagated through letters

containing anthrax spores mailed to offices of news media and two U.S. senators, led to 5 deaths and 17 serious infections.[4]

From a public health perspective, these attacks created two distinct populations requiring evaluation and treatment: those directly affected (numbering thousands of people, including exposed patients, health care workers, and law enforcement professionals); and the numerous anxious and worried individuals who flooded local hospital emergency rooms, physicians' offices, and health information hotlines around the country. The actions of local public health agencies in triaging and providing countermeasures to the affected individuals in the aftermath of such an event will determine its health and social impact. However, since it is infeasible to wait for further attacks or to conduct truly realistic exercises involving thousands of actors, local public health agencies have increasingly looked to quantitative computer models to make informed decisions about such high-risk, low-probability events. For example, the CDC has developed several influenza-related models (for example, FluAid, FluSurge) designed specifically to assist local public health agencies in planning for pandemic response. These models were downloaded more than 100,000 times during the 2009 pandemic A (H1N1) response. In this chapter, we focus on anthrax, classified in the most serious risk category for bioterrorism agents, as an example to illustrate the use of such analytical models to represent the impact of a bioterrorism attack over a large population, the consequences of different response policy options, and the logistical considerations for large-scale public health interventions, such as mass-prophylaxis, in mitigating its effect.

The release of anthrax spores over a major metropolitan region, an event with the potential to cause massive casualties on the scale of a small nuclear detonation, had received considerable federal attention even before the 2001 U.S. mail attacks. With the creation of the Strategic National Stockpile (SNS) in 1999, the U.S. government directed significant financial and human resources to the development of countermeasures and strategies to mitigate the effects of intentional and natural outbreaks of disease.[5] In 2004, the Department of Health and Human Services (DHHS) convened the Anthrax Modeling Working Group (AMWG) of the Secretary's Council on Public Health Emergency Preparedness to provide quantitative guidance on the purchase and use of medical countermeasures to minimize the health impact of anthrax release over civilian populations. The CDC unveiled the Cities Readiness Initiative (CRI), the nation's first attempt to provide quantitative goals for public health actions, such as countermeasure dispensing, to minimize casualties in the aftermath of such an event.[6]

The CRI mass prophylaxis recommendations focus on a "48-hour goal" for countermeasure distribution.[7, 8] Taking the lead from CRI grantees, which now include the 72 largest U.S. cities, local public health agencies across the national have endeavored to develop the ability to dispense countermeasures to all designated at-risk populations within 48 hours of the decision to do so.[8] Given this effort, it is notable that when the initiative started in 2004 no published reports had quantitatively evaluated how outcomes varied in relation to meeting or missing the 48-hour goal. Achieving such rapid countermeasure dispensing capability has important implications for resources (human and material) and logistics, which depend in part on the details of how local public health agencies are intending to carry out mass-prophylaxis campaigns (that is, using newly established points of dispensing

[PODs] or existing health care facilities and providers). The following models attempt to provide a systematic, quantitative basis for assessing these requirements.

A STATE TRANSITION MODEL OF HEALTH OUTCOMES AFTER A BIOTERRORISM ATTACK

The first model presented here, first developed in 2002–2003 for the Agency for Healthcare Research and Quality (AHRQ), was used during the initial formulation of the CRI to provide quantitative assessments of the effect of mass prophylaxis interventions on expected hospital surge arrivals in the aftermath of an aerosol anthrax attack.[9] We use this model to explore two interrelated questions: What are the consequences of missing the CRI 48-hour goal for postexposure prophylaxis? What impact does initial response time, anthrax incubation period, and antibiotic effectiveness have on expected hospitalizations?

We created a discrete-time, state-transition model representing the dynamic interaction between the rate of progression to symptomatic inhalational anthrax and the rate of successful dispensing and effective use of prophylactic antibiotics in a defined population after a large-scale anthrax attack. We focus on the first week of response after population exposure. The formulation of a state-transition model involves identifying the underlying "states," in this case phases of progression of disease or removal into a nonprogressing group, determining how the states are linked and quantitatively describing the transitions between them.

We incorporate transitions between asymptomatic-but-exposed, symptomatic, and prophylaxed states on a daily basis, and calculate the area under the incubation distribution curve (that is, the number of potentially symptomatic individuals). The transition parameters are based on recently published epidemiological analyses of the 1979 aerosol anthrax release in Sverdlovsk, Russia, which represents the only documented population-level anthrax exposure in modern history.[10–12] The model applies to a hypothetical population that has been exposed to sufficient anthrax spores to cause symptomatic illness. The probability of becoming symptomatic changes because of three factors: (1) the probability distribution for the incubation period for inhalational anthrax (see below), (2) the particular prophylaxis strategy undertaken, and (3) the effectiveness of the countermeasures given.

A prophylaxis strategy is defined by two tactics: the delay until countermeasure dispensing starts ("time to first pill") and the length of time needed to successfully prophylaxis all those eligible for treatment ("time to last pill"). Once prophylaxis begins, we assume a uniform rate of countermeasure dispensing with no ramp-up period.[13] Some patients in this at-risk cohort will, by chance, be offered prophylaxis on the same day that they happen to become symptomatic (probabilistically, due to the incubation distribution). For these patients, we assume that 38.1 percent will be hospitalized, based on data suggesting that this is the rate of progression to critical illness for inhalational anthrax if antibiotics are initiated during the prodromal phase of infection.[14]

The model output is the number of individuals who are expected to develop symptomatic inhalational anthrax requiring hospital-based intravenous antibiotic treatment. The inverse of this is the "save rate," or the percentage of exposed and "at-risk" individuals who successfully avoid developing illness due to the prophylaxis

campaign. The model does not explicitly consider spore dispersal; modeling aerosol plumes involves complex interactions between spore size, charge, and additives; weather conditions; release characteristics; population demographics; building protection factors; time of day; and other factors. Recently, Brookmeyer et al. demonstrated the relative insensitivity of the incubation period distribution to dose of inhaled anthrax spores over a wide range of exposure intensities.[15] The model user defines a subset of the target population as having sufficient anthrax spores to cause symptomatic infection in the absence of treatment, 95 percent of these exposed individuals would develop inhalational anthrax in the absence of intervention.[16]

The baseline probability distributions for the incubation (or latency) period for inhalational anthrax come from analyses of the Sverdlovsk outbreak by Brookmeyer et al. that have recently been revised by Wilkening.[10, 11] Brookmeyer fit the timing of hospitalization of 70 cases of inhalational anthrax to a lognormal distribution with a mean of 2.398 and a standard deviation of 0.713, corresponding to a median time to onset of symptoms of 11 days with a dispersion factor (e^{σ}) of 2.04 days. While there has been controversy about the integrity of the Sverdlovsk data and consequently the true shape of an inhalation anthrax latency curve,[17] recent comparative analyses by Wilkening provide strong support for using a modified Brookmeyer curve (with a median incubation period of 9.5 days and a dispersion factor of 1.91 days) as a basis for population-based exposure modeling.[11, 12, 17] Prophylactic interventions for anthrax will have a theoretical efficacy (the ability to halt disease progression under ideal circumstances) and a practical effectiveness (how this efficacy plays out under lifelike conditions, including such patient-related factors as adherence). Because effectiveness ultimately is what matters in calculating both patient-level and population-wide outcomes (that is, probability of symptomatic illness and regional hospital surge), we include only the latter as a variable in the model. For consistency with prior studies, we assume 90 percent effectiveness of interventions for individuals who receive their antibiotics before becoming symptomatic.[6, 18] To maximize usability, our model is a Microsoft Excel™ workbook running a Visual Basic for Applications (VBA) macro. The model is available at http://www.simfluenza.net.

Using this model with baseline assumptions of a CRI-compliant 48-hour mass prophylaxis campaign after two-day delay between exposure and initiation of response, we estimate that 86 percent of exposed individuals would be protected from developing symptomatic inhalational anthrax (see table 8.1). If antibiotic effectiveness were 100 percent, population protection with these same response parameters would increase to 95.6 percent (see table 8.2). In contrast, with long delays (that is, if detection and response were delayed by one week and the campaign took seven days instead of 48 hours), population protection would be reduced to 39.5 percent.

Figure 8.1 illustrates the increase in expected hospitalizations due to missing the CRI goal of a 48-hour dispensing campaign. This figure shows that the detrimental effect of longer campaigns vary with the delay to campaign initiation. On average, the first additional campaign day beyond the CRI goal leads to an additional 2.7 percent of the exposed population requiring hospitalization (with a range between 0.6 and 3.6 percent, varying with the delay to campaign initiation from one to seven days). In contrast, a single day's delay in initiation of a CRI-compliant 48-hour campaign would lead to an average of 6 percent more of the exposed population requiring hospitalization (range from 2.8 to 7.4 percent, depending on the original delay until

Table 8.1
Proportion of Exposed Population Protected against Hospitalization under Different Mass Prophylaxis Tactics Varying Start and Duration of Campaign (90 percent antibiotic efficacy, ID10, 95 percent attack rate, Wilkening curve)

		Delay in initiating dispensing campaign after exposure							
		Immediate	1 day	2 days	3 days	4 days	5 days	6 days	7 days
Campaign duration	1 day	90.0%	89.7%	88.0%	84.0%	78.1%	71.0%	63.6%	56.2%
	2 days	89.8%	88.8%	86.0%	81.1%	74.6%	67.4%	60.0%	52.9%
	3 days	89.2%	87.2%	83.4%	77.8%	71.0%	63.8%	56.6%	49.8%
	4 days	87.9%	85.0%	80.3%	74.3%	67.4%	60.3%	53.4%	46.9%
	5 days	86.0%	82.2%	77.0%	70.8%	63.9%	57.0%	50.4%	44.2%
	6 days	83.5%	79.1%	73.6%	67.3%	60.6%	53.9%	47.6%	41.8%
	7 days	80.7%	75.9%	70.3%	64.0%	57.4%	51.0%	45.0%	39.5%

Source: Hupert, Wattson, and Wei Xiong

Table 8.2
Proportion of Exposed Population Protected against Hospitalization under Different Mass Prophylaxis Tactics Varying Start and Duration of Campaign (100 percent antibiotic efficacy, ID10, 95 percent attack rate, Wilkening curve)

		Delay in initiating dispensing campaign after exposure							
		Immediate	1 day	2 days	3 days	4 days	5 days	6 days	7 days
Campaign duration	1 day	99.99%	99.6%	97.7%	93.4%	86.8%	78.9%	70.6%	62.5%
	2 days	99.8%	98.7%	95.6%	90.1%	82.9%	74.8%	66.6%	58.8%
	3 days	99.1%	96.9%	92.6%	86.4%	78.9%	70.8%	62.9%	55.3%
	4 days	97.7%	94.4%	89.2%	82.5%	74.9%	67.0%	59.3%	52.1%
	5 days	95.5%	91.3%	85.6%	78.6%	71.0%	63.3%	56.0%	49.1%
	6 days	92.8%	87.9%	81.8%	74.8%	67.3%	59.9%	52.9%	46.4%
	7 days	89.6%	84.3%	78.0%	71.1%	63.8%	56.7%	50.0%	43.8%

Source: Hupert, Wattson, and Wei Xiong

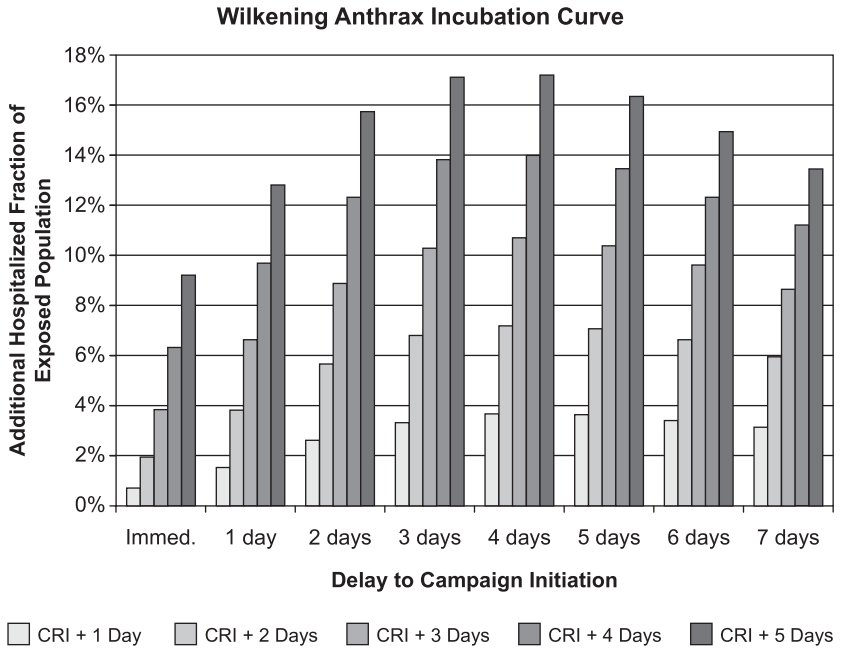

Figure 8.1 Increase in hospital surge with mass prophylaxis campaigns that exceed the Cities Readiness Initiative (CRI) 48-hour campaign goal (Wilkening incubation distributions, 90% antibiotic effectiveness, 95% attack rate).

campaign initiation). For example, if a 48-hour dispensing campaign were carried out after a delay of four days instead of two days, an additional 11.4 percent of the exposed proportion would be expected to require hospitalization. The greatest marginal increase in hospitalizations is seen with delays of four to five days, after which the marginal effect of each additional day duration declines.

Population protection declines in a roughly linear fashion with decreasing antibiotic effectiveness, but this relationship is modulated by delay to campaign initiation. For the baseline case of a two-day delay to campaign initiation and a 48-hour campaign, every 1 percent decrease in antibiotic effectiveness leads to a 0.95 percent increase in expected hospitalizations. The timing of these hospitalizations, however, is highly nonlinear. For the baseline scenario with 90 percent effectiveness and a two-day delay, hospitalizations are temporally distributed as a "double hump" peaking on postexposure days three and seven with an extended right-hand "tail" of late hospitalizations. Delay to campaign initiation determines the size of the initial surge, which may merge into the second "hump" if antibiotic dispensing does not occur within the first seven days.

Depending on response time, there is a range in the rate of expected hospitalizations with a 48-hour campaign with varying delay to campaign initiation and a median anthrax incubation period (from 8 to 14 days using a lognormal distribution). With an immediate response, population protection is highly insensitive to incubation duration, ranging from 89.5 percent (median eight-day incubation period) to 90.1 percent (median 14-day incubation period). As delay to response

increases, so too does variation in predicted outcomes. For a seven-day delay, protection varies from 46.1 percent under the fastest incubation assumption to 71.3 percent under the slowest. For the base case assumption of a two-day delay in response, population protection ranges from 82.7 to 88.7 percent (for 8- to 14-day incubations). This yields an absolute difference of 6 percent change in expected hospitalizations despite an almost doubling of the anthrax incubation period for the base case mass prophylaxis campaign.

In this section we showed how a computer model can be used to quantify the ability of mass countermeasure-dispensing campaigns in preventing serious health outcomes after an intentional release of a bioterrorism agent. The findings of this model led to a critique of the current CRI emphasis on reducing campaign duration to 48 hours or less, since this appears to constitute important but insufficient guidance to ensure adequate protection of exposed populations in the aftermath of such an attack. Specifically, our model sheds light on two additional factors relating to antibiotic prophylaxis—delay in campaign initiation and effectiveness of dispensed medications (which, in turn, is dependent on patient adherence and efficacy of the treatment)—that have a greater relative impact on projected hospitalizations than campaign duration. If an optimistic public health goal is the prevention of more than 80 to 85 percent of expected hospitalizations in such a scenario, this model suggests that the delay until commencement of a CRI-compliant campaign should be *no more than three days.*

A QUEUING NETWORK MODEL OF WORKFORCE REQUIREMENTS FOR MASS PROPHYLAXIS USING PODs

With a better understanding of the consequences of different policy options, the next challenge facing local public health agencies in planning effective responses to a bioterrorism attack or other outbreak of preventable infectious disease is to prepare for the logistic requirements in dispensing medical countermeasures to a large population within a limited time frame. In this section, we describe the formulation of a simulation and queuing-network model, the Bioterrorism and Epidemic Outbreak Response Model (BERM), to determine the logistical and staffing needs of a large-scale prophylaxis campaign designed to cover a target population. The model quantifies long-run performance with the equations of queuing theory, which helps the user assign resources appropriately. (A Web-application of the BERM model is available at http://www.simfluenza.org/BERMweb/input.aspx). The BERM assumes that, at the onset of an outbreak scenario, the community's emergency response planners will initiate the formation of several prophylaxis PODs throughout the community. An individual POD can be modeled as a network of stations through which patients flow. We consider a relatively simple, high-flow, four-station POD setup in this section,[19] but the same methodology can be applied to alternative setups. We define the four stations as follows:

- Greeting/Entry: Staff performs a rapid, initial check to route visibly ill patients immediately to the medical evaluation area; other patients are routed to triage.
- Triage: Staff follows a triage protocol to determine whether patients should receive a more detailed medical assessment at the medical evaluation station; other patients are deemed healthy and are routed to the drug dispensing station.

- Medical Evaluation: Medical professionals perform a detailed evaluation of the patients routed there and stabilize and send the ill patients to the health center; patients deemed healthy are routed to the drug dispensing station.
- Drug Dispensing: Staff disseminates an appropriate amount of prophylactic drugs to each patient, and the patient leaves the POD.

The number of individual staff required to process patients at each POD station can be determined with one of the following techniques:

1. Deterministic Calculation: A deterministic calculation is a simple algebraic operation that yields the number of staff assuming no variability in patient arrival or processing time. For example, if two patients are expected to arrive at a station each minute and a single staff member can serve a single patient in one minute, a deterministic calculation would conclude that the station should be staffed with two individuals. The mathematical formula is as follows:

   (2 patients/minute) / (1 minute/patient/staff member) = 2 staff members

   Advantages: Simple, fast calculation.

   Disadvantages: Does not account for the variability inherent in the arrival and service rates, which leads to an underestimation in the number of staff needed (in the calculation above, although the system appears "balanced," significant queuing may occur under variability because the two staff members cannot "store" extra capacity when idle and later "spend" it when there is a temporary surge of patients).
2. Trial-and-Error with Stochastic Simulation: If a stochastic simulation model (that is, a simulation model that accounts for variability in patient arrival and service times) of the system is built, a user can guess at the station staffing levels, run the simulation, and continually adjust the staff and rerun the simulation until the performance of the system is adequate.

   Advantages: Accounts for variability and other real-life properties of the system.

   Disadvantages: Cumbersome and time-consuming, requires the construction of a complex simulation model.
3. Queuing Theory: The engineering field of queuing theory provides equations that can be used to calculate the long-run (that is, steady-state) performance of a staffed system under variable arrival and service rates. The equations yield various helpful measures of system performance, including the average length of the queue, average wait time in the queue, probability that the staff are idle, and so forth. The number of staff can be determined by specifying a maximum or minimum acceptable value for a performance measure and calculating the number of staff necessary to meet that standard.

   Advantages: Simple, fast calculation that also accounts for the stochastic nature of the system.

   Disadvantages: Determines only long-run results and is not as descriptive as a discrete-event simulation.

The BERM utilizes a queuing theory approach to calculate the number of staff to be allocated to each station. It then permits the user to run a discrete-event stochastic simulation with the given POD setup to view the more realistic consequences of the staffing plan. Several similar models developed at academic, public health, medical, and engineering centers in the United States are available for Internet download (for example, the University of Maryland Clinic Model, the Georgia Tech RealOpt model). The model assumes that patients arrive at a POD according

to a Poisson process, whereby the individual interarrival times vary according to an exponential distribution with a mean of the expected throughput.[1] This arrival process can be stable (that is, with a single mean value for the entire duration of clinic activity) or time varying. For simplicity, we describe results for a stable Poisson arrival process, but the implications of dynamic Poisson processes on staffing will be considered below. Based on expectations of patient arrival rates at each POD and interior organization of the stations within each POD, public health planners may use models like BERM to estimate the number of staff. Staff members served the patients according to exponentially distributed service times. All staff members are assumed to be of similar capability, and thus will take the same average amount of time to serve patients (though with patient-to-patient variability). If a patient arrives at a station and all staff are busy serving others, the patient enters a first-come, first-served queue. The model assumes that patients will not balk when they arrive, nor will they renege while waiting in the queue.

Using the common Kendall-Lee notation for queues, this station setup would be described as having an *M/M/s* queue. "M" stands for "Markovian," which implies the exponential distribution (the first *M* describes the arrival process and the second describes the service process) and the *s* stands for the number of staff at the station. The patients are routed from station to station in the POD according to a predetermined set of routing probabilities. This network of *M/M/s* stations is otherwise known as an *Open Jackson Network* (James Jackson's theorem in queuing theory). The queuing theory equations permit the user to determine various steady-state performance measures at each station, which can in turn be interpreted to decide on appropriate staffing levels. The equations used in the model require the following three inputs for each POD station:

- The mean arrival rate, λ (which implies a mean inter-arrival time of $1/\lambda$)
- The mean service rate, μ (which implies a mean service time of $1/\mu$)
- The number of staff, s

Important performance measures for each POD station are calculated as follows:

- Expected Number in Queue (L_q) =

$$\left(\sum_{n=0}^{s-1}\frac{(\lambda/\mu)^n}{n!}+\frac{(\lambda/\mu)^s}{s!}\left(\frac{1}{1-(\lambda/s\mu)}\right)\right)^{-1}\times\left(\frac{(\lambda/\mu)^{s+1}}{(s-1)!(s-\lambda/\mu)^2}\right) \qquad (8.1)$$

1. The density function of an exponential distribution is $f(x) = 1/\beta \cdot \exp(-x/\beta)$ for $x \geq 0$, with a mean of β. The true distributions of the interarrival and service times may not be precisely exponential, but this is a reasonable assumption given the relative dearth of data on mass prophylaxis arrival and service rates. The exponential distribution is widely used in queuing theory and simulation modeling as an approximation for interarrival and service times. There is reason to believe, based on limited data from live mass prophylaxis exercises, that the interarrival times are indeed exponentially distributed but the actual distributions of the service times may be slightly less variable than the exponential distribution. The BERM's outputs for the average queue lengths and wait times therefore are likely to be greater than the real-life queuing values, which means that the model errs on the side of caution.

- Expected Wait Time in Queue $(W_q) = \frac{L_q}{\lambda}$, a relationship known as Little's Law
- Expected Staff Utilization $(\rho) = \frac{\lambda}{s\mu}$

The BERM permits the user to specify a maximum acceptable *Lq* or *Wq* at each station, and an algorithm continually increases the number of staff assigned to each station until this standard is met. The mean service rate (μ) at each station can be determined by taking the reciprocal of the expected processing times. The mean arrival rate (λ) at each station, however, is more difficult to determine in the *Open Jackson Network* framework. The λ at the greeting/entry station is the incoming arrival rate to the POD, but the arrival rates at the other stations may differ from this because not all patients are necessarily routed to each station. The true arrival rates can be determined by using matrix algebra to create a vector that contains the arrival rates to each station. This arrival rate vector, Λ, follows the following relationship:

$$\Lambda = \left[(I - R)'\right]^{-1} A \tag{8.2}$$

In this equation, *I* is the identity matrix with the same dimension as the number of stations in the POD (four, in this case). *R* is the routing probability matrix, which is a square matrix, also with the same dimension as the number of stations in the POD. The stations are assigned to the rows and columns in the same order, and each cell in the matrix represents the probability that a patient will be routed from the row station to the column station.

The routing probability matrix is subtracted from the identity matrix and the result is transposed, inverted, and multiplied by *A*, the external arrival rate vector. The column vector *A* contains the arrival rates into the POD at each station; in this case, because patients arrive only at the greeting/entry station, only the first row of the vector is populated. For instance, if the mean arrival rate into the POD is 10 patients per minute, *A* is as follows:

$$\begin{bmatrix} 10 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The calculation produces the following arrival rate vector, Λ, which contains the individual values for λ (in patients per minute) at each station:

$$\begin{bmatrix} 10.0 \\ 9.5 \\ 0.975 \\ 9.95125 \end{bmatrix}$$

As an example, if we assume an overall POD throughput of 10 patients per minute, the mean arrival rate to the POD also will be 10 patients per minute and the arrival rates at each POD station are the same as those determined in the previous

calculation of Λ. The mean service times are expected to be 0.3 minutes for greeting/entry, 3 minutes for triage, 10 minutes for medical evaluation, and 1 minute for drug dispensing (these figures were based on observed service times from several mass prophylaxis live exercises).

Under the assumption of a stable mean arrival rate, optimal staffing of this POD would require 52 active staff. If variable arrival and service times are considered, however, substantial queuing would occur with this staffing arrangement. Additional on-call staff are needed at each station to handle the periodic patient surges and subsequent patient queuing. The lower the tolerance for queuing at the system's stations, the more staff are required. For instance, if the maximum acceptable average queue length (Lq) at each station is one patient, 66 active staff are required in the POD (an increase of 27 percent over the deterministic scenario). As the threshold gets particularly low (for example, less than one patient), the number of staff required increases dramatically. It is thus unrealistic to assign so many staff that no queuing will occur, so an acceptable level of queuing must be decided upon.

Public health planners can use models such as BERM to reproducibly generate quantitative estimates of staff requirements at individual emergency dispensing clinics. By multiplying these estimates across a number of clinics for a designated geographic or population area, planners can generate campaign-wide staffing estimates that can be combined with past experience, expert opinion, and the results of exercises to form the basis for robust emergency response plans. Computer models such as these are best seen as a starting point for planning, because their accuracy depends on assumptions made about the lengths of time required to process different patient types (families of different sizes, elderly people, disabled individuals, and so on), the percentages of various patient types that follow each possible route through the POD, and the pattern of arrivals throughout the time the POD is in operation. For example, most available CRI modeling programs assume a homogeneous arrival process. POD-based mass prophylaxis operations initiated in response to public health emergencies are almost certain to be conducted against a backdrop of considerable uncertainty regarding population behavior, availability of countermeasures, and operational capability of dispensing centers.

SUMMARY

Although all model-based studies have their potential limitations, models may serve many functions in emergency preparedness and planning that cannot be provided through other means. We believe these models and others like them can assist emergency planners in understanding the scope of the problem and giving a picture of the downstream effects of proposed interventions. Models may help bridge the gap between public health and hospital-based emergency response planning for anthrax and other bioterrorism attacks, with specific applications for two types of users: (1) public health and emergency management personnel with responsibility for developing effective regional mass prophylaxis campaigns; and (2) hospital or health system managers whose facilities would have responsibility for managing the resulting surge in inhalational anthrax cases. Additional audiences may include individuals in the public or private sectors with other roles in responding to terrorism and other public health emergencies (for example, law enforcement and public communications),

providing mass prophylactic treatment (for example, pharmacists, community-based primary care physicians, public and private school officials), providing mass casualty care (for example, ambulance services, medical suppliers), and dealing with the consequences of mass fatalities (for example, coroners and social service providers).

REFERENCES

1. Franco C. Billions for biodefense: federal agency biodefense funding, FY2008-FY2009. *Biosecur Bioterror*. 2008;6(2):131–146.
2. Torok TJ, Tauxe RV, Wise RP, et al. A large community outbreak of salmonellosis caused by intentional contamination of restaurant salad bars. *JAMA*. 1997;278(5):389–395.
3. Agency for Healthcare Research and Quality. *Pediatric Terrorism and Disaster Preparedness: A Resource for Pediatricians*. Rockville, MD: Agency for Healthcare Research and Quality; 2006. Published October 2006. AHRQ Publication Nos.06(07)-0056/06(07)-0056-1.
4. Jernigan DB, Raghunathan PL, Bell BP, et al. Investigation of bioterrorism-related anthrax, United States, 2001: epidemiologic findings. *Emerg Infect Dis*. 2002;8(10):1019–1028.
5. Franco C, Deitch S. Billions for biodefense: federal agency biodefense funding, FY2007-FY2008. *Biosecur Bioterror*. 2007;5(2):117–133.
6. Wein LM, Craft DL. Evaluation of public health interventions for anthrax: a report to the secretary's council on Public Health Preparedness. *Biosecur Bioterror*. 2005;3(4):348–356.
7. Centers for Disease Control and Prevention. Appendix 3: cities readiness initiative guidance: cooperative agreement guidance for public health emergency preparedness. http://emergency.cdc.gov/planning/guidance05/pdf/appendix3.pdf. Published 2005. Accessed September 11, 2008.
8. Koh HK, Elqura LJ, Judge CM, et al. Implementing the cities readiness initiative: lessons learned from Boston. *Disaster Med Public Health Prep*. 2008;2(1):40–49.
9. Agency for Healthcare Research and Quality. Mass prophylaxis: building blocks for community preparedness. bioterrorism and health system preparedness. Rockville, MD: Agency for Healthcare Research and Quality; 2006. http://www.ahrq.gov/news/ulp/btbriefs/btbrief10.htm. Issue Brief No.10. AHRQ Publication No.06-0050. June 2006.
10. Brookmeyer R, Blades N, Hugh-Jones M, Henderson D. The statistical analysis of truncated data: application to the Sverdlovsk anthrax outbreak. *Biostatistics*. 2001;2(2):233–247.
11. Wilkening DA. Modeling the incubation period of inhalational anthrax. *Med Decis Making*. 2008;28(4):593–605.
12. Wilkening DA. Sverdlovsk revisited: modeling human inhalation anthrax. *Proc Natl Acad Sci U S A*. 2006;103(20):7589–7594.
13. Craft D, Wein L, Wilkins A. Analyzing bioterror response logistics: the case of anthrax. *Manage Sci*. 2005;51:679–694.
14. Holty JE, Bravata DM, Liu H, Olshen RA, McDonald KM, Owens DK. Systematic review: a century of inhalational anthrax cases from 1900 to 2005. *Ann Intern Med*. 2006;144(4):270–280.
15. Brookmeyer R, Johnson E, Barry S. Modeling the incubation period of anthrax. *Stat Med*. 2005;24(4):531–542.
16. Fowler RA, Sanders GD, Bravata DM, et al. Cost-effectiveness of defending against bioterrorism: a comparison of vaccination and antibiotic prophylaxis against anthrax. *Ann Intern Med*. 2005;142(8):601–610.

17. Inglesby TV, Henderson DA, Bartlett JG, et al. for Working Group on Civilian Biodefense. Anthrax as a biological weapon: medical and public health management. *JAMA*. 1999;281(18):1735–1745.
18. Baccam P, Boechler M. Public health response to an anthrax attack: an evaluation of vaccination policy options. *Biosecur Bioterror*. 2007;5(1):26–34.
19. Hupert N, Cuomo J, Callahan M, Mushlin A, Morse S. National guidelines for mass prophylaxis using dispensing/vaccination centers (DVCs). Washington, DC: Department of Health and Human Services, Agency for Healthcare Research and Quality; 2003.

CHAPTER 9

Evolving Public Health Approaches to the Global Challenge of Foodborne Infections

Robert V. Tauxe, MD, MPH, Michael P. Doyle, PhD, Tanja Kuchenmuller, Jørgen Schlundt, MD, and Claudia E. Stein, MD, MSc, PhD

INTRODUCTION

Foodborne diseases result from the ingestion of contaminated foods and food products and include a broad group of illnesses caused by pathogens, chemicals, and parasites, which contaminate food at different points in the food production and preparation process. Diarrheal diseases alone, a considerable proportion of which is foodborne, kill 1.8 million children every year worldwide.[1] Although most of these diarrheal deaths occur in poor countries, foodborne diseases are not limited to developing countries, nor to children. It is estimated that in the United States foodborne diseases result in 76 million illnesses, 325,000 hospitalizations, and 5,000 deaths each year.[2] The resulting medical costs and productivity losses are in the range of $6.6 billion to $37.1 billion.[3] This cost estimate is conservative and does not cover the full spectrum of foodborne diseases but merely seven important foodborne infections in the 1996 U.S. population of 250 million. The full extent of the burden and cost of unsafe food is still unknown, but its impact on global health, trade, and development is likely to be profound.

THE GLOBAL CHALLENGE

Growing international trade, migration, and travel accelerate the spread of dangerous pathogens and contaminants in food, thus increasing universal vulnerability. In today's interconnected and interdependent world, local foodborne disease outbreaks have become a potential threat to the entire globe. In 1991, the Latin American cholera epidemic, which was thought to have begun with contaminated water and seafood

This chapter was originally published as R. V. Tauxe et al, "Evolving Public Health Approaches to the Global Challenge of Foodborne Infections," *Int. J. Food Microbiol*, Vol. 139, Suppl.1; S16–S28. Reprinted, with permission, from the International Journal of Food Microbiology, Volume 139 © by Elsevier, www.elsevier.com.

in Peru, rapidly spread across Latin America resulting in approximately 400,000 reported cases and more than 4,000 reported deaths in several countries.[4]

Through the globalization of food marketing and distribution, contaminated food products can affect the health of people in numerous countries at the same time. The identification of one single contaminated food ingredient can lead to the recall of literally tons of food products, to considerable economic losses in production and from trade embargoes, as well as damage to the tourist industry. In early 2008, an outbreak of avian influenza in Bangalore, India, led to an import ban of Indian poultry products in the Middle East, resulting in losses totaling hundreds of thousands of U.S. dollars to the Indian economy.[5] Some global foodborne diseases result from exports of food or food animals from developed countries, for example, the global spread of *Salmonella* serotype Enteritidis (now the most important *Salmonella* serotype in most countries) and potentially the spread of multiply antibiotic-resistant strains of *Salmonella* serotype Typhimurium.[6]

Foodborne diseases can spread faster, once they emerge. A recent estimate suggests that approximately 30 percent of all globally emerging infections over the past 60 years included pathogens commonly transmitted through food.[7] This is compounded by the growing industrialization of food production, which catalyzes the appearance and spread of new or antibiotic-resistant pathogens. Beyond its macroeconomic effects, foodborne diseases impede development at a country level. Without concerted action to reduce the burden of foodborne diseases, international efforts to achieve the Millennium Development Goals (MDGs), including the overarching goal of poverty reduction, will be jeopardized, particularly those goals relating to children and the poor. Pregnant women are especially susceptible to infectious risks, including foodborne diseases. Infection with Listeria mono-cytogenes and Toxoplasma gondii are particular hazards in pregnancy. These infections can cause serious illness in the mother and fetus, as well as miscarriages, premature delivery, and stillbirth, all of which increase the risk of maternal mortality. People living with HIV/AIDS are prone to opportunistic infections, including those resulting from contaminated food. Serious complications and chronic symptoms as a result of foodborne diseases, including cerebral toxoplasmosis, are common in immunocompromised people.

RECENT MICROBIOLOGICAL FOOD SAFETY CHALLENGES

The landscape of foodborne infections is in flux. New pathogens emerge, established pathogens may acquire new characteristics and appear in unexpected food vehicles, and many existing problems remain unsolved. Consumer tastes and requirements change, populations age and migrate, and the technologies and trade that go into the products are changing as well. Protecting the public health and minimizing the burden of foodborne illness require expecting the unexpected, and being prepared to understand the unexpected when it occurs.

In the United States, the past decade has been filled with more microbiological food safety challenges for the food industry than ever before. Many of these microbial threats were not new but rather presented themselves under unanticipated circumstances. Foods such as peanut butter and microwavable products like turkey pot pies and chicken nuggets have been consumed for decades before *Salmonella* contamination was discovered through major outbreaks of illness. Even ready-to-cook pizzas were associated with an outbreak of *Escherichia coli* 0157:H7 infection.

Commercially canned foods in the United States have for more than three decades had an unscathed record of safety, but recently were associated with illness due to *Clostridium botulinum* toxin. Fresh vegetables have long been considered among the safest of foods but recently have been recognized as the leading U.S. vehicle of illnesses associated with recent foodborne outbreaks. Fresh leafy produce that is ready to eat has been responsible for at least 12 outbreaks of *E. coli* 0157:H7 or *Salmonella* infections since 1998. Considering the emphasis that reducing *Salmonella* contamination of poultry and meat products has received, relatively little progress has been made in reducing the incidence of human cases of salmonellosis during the past decade. Although poultry, eggs, and meat are important vehicles of salmonellosis, additional vehicles that are largely sporadic, and most often associated with undercooked poultry, have continued unabated after a modest decline in the late 1990s.

In addition to issues previously recognized and well-established foodborne pathogens, many additional microbial threats will challenge the food industry and public health in the coming years. Emerging and reemerging pathogens are finding their way into the food supply or are becoming greater foodborne threats. Another emerging issue confronting many countries is the growing population of exquisitely sensitive individuals who have reduced immunity to counteract foodborne infections. This increased sensitivity to foodborne microbes can widen the spectrum of food-associated pathogens to include opportunistic pathogens that frequently are associated with soil and vegetation, including fresh produce. An additional growing food safety threat that has been under the radar, but becoming a heightened concern, is food exported by countries that do not use adequate sanitary practices for food production and processing. Examples include countries that grow vegetables with soil amendments or irrigation water containing untreated human or animal manure. Produce grown under such conditions can be contaminated with pathogens for weeks to months.

LESSONS LEARNED: CONTROLLING FOOD SAFETY HAZARDS IN THE UNITED STATES

Among the earliest of recognized foodborne pathogens, *Salmonella* enterica serotypes continue to be major etiologic agents of foodborne illness. In the United States, salmonellae are among the most prevalent causes of bacterial foodborne disease, with more than an estimated million cases annually.[2] The U.S. Centers for Disease Control and Prevention (CDC) Foodborne Diseases Active Surveillance Network (FoodNet), which has tracked annually the prevalence of salmonellosis in the United States, has reported less than a 10 percent reduction in human cases between 1996–1998 and 2007.[8] The 2010 U.S. National Health Objective is 6.8 reported cases of salmonellosis/100,000 persons; however, as of 2007, the rate was 14.9 reported cases/100,000 persons.[8] Case-control studies of sporadic cases of salmonellosis from FoodNet data have revealed primary risk factors for *S. Enteritidis* infections from 2002 to 2003 included eating chicken outside the home (35 percent population attributable fraction, PAF), traveling internationally (30 percent PAF), and eating undercooked eggs inside the home (15 percent PAF).[9] Eating eggs prepared outside the home was also the primary risk factor for acquiring *S. Heidelberg* (39 percent PAF) and multidrug resistant (MDR) *S. Typhimurium* (13 percent PAF) infections.[10, 11]

Considerable attention has been paid by the meat and poultry industry, in part at the behest of the U.S. Department of Agriculture (USDA) through its Pathogen

Reduction Rule, in reducing *Salmonella* contamination of fresh meat and poultry. Results of the USDA-FSIS (Food Safety and Inspection Service) Pathogen Reduction/HACCP (Hazard Analysis and Critical Control Point) Verification Testing Program reveal, with the exception of ground chicken, major reductions in *Salmonella* contamination compared with the original baseline of pre-1998 levels. Although these trends generally have been favorable, they are not fully informative as many *Salmonella* serotypes associated with meat and poultry are not frequent causes of human illness. For example, *S. Kentucky* is one of the most common serotypes associated with broiler chickens (17 percent of isolates from that source), but is an infrequent cause of human illness (0.1 percent of isolates from humans).[12] More emphasis is needed in further differentiating *Salmonella* isolates from human cases and animal and environmental sources to better identify sources of human infections. This will enable more effective uses of resources to prevent human disease.

Recent outbreaks of salmonellosis associated with microwavable ready-to-cook foods have exposed a major weakness in the U.S. food safety net. Outbreaks have been attributed to contaminated breaded, sometimes prebrowned chicken nuggets and chicken entrees [13–16] and turkey pot pies.[17] Although these foods are intended to be fully cooked by microwave or oven heating, testing of microwave ovens has revealed that many do not deliver the specified energy output. Furthermore, many consumers do not fully cook such products before consumption, but rather warm them to a good temperature for eating that may not be sufficient to kill pathogens like *Salmonella.* Microwave equipment manufacturers, food processors, and consumers need to better understand and address the risks associated with ready-to-cook foods or other options may need to be considered for foods intended to be heated in the home with microwaves.

Recently added to the list of unanticipated foods associated with salmonellosis is the processed food peanut butter, made from roasted peanuts. Although peanut butter was linked to a *Salmonella* outbreak in Australia in 1996, its otherwise long history of safety had given many food safety experts an assurance of being of low risk.[18] An outbreak of 628 reported *S. Tennessee* infections in 47 states from August 2006 to May 2007 heightened concerns regarding the microbiological safety of peanut butter.[19, 20] The outbreak strain was isolated from several jars of peanut butter and contributing factors to the outbreak included water from a leaky roof and a sprinkler system in the oven roaster area. This outbreak illustrates not only how large and widely distributed an outbreak can occur when a highly stable product with national distribution is contaminated with harmful microbes, but also that even products with long histories of safety can be tainted when perturbations exist in the food safety net.

Similarly, until recently, commercially canned foods had enjoyed a long history of safety and not had an outbreak associated with such products since the 1970s. During 2007–2008, however, major recalls of commercially canned foods of two different companies resulted following an outbreak of botulism and the finding of viable spores of *C. botulinum* in products.[21, 22] In one instance, nine cases of botulism were reported from eating contaminated hot dog chili sauce [23] that was underprocessed and, in the other, many canned products, including beans, peas, and asparagus that were contaminated with *C. botulinum* spores [24] likely from post-process cooling water. The widespread distribution of these products prompted

national recalls, which were preventable had good manufacturing practices and federal regulations been applied. These may be isolated incidents but are more concerning examples of an infrastructure that is breaking down because of lack of sufficient oversight.

Fresh vegetables and fruits have during the past decade come to the forefront as important vehicles of foodborne illnesses, accounting for 13 percent (713/5,416) of reported outbreaks between 1990 and 2005 with an identified food source.[25] Salad greens, lettuce, sprouts, and melons were the leading vehicles of illness, with norovirus, *Salmonella*, and *E. coli* 0157 as the most frequently identified pathogens. Although increased consumption of fresh produce and better surveillance and detection of foodborne outbreaks are likely contributing factors to the increased recognition of vegetables and fruits as vehicles of illnesses, CDC studies have revealed that the increased occurrence of outbreaks associated with leafy greens cannot be accounted for by these two factors alone.[26] Major outbreaks have during the past decade been associated with fresh-cut, bagged produce, with 18 such outbreaks (mostly leafy greens) occurring between 1998 and 2006. Preparing fresh-cut produce involves cutting, shredding, dicing, or peeling, thereby wounding the plant tissue. Microbes attach more easily to cut or bruised surfaces than intact produce, cut produce surfaces release large amounts of nutrients that are readily utilized by attached bacteria, and cut plant tissue and leaking juices interfere with disinfectant bactericidal activity during the initial wash steps.[27] Studies have revealed that disinfectants such as chlorine typically used to wash fresh-cut produce reduce pathogen contamination by only 10- to 100-fold, hence neither processors nor consumers have available treatment to ensure the microbiological safety of fresh produce contaminated with large populations of foodborne pathogens. With fresh produce being increasingly responsible for outbreaks of foodborne illnesses, more effective food safety interventions are needed throughout the production, processing, and distribution of fresh vegetables and fruits.

ADDRESSING THE CHALLENGE OF NEW AND EMERGING PATHOGENS

When a new foodborne pathogen emerges, a number of questions emerge along with it.[28] Some are clinical, concerning the range of illness that it causes, when to suspect and diagnose it, the pathologic processes involved, how best the infection can be treated, and to which sequelae it might lead. Some are microbiological, concerning the means of detecting the pathogen in samples from patients, foods, and other environments, and how to measure the immune response it evokes, as well as defining the virulence characteristics that cause the illness. Some are epidemiological, concerning the frequency of the infection in the population, establishing the route, likely vehicles of and risk factors for transmission, the infectious dose and incubation period, and the likelihood of transmission from the infected person to another. Some are concerned with prevention and risk management options. Vaccination is rarely an option for foodborne infections, so prevention usually depends on interrupting transmission before exposure occurs, or controlling the pathogen at its source. Understanding transmission well enough to interrupt it can be difficult, and it is likely to require determining the potential reservoirs and ecologies of the pathogen, the mechanisms by which it may contaminate the food supply, the frequency with which various food and other sources, and the impact of processing,

treating, and handling the food. Some are economic, including the social burden of the infection, the effect of poverty on food choices, and the costs of various control and prevention strategies.

When an established pathogen appears in a new food vehicle, many of the same questions arise. The appearance of *Salmonella Enteritidis* inside intact eggs, of *Yersinia enterocolitica* in lettuces, of norovirus in frozen raspberries, or of *E. coli* 0157 in fresh-bagged spinach has triggered important questions about how to improve control and prevention.[29–31]

Answering the questions and improving prevention is a cyclical process of continuous improvement. When formal or informal surveillance detects a new problem, an epidemiological investigation may clarify something important about its etiology, source, and impact. In some cases, particularly if the pathogen and its transmission are already well understood, control and prevention measures can be directly applied. In other cases, applied research is needed to answer questions raised by the appearance of the infection and the epidemiological findings, and the results of that research can again be used to develop and implement control and prevention strategies. In either case, surveillance helps document that the control and prevention effort has been successful. With successive turns of the cycle, one hopes to see the incidence of the infection decrease, as prevention becomes increasingly successful. Assembling this information is time-consuming, often fragmented, and may take decades. Information may be gathered during outbreak investigations, in specialized research studies, and from public health surveillance. With greater attention and support, the cycle of public health control and prevention can spin faster.

THE IMPORTANCE OF PUBLIC HEALTH SURVEILLANCE

Routine surveillance of the notifiable infections, and subtype-based surveillance of particularly important pathogens, is reinforced by informal communication between caregivers, microbiologists, epidemiologists, and other public health authorities about unusual events in general.[32] This surveillance in itself is not research, and should be supported as part of the routine cost of basic public health. Some surveillance efforts are conducted throughout a country and provide a general net for the entire population. Other surveillance focusing on characterization of sporadic cases by gathering more detailed information is based on sentinel sites thought to be reasonably representative of the population as a whole; examples in the United States include the hepatitis Sentinel County surveillance system and FoodNet.[33, 34]

Public health surveillance for human illness is highly dependent on the local norms of clinical and microbiological diagnosis, which may miss whole categories of pathogens. For example, norovirus is known to cause substantial clinical burden and is a common cause of outbreaks of acute gastroenteritis in many developed countries.[35] Because it is not routinely diagnosed virologically in clinical laboratories, however, illness estimates must depend on special surveys, not on routine surveillance. Surveillance is constantly evolving, as new diseases emerge and are judged worthy of notification, as new diagnostic tests change the ease and specificity of routine diagnosis, and as social interest in particular issues waxes and wanes.

Routine surveillance, paid for by local jurisdictions, will always vary from one jurisdiction to another. Constant dialogue across linked states or provinces can help make definitions consistent, but consistent comparisons of rates of a disease across geographic divisions can be dogged by artifacts that are related to differences in the underlying system that produce surveillance data, rather than by the differences in actual incidence.

In the United States, the FoodNet system has been extremely useful to generate consistent and comparable information despite the variation in local notification requirements to monitor the burden, and to conduct swift assessments of new threats. Similar active surveillance is part of the OzFoodNet system in Australia as well as a new foodborne disease surveillance system being formed in Japan.[36, 37] Other multistate or multinational jurisdictions may wish to consider something similar in order to have transjurisdictional burden and trend data that are robust and interpretable.

The standardization of subtype-based surveillance for the SATEAC (Satellite Energy Advice Centre) *Salmonella* and *Listeria* in PulseNet has inspired the formation of other subtype or genotyping systems, including for *Campylobacter*, norovirus, and hepatitis A. For each of these, the substantial effort required to develop standard methods, subtype databases, and the commitment to rapid routine subtyping needs to be justified by the detection of outbreaks that otherwise would have gone unnoticed. An early indicator of utility is the identification of unsuspected links between outbreaks, such as those observed for the 1998 constellation of shigellosis outbreaks that all proved to be related to parsley imported from one farm in Mexico, hepatitis A infections traced to green onions from a source in Mexico in 2003, and outbreaks of norovirus infections traced to imported frozen raspberries from Poland in Denmark.[38–39]

THE IMPORTANCE OF THE OUTBREAK INVESTIGATION

The investigation of a large foodborne outbreak can lead to the most rapid advances in knowledge. Many foodborne pathogens were discovered in the setting of an outbreak, and a series of cases made it possible to search for the organism systematically and to fulfill several Koch-Henle postulates at once.

Outbreaks are important opportunities to gather new information about how transmission occurs, for both new and emerging pathogens and well-established ones. Although an outbreak is by definition an unusual event, they do not represent random noise. When a group becomes ill following some exposure in common, an outbreak is a surprising and ordered event. By defining the drivers of that order, we learn to prevent outbreaks.[40]

The foodborne outbreak investigation is a major activity in public health, a rapidly conducted scientific endeavor that can lead to direct preventive action. In the past, regulatory authorities often waited to take action against a particular food, until the pathogen was demonstrated in the implicated food. In recent years, in the United States, it has become routine for a well-conducted epidemiological investigation to trigger control and preventive measures by regulatory authorities, even without demonstration of a pathogen in a suspected food.

Outbreaks can be local or widespread. With growing use of subtype-based surveillance networks, such as PulseNet in the Untied States and other locations around the world, and EnterNet in Europe, more geographically dispersed outbreaks are detected.[41] These dispersed outbreaks often reflect contamination events early in the production and distribution of a food, events that otherwise may be hard to detect, and that can have a large impact on the public health. Detecting those events is important to identifying and addressing gaps in the food safety system. A gray zone exists between truly sporadic cases and the highly dispersed outbreak, and it is possible that many cases of apparently sporadic illness actually are part of highly dispersed outbreaks.

At the local level, rapid detection and investigation of outbreaks depends on local capacity, attentive use of surveillance, and good informal communications between care providers, clinical laboratories, and the public health authorities. This also is the case at more central levels of surveillance and includes the food and agriculture epidemiological approaches across the several jurisdictions. At the outset, size distribution and source may all be unclear—it may not even be clear that there is a common source of clusters of infections. There is a dynamic tension between the competing needs. It is critical to consider a broad array of possible exposures to generate hypotheses, and it also is critical to rapidly gather detailed exposure and source information about each specific food, although it can be difficult to do both in one interview with limited resources. While no one solution has been applicable to all settings, the trend has been to use more standard "trawling questionnaires" early in the evaluation of a suspected cluster and to use open-ended questions to collect additional hypotheses that can be added to subsequent interviews.

Though the inherent delays in outbreak recognition and investigation mean that it is sometimes possible to implicate and recall a suspect food while it is still in commerce, careful investigation of some outbreaks can be critically important to the entire industry to learn how to prevent the next outbreak from occurring. Given the advanced technologies now found in the food industry, the complex ecologies that may be operating at the farm or factory level, and the potential importance of applying control measures at the source, it would be highly appropriate to include several disciplines on the investigative team. As with the investigation of air crashes, it may be productive to include engineering or scientific experts from the relevant corporations on the investigative team as well, providing they can be insulated from legal or corporate consequences of sharing their expertise.

FUTURE TRENDS

It is difficult to predict future trends regarding the microbiological safety of foods with confidence. However, several issues are emerging that are certain to be challenges for food safety professionals in the coming years. It is highly probably there will be many newly emerging as well as reemerging foodborne pathogens. Major demographic changes in the world's population are projected to occur during the next 50 years as the world's elderly population grows.[27] Estimates indicate there will be in 2050 three times more elderly (age 65 years or older) than were alive in 2002, accounting for 17 percent of the global population.[44] For many reasons, including weakened immune systems, more underlying illnesses, decreased protection by

vaccines, longer hospitalizations, permanent catheterization, decreased absorption of nutrients, renal insufficiency, and problems with drug interactions, the elderly are at increased risk to pathogenic microbes.[42]

Immunocompromised persons form another population at increased risk of foodborne illnesses. These individuals have increased susceptibility to infections, a greater likelihood of more severe illnesses, including death, and increased potential for illness with an opportunistic pathogen.[43] Conditions of the immunocompromised that predispose them to foodborne infectious agents include inflammatory bowel disease, malignancy, immunosuppressive medications, chronic medical conditions, and HIV/AIDS.[44] The global immunocompromised population continues to increase due to the HIV/AIDS epidemic, life-prolonging treatment of immunodeficiency diseases, and the use of chemotherapeutics and immunosuppressive drugs in cancer and transplant patients. This is a public health concern that cannot be ignored considering the magnitude of this emerging issue.

Food imports are increasing for many countries at unprecedented rates, as the supply of fresh food becomes year-round. Microbiological food safety issues associated with imported foods include inadequate sanitary practices used for food production and preparation in many countries exporting perishable foods and the movement of pathogens from areas where they are indigenous to other areas where they are rare or do not exist. Foods like fresh produce, or fresh and frozen seafood that are handled as ready to eat, are of particular concern. Produce can become contaminated from a variety of sources, including sewage and manure used as a soil amendment or through environmental contamination, contaminated water used to spray plants or in processing, and poor hygienic practices of infected food handlers.

Recognizing the current data gaps, the World Health Organization (WHO) convened a group of international scientists at the "WHO Consultation to Develop a Strategy to Estimate the Global Burden of Foodborne Disease" in September 2006.[45] The consultation provided the strategic framework for the assessment of burden of foodborne diseases, including a road map for assembling existing information on the disease burden and a time frame outlining the individual strategic activities. The consultation marked the formal launch of the Initiative to Estimate the Global Burden of Foodborne Diseases from all major causes using summary health metrics that combine morbidity, mortality, and disability in the form of a Disability-Adjusted Life Year (DALY).

One of the major recommendations of the 2006 consultation was the establishment of the Foodborne Disease Burden Epidemiology Reference Group (FERG), which is charged with implementing the recommendations of the consultation and estimating the global burden of foodborne diseases. Due to the multifactorial nature of foodborne diseases, the FERG is highly multidisciplinary and includes a large number of members. All FERG products undergo a rigorous peer-review process involving scientists outside the expert group to ensure the highest quality and policy impact.

CONCLUSION

Foodborne diseases caused by a broad variety of pathogens and toxins can be expected to continue as a major challenge to public health in the 21st century.

As the structure of the food industry has changed and globalized, so too may the likely challenges change and be increasingly international. Thus food safety, which has long been primarily a local and highly industry-specific effort, also will need to evolve. Robust and harmonized multijurisdictional surveillance, increasingly based on molecular and gene sequence subtyping, can identify the widespread outbreaks. Investigation of those outbreaks can identify gaps high in the food chain, which may reflect systemic problems in need of correction. Developing models to attribute the burden of foodborne illness across the pathogens, across commodity types, and at different levels along the food chain will require a long-term effort of great conceptual importance. Long-term trends likely to influence food safety in the next 50 years include the aging of the population, leading to more people in the high-risk groups; the growing global trade in foods, bringing greater interdependence on the food safety systems of a number of countries; and the appearance of new and unexpected pathogens in foods. We can meet these challenges with robust national surveillance systems; programs to more quickly identify, characterize, and learn about the prevention of new food safety threats; and coordinated global attention to the public health challenges of foodborne disease.

REFERENCES

1. World Health Organization. *The World Health Report, 2007*. Geneva, Switzerland: World Health Organization; 2007. http://www.who.int/whr/2007/en/index/html. Accessed August 16, 2008.
2. Mead P, et al. Food-related illness and death in the United States. *Emerg Infect Dis*. 2007;5:607–625.
3. Buzby J, Roberts T. ERS updates: US foodborne disease costs for seven pathogens. *Food Rev*. 1996;19:20–25.
4. Pan American Health Organization. Cholera in the Americas: epidemiological bulletin of the Pan American Health Organization; 1995;16(2). http://www.paho.org/english/sha/epibul_95-98/be952choleraam.htm.
5. International Food Safety Authorities Network. Verification request: poultry product exports 2008. Geneva, Switzerland: INFOSAN; 2008.
6. Tauxe, R. Salmonella Enteritidis and Salmonella Typhimurium DT104: successful subtypes in the modern world. *Emerg Infect*. 1999;3(3):37–52.
7. Jones KE, et al. Global trends in emerging infectious diseases. *Nature*. 2008; 451:990–993.
8. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food-10 states. *MMWR*. 2007;56.
9. Kimura SC, et al. Chicken consumption is a newly identified risk factor for sporadic Salmonella enterica serotpe Enteritidis infections in the United States: a case-control study in FoodNet sites. *Clin Infect Dis*. 2004;38(Suppl 3):S5244–S5252.
10. Glynn MK, et al. Prior antimicrobial agent use increases the risk of sporadic infections with multidrug-resistant Salmonella enterica serotype Typhimurium: a FoodNet case-control study, 1996–1997. *Clin Infect Dis*. 2004;38(Suppl 3):S5227–S5236.
11. Shiferaw B, Angulo FJ. Egg consumption is the principal risk factor for sporadic Salmonella serotype Heidelberg infections: a case-control study in FoodNet sites. *Clin Infect Dis*. 2004;38(Suppl 3):S237–S243.

12. Sarwari AR, et al. Serotype distribution of Salmonella isolates from food animals after slaughter differs from that of isolates found in humans. *J Infect Dis*. 2001; 183:1295–1299.
13. MacDougall L, et al. Frozen chicken nuggets and strips: a newly identified risk factor for Salmonella Heidelberg infection in British Columbia. *Can J Food Prot*. 2004;67:1111–1115.
14. Currie A, et al. Frozen chicken nuggets and strips and eggs are leading risk factors for Salmonella Heidelberg infections in Canada. *Epidemiol Infect*. 2005;133:809–816.
15. Eglezos S, et al. Bacteriological profile of raw, frozen, chicken nuggets. *J Food Prot*. 2008;71:613–615.
16. Williamson K, Allen G, Bolton FJ. Report of the greater Manchester/Lancashire/Preston PHL Liaison Group survey on the microbiological examination of prepacked frozen raw chicken products in bread crumbs/batter. http://www.tameside.gov.uk/foodsafety/samples/results/thawedchichek.
17. Centers for Disease Control and Prevention, National Center for Zoonotic, Vector-Borne, and Enteric Diseases. Investigation of outbreak of human infections caused by Salmonella. http://www.cdc.gov/salmonella/agona/. Last modified May 14, 2008. Accessed May 22, 2010.
18. Schell W, et al. A South Australian Salmonella Mbandaka outbreak investigation using a database to select controls. *Aust N Z J Public Health*. 1998;22:536–539.
19. Multistate outbreak of Salmonella serotype Tennessee infections associated with peanut butter—United States, 2006–2007. *MMWR*. 2007;56.
20. Food and Drug Administration. Update on Salmonella outbreak and Peter Pan peanut butter and Great Value peanut butter. http://www.fda.gov/bbs/topics/NEWS/2007/NEW1565.html. Published 2007.
21. Ongoing multistate outbreak of Escherichia coli serotype 0157:H7 infections associated with consumption of fresh spinach—United States. *MMWR*. 2006;55:1045–1046.
22. Food and Drug Administration. FDA survey of domestic fresh produce: interim results. http://www.fsan.fda.gov/-dms/prodsur9.html. Published July 31, 2001.
23. Centers for Disease Control and Prevention. Botulism associated with canned chili sauce. http://www.cdc.gov/botulism.htm.
24. Food and Drug Administration. New Era canning company expands nationwide recall. http://www.fda.gov/bbs/topics/NEWS/2008/NEW1795.html.
25. DeWaal CS, Bhuiya F. Outbreaks by the numbers: fruits and vegetables 1990–2005. International Association for Food Protection Poster Presentation. http://www.cspinet.org/foodsafety/prodce%20Outbreaks%20.(IAFP%20Poster%2007.pdf.
26. Herman KM, Ayers TL, Lynch M. Foodborne outbreaks associated with leafy greens, 1973–2006: Abstracts of the 2008 International Conference on Emerging Infectious Diseases; March 14, 2008; Atlanta, GA:81.
27. Doyle MP, Erickson MC. Emerging microbiological food safety issues related to meat. *Meat Sci*. 2006;74:98–112.
28. Holmberg SD, Feldman RA. New and newer enteric pathogens: stages in our knowledge. *Am J Public Health*. 1984;74:205–207.
29. St. Louis ME, et al. The emergence of Grade A eggs as a major source of Salmonella Enteritidis infections: implications for the control of salmonellosis. *JAMA*. 1988;259:2103–2107.
30. Nuorti J, et al. A widespread outbreak of Yersinia pseudotuberculosis 0:3 infections from iceberg lettuce. *J Infect Dis*. 2004;189:766–774.
31. Ealkenhorst G, et al. Imported frozen raspberries cause a series of norovirus outbreaks in Denmark. *Euro Surveill*. 2005;10(9):E050922.2.

32. The Science of Public Health Surveillance. *Scientific Criteria to Ensure Safe Food*. Washington, DC: The National Academies Press; 2003:28–68.
33. Allos B, et al. Surveillance for sporadic foodborne disease in the 21st century: the FoodNet perspective. *Clin Infect Dis*. 2004(Suppl 3):S115–S120.
34. Bell B, et al. The diverse patterns of hepatitis A epidemiology in the United States: implications for vaccination strategies. *J Infect Dis*. 1998;178:1579–1584.
35. Kroneman A, et al. Data quality of 5 years of central norovirus outbreak reporting in the European Network for foodborne viruses. *J Public Health*. 2008;30:882–890.
36. Ashbolt R, et al. Enhancing foodborne disease surveillance across Australia in 2001: the OzFoodNet Working Group Communicable Disease Intelligence. *Commun Dis Intell*. 2002; 26:375–406.
37. Kubota K, et al. Estimation of the burden of diarrheal diseases in Miyagi Prefecture, Japan, 2005–2006: Abstract P3-75: 95th Annual Meeting of the International Association of Food Protection; August 3–6, 2008; Columbus, OH.
38. Naimi T, et al. Concurrent outbreaks of Shigella sonnei and enterotoxigenic Escherichia coli associated with parsley: implications for surveillance and control of foodborne illness. *J Food Prot*. 2003;66:535–541.
39. Amon J, et al. Molecular epidemiology of foodborne hepatitis A outbreaks in the United States, 2003. *J Infect Dis*. 2005;192:1323–1330.
40. Reingold A. Outbreak investigations: a perspective. *Emerg Infect Dis*. 1998;4:21–27.
41. Tauxe R. Molecular subtyping and the transformation of public health. *Foodborne Pathog Dis*. 2006;3:4–8.
42. Ohlsen K, Hacker J. Infections in the elderly. *Int J Med Microbiol*. 2005;294:471–472.
43. Neill MA. The face of change: demographic trends in the US population and implications for foodborne illness: Annual Meeting of the International Life Science Institute; January 17, 2005; Washington, DC.
44. Trevejo T, Barr MC, Robinson RA. Important emerging bacterial zoonotic infections affecting the immunocompromised. *Vet Res*. 2005;36:493–506.
45. World Health Organization. WHO Consultation to develop a strategy to estimate the global burden of foodborne diseases. Geneva, Switzerland: World Health Organization. http://www.who.int/foodsafety/publications/foodborne_disease/burden_sep06/en/index.html.

CHAPTER 10

Managing Benefits and Risks of Pharmaceutical Drug Treatment

Joanna F. Haas, MD, MS, FACP, FACPM, FISPE

No effective treatment is without its risks. Indeed, drug risks make for good headlines, capturing the attention of patients, prescribers, lawyers, and investors. Recently, perhaps prompted by well-publicized pharmaceutical cases (for example, Vioxx, hormone replacement therapy), safety problems associated with drug treatment are in the public consciousness.[1] Part of the problem is that television commercials and other media sources begin with glowing reports of a drug's benefit and end with a long list of potential side effects. Even though the Food and Drug Administration (FDA) puts out preliminary drug safety advisories, such information usually does not reach the consumer. In the past, mention of drug risks was perfunctory at best, now one could argue that the pendulum has swung far in the other direction.

Targeted, effective, disease-modifying drugs for common diseases are a product of the second half of the 20th century.[2, 3] For example, sulfonamides, sometimes called simply sulfa drugs, were the first "miracle drugs" widely available only since the 1930s, and penicillin since 1945 when sufficient stock for mass use could be produced. Pharmaceutical treatments for most 21st century chronic diseases were not yet developed or in widespread use for much of the 20th century. With the development of effective drugs for chronic diseases, a new era of pharmaceutical therapy was ushered in. Today, such drugs are essential to, and synonymous with, medical treatment. But, as exposure to pharmaceuticals becomes ubiquitous, so does the potential for drug-related harm. As such, the risks and harms associated of medical products must be evaluated in terms the benefits and the risks of treatment. The goal is to consistently get the right drug, to the right patient, for the right disease, at the right time, and at the right dose. Given the immense global market for pharmaceuticals composed of more than 5,800 active substances in 128,000 formulations, made by 10,900 manufacturers, this is a Herculean task.[4]

Individuals know that invasive therapeutic interventions such as surgery carry risks.[5] Although there is widespread understanding that drugs have risks as well as benefits,[6] many people still believe that if a drug has been approved by the

FDA it has no risks. At the other extreme are patients who refuse needed medicines out of fear of possible side effects.[7] Communicating a functional understanding of subtle, variable, and unpredictable drug treatment–related risks is difficult, yet achieving optimal therapeutic results requires that patients have a balanced understanding of the benefits and risks of the drugs that they are prescribed. This chapter focuses on understanding how the benefits and risks of pharmaceuticals are managed. The implications for global drug safety are considerable.

SAFETY ASSESSMENT DURING DRUG DEVELOPMENT

Marketed drugs emerge from an extensive, highly structured development process lasting years. The paradigm for risk evaluation during drug development begins with characterizing toxicity in nonclinical models and then in humans (see table 10.1). All drugs undergo extensive preclinical testing before the first human exposures. Animal toxicological studies, of relatively standardized design, are undertaken to identify major potential target organ toxicities as well as to find a dose that does not produce deleterious effects in animals, the so-called NOAEL (No Adverse Effects Level). A dose substantially lower than the NOAEL is selected for the first human exposures in Phase I studies.

Table 10.1.
Safety Assessment during the Drug Life Cycle

Life Cycle Phase	Typical Safety Assessment Activities
Preclinical	Acute toxicology in two species; Short term, repeat dose toxicity studies; Pharmacology and metabolism; Mutagenesis, genotoxicity; Reproductive studies; Carcinogenicity studies in animals
Phase I	Single dose studies, including dose escalation, pharmacokinetics; Repeat dose escalation, pharmacokinetic studies; Identification of maximum tolerated dose (MTD), acute and repeat dose
Phase II	Proof of concept: biomarkers or clinical response as signs of efficacy; Tolerability and safety at doses needed for efficacy; Identify doses and mode of administration for future trials
Phase III	Controlled studies to confirm efficacy and safety; Safety database of sufficient size and at appropriate exposure levels; Basis of marketing approval application
Phase IV	Further characterize effectiveness and safety within the approved indication
Postmarket Surveillance	Evaluation of safety in real life usage; Identification of rare adverse reactions; Safety profile in previously unstudied populations; Safety experience outside the approved label (off-label usage)

Source: Haas

Phase I studies are performed in healthy human volunteers unless anticipated risks make that unacceptable, in which case a target patient population is used, as is often the case for cancer drugs. The goal of Phase I studies is to understand tolerability and adverse effects associated with increasing doses of the drug and to study pharmacokinetics and patterns of drug absorption, distribution, metabolism, and excretion (ADME). Once preliminary evidence indicates adequate safety and tolerability at a plausible dose, Phase II studies are begun on patients with the target condition. These proof-of-concept studies focus on finding some evidence supporting a therapeutic effect whether pharmacodynamic, biomarker, or clinical improvement. Such studies also guide decisions on how the drug will be administered in subsequent studies (dose, route, frequency). A successful Phase II program will identify one or more doses that are likely to be efficacious and well-tolerated and will inform the design of "pivotal trials" that are the basis for drug approval.[8]

The Phase III program must use adequate and well-controlled studies (pivotal trials) to demonstrate drug efficacy. Pivotal trials provide the information that defines the basis for drug approval: the exact indication, the mode of administration, dosing schedule, the benefit claims that can be made, and the anticipated safety profile. Phase III studies are large, long, complex, and expensive. They are not undertaken unless the likelihood is strong that the drug will demonstrate efficacy. Efficacy endpoints should be extremely well defined in pivotal trials because meeting them is a condition of FDA approval. By contrast, predefined specific safety endpoints are unusual. Safety data are collected using an open-ended approach based on findings from physical examinations and laboratory tests as well as answers to questions such as, "How have you been feeling since your last visit?"

Presentation of adverse event data from clinical trials is largely descriptive, showing the frequency and severity of adverse events in different treatment arms. Adverse events often are displayed by the investigator's assessment of causal relationship to the study drug, although it may be difficult to distinguish adverse drug effects from adverse events that are related to the underlying disease. For example, if a 65-year-old heavy smoker with hypertension and elevated cholesterol has a myocardial infarction while taking a new drug, it is impossible to know whether the cardiac event would have occurred in the absence of the drug.

During clinical development, only relatively frequent adverse reactions can be identified and characterized. Even large safety databases have substantial limits. For example, some 3,000 patients would need to be exposed during clinical development to be reasonably confident of seeing an adverse reaction with a 1 in 1,000 frequency.[9] As a result, when a drug is marketed, infrequent or rare adverse drug reactions may not have been identified. The International Conference of Harmonization (ICH) (an organization that develops standards and harmonizes requirements for medicinal product approval, including international consensus standards for drug development in different regions of the world) suggests a minimum safety database (the number of patients who ought to be exposed to a drug prior to approval) of 1,500 patients with at least 100 patients treated for one year if the drug is intended for chronic treatment.[10] The tendency has been for regulatory agencies to request substantially higher patient exposure numbers during clinical trials, particularly if a large number of people will be exposed once the medication is commercially

available. Nonetheless, the number of patients studied before FDA approval will almost always be very small compared with the number treated postapproval.

Another concern is that patients studied in clinical trials for new drugs are often different from those who receive the drug once it is on the market. Clinical development protocols exclude many patients, notably the very old and those with complex comorbidities.[11] Patients with certain prior conditions, concomitant diseases, or medications may be excluded, especially if these are potentially associated with increased risks. Furthermore, patient monitoring, compliance, and clinical supervision are often more intensive in clinical trials, and overall outcomes in clinical trial patients often are better, regardless of the disease or treatment, than similar patients not enrolled in clinical studies.[12] All of these factors limit the generalizability of premarket experience so that an understanding of the product's safety profile necessarily will continue to develop during postmarketing. Postmarketing safety surveillance is designed to identify safety issues not yet appreciated during clinical development.[13]

Drug development and drug approval processes have become increasingly global. The ICH's purpose, for example, is to foster agreement of standards of preclinical research, clinical trial assessment of efficacy and safety, and management of drug quality and purity.[14] The ICH technical working groups are composed of pharmaceutical industry organization representatives, as well as drug regulatory authorities from the European Union, Japan, and the United States. The standards established within the ICH framework are the basis of a worldwide system of surveillance, collection, evaluation, and reporting of adverse drug experiences. The consultation processes are designed to result in recommendations for common data and documentation standards so that a single drug application can be submitted for marketing approval throughout the world. The recommendations of ICH, however, have no legal standing in their own right. To have the force of law, a recommendation must be incorporated into regional and national pharmaceutical product regulatory systems. Although the ICH process is far from perfect, it has improved the transparency and comparability of applications for market approval and has defined standards for postmarketing safety surveillance.

The ICH guidelines define how information on individual adverse event case reports is categorized and stored in standard data structures utilized in pharmaceutical industry and regulatory agency databases throughout the world. An international adverse event terminology thesaurus known as MedDRA (Medical Dictionary for Regulatory Activities) was developed and adopted in each ICH region. This allows medical concepts to be coded consistently and automatically translated. Free movement of adverse events data was made possible by adoption of a common information technology data exchange protocol. The global network for safety information sharing is a major success of the ICH collaboration.

Analyzing and presenting such information requires careful consideration because the way risks are presented has a substantial effect on the behavior of both health professionals and patients. The next section considers various measures of risk and discusses some factors affecting drug risks' acceptability.

BALANCING BENEFITS AND RISKS OF PHARMACEUTICALS

It is universally accepted that the safety of a medicine must always be seen in the context of its utility, weighing potential benefit against potential risks and tolerability.

Each drug risk is characterized by its nature, severity, and frequency. With the exception of situations in which deaths attributable to treatment can be directly balanced against deaths averted in a similar time frame, the risks and benefits of treatment often are difficult to compare on a common scale. That being said, the magnitude of risk may be expressed in various ways (see table 10.2).[15]

The *relative risk* is the ratio of the event rate in patients receiving a drug compared to that in patients not receiving the drug. A relative risk of three means that the adverse effect occurs three times more often in patients receiving the treatment compared to those who did not receive that treatment. Relative risk does not take into account the absolute frequency of the undesirable effect. Thus, a relative risk of three may represent a risk of an adverse effect in 3 of 100 exposed patients, compared to 1 in 100 in the unexposed, but it also describes the ratio of a risk of 3 in 1 million versus 1 in 1 million. In the first instance a common problem, for example, headache, becomes more common. In the second example, a rare condition, such as liver failure, triples in frequency but remains rare. The two situations are quite different with regard to decision making for the individual patient and also with regard to the potential public health impact. For this reason, relative risk alone is not sufficient to assess safety risk; the actual frequency of the undesirable effect needs to be taken into account.

The difference in risk in the exposed versus the unexposed group provides valuable information about the *absolute risk* averted or incurred (Absolute Risk Reduction [ARR], Absolute Risk Excess). In the example above the difference in absolute rates of liver cancer is two cases per million. There will be two additional cases of liver failure for every million patients treated with the drug: this is the absolute risk excess. This same concept can be used to quantify the benefits of treatment. For example, if a study shows the rate of fatal heart attacks to be 3 per 100 in patients who receive active drug and 7 per 100 in those on placebo, this corresponds to four fatal heart attacks averted for every hundred patients treated: the ARR due to treatment was 4 per 100.

Another way of looking at this information is that the four fatal myocardial infarctions averted represent a 57 percent reduction from the baseline risk (four deaths averted of seven expected). This may be referred to as the relative risk reduction (RRR) due to treatment. The reciprocal of the ARR turns out to be useful in communicating the impact of drug treatment. How many patients will have to be

Table 10.2.
Measures of Risk

Frequency in Exposed	Frequency in Unexposed	Absolute Risk Reduction (ARR) Absolute Risk Excess (ARE)	Relative Risk	Number Needed to Treat (NNT)[a] Number Needed to Harm (NNH)[a]
3 per 100	1 per 100	2 per 100	3	50
3 per 1,000,000	1 per 1,000,000	2 per 1,000,000	3	500,000

Source: Haas

[a.] Needed to avert or produce one outcome.

treated, for example, to avert one death due to myocardial infarction? In the above example, treating 100 patients averted four fatal myocardial infarctions, thus 25 patients will need to receive this treatment to avoid one death. This is referred to as the Number Needed to Treat (NNT). Analogous calculations can be used for risk, often referred to as Number Needed to Harm (NNH). In the liver failure example above, two excess cases occurred for every million patients exposed to the drug. Thus about 500,000 patients would need to be exposed to the drug to have one excess case of liver failure. The concepts of NNT and NNH can help summarize and compare the benefits and risks of a treatment in the context of patient exposures.

Measures of risk are somewhat one dimensional in that they provide no weight to the benefits or harms that they quantify. They do not differentiate between the impact of a headache and a stroke. A widely used measure that attempts to take into account other aspects of treatment "utility" is the Quality-Adjusted Life Year (QALY), a summary measure adjusting quantitative risk measures by associated changes in quality of life. This measure has found particular favor in pharmacoeconomics where it provides a single summary measure of utility that can also be valued, as in "cost per QALY." The British National Institute for Health and Clinical Excellence (NICE) uses this measure to decide which drugs will be offered by the National Health Service (NHS): the higher the cost per QALY, the less the likelihood that the drug would be offered.[15] Under the prevailing NICE standards, treatments that cost less than £20,000 per QALY are generally considered acceptable, whereas those costing above £30,000 require special justification to be offered under the NHS.

The acceptance of drug-related risk is always framed by the disease being treated, the potential benefits of the drug, and available alternative treatments. Major risks may be accepted to achieve even modest benefits if the disease is progressive and unrelenting. For example, substantial drug-related morbidity and mortality is accepted with bone marrow transplant for leukemias that are uniformly rapidly fatal in the absence of successful treatment. Patients and physicians are willing to accept substantial treatment risks and discomfort as a trade-off for a few extra months of life. By contrast, when epidemiologic evidence suggested a small increase in the risk of cerebral hemorrhage in young, first-time users of over-the-counter cold medications that contained phenylpropanolamine, the FDA quickly called on manufacturers to voluntarily remove this ingredient.[16] The increase in relative risk was modest (less than 1.5 for users for over-the-counter cold remedies) and the absolute risk excess was estimated at 1 per 300,000 persons to 1 per 3 million persons. Even though the absolute number of people who might be harmed was very small, it was deemed unacceptable for medications used by millions of healthy Americans seeking symptomatic relief for a self-limited disease.[17]

Acceptability of risks also depends on the availability and nature of alternative treatments. When there are effective alternative treatment modalities, particularly within a drug class, less risk is accepted. Cerivastatin, a statin used to treat elevated blood cholesterol levels, was withdrawn in 2001 when it became clear that the risk of rhabdomyolysis (the rapid, and potentially fatal, breakdown of skeletal muscle) was higher among those taking this drug compared to those taking other statins of similar efficacy.[18] Similarly, troglitazone, the first commercially available

member of the new thiazolidinedione class of insulin sensitizing, antidiabetic drugs, was known to be associated with rare cases of severe liver injury. Yet, it remained on the market until other agents in the same class (rosiglitazone and pioglitazone) became available as the FDA no longer considered it justifiable to keep troglitazone on the market. The drug was removed.[19] Thus, the "safety" of a drug is not an absolute characteristic of a medicine that can be assessed in a vacuum. Acceptance of drug-related risk is always seen in light of its utility and the nature and availability of alternative treatments.

HOW A DRUG CRISIS CAN GALVANIZE CHANGE

In the drug safety world, as in other spheres, it may take a crisis to galvanize major change.[20, 21] A defining experience in the history of pharmacovigilance (pharmacovigilance refers to the detection, assessment, understanding, and prevention of adverse effects of drugs, particularly long-term and short-term side effects of medicines) was the unforgettable sight of thalidomide babies with their flipper-like, vestigial limbs. It did not take a medical professional to recognize that something was wrong with these children, but it took years for the medical profession to recognize the association between in utero exposure to thalidomide and the dramatic and distinctive congenital deformities called phocomelia. Thalidomide was used widely in many countries and was specifically promoted for morning sickness in pregnant women. In the United States, millions of babies were spared this fate thanks to FDA official, Dr. Frances Kelsey, who leveraged the limited powers then available to her and delayed the introduction of thalidomide to the U.S. market. The system of pharmacovigilance created in the wake of this experience was designed to ensure that a tragedy like thalidomide would not be repeated. A key lesson learned was the need to concentrate reports of suspected adverse drug reactions from around the world. Unexpected events that were serious and possibly related to a specific drug were to be reported promptly by the manufacturer to special drug evaluation units in the FDA. Nonetheless, it was not until 1982 that FDA finalized reporting requirements, making pharmaceutical companies responsible to report adverse drug experiences.[1] (See appendix for a listing of milestones in drug safety regulation.)

PROGRAMS FOR DRUG SAFETY

In the 21st century, the pharmaceutical industry is required to evaluate all adverse drug experiences that come to its attention. The FDA receives most adverse drug experience reports via pharmaceutical companies. Each company that holds a license to market a drug or biological product is required to have a system in place

1. "Side effects," the lay term for adverse reactions, is disdained by pharmacovigilance professionals, probably because of its clarity. The terms adverse event, adverse drug experience, and adverse reaction have distinct meanings in pharmacovigilance jargon. "Adverse event" refers to an undesirable occurrence but does not imply any relationship to a drug. "Adverse drug experience" is a term in FDA regulations that refers to an adverse event associated with a drug, but it does not necessarily imply that it was caused by that drug. "Adverse reaction" refers to an adverse event suspected to be causally associated with a treatment. I have used these terms somewhat loosely.

for surveillance, collection, evaluation, and reporting of adverse drug events. Elaborate, highly structured business processes and computer systems support a worldwide system to collect adverse drug experience reports, enter information into specially designed databases, evaluate the medical significance of each case, and report cases to the FDA and other regulatory agencies, as required by law. The FDA regularly conducts unannounced, in-depth inspections of these systems to ensure that they are in compliance with the regulations. Adverse event reporting by health care professionals is voluntarily, however. Case reports may be sent directly to FDA under its MedWatch program. The MedWatch Web site provides information on how to report adverse drug experiences.[22] It has been expanded to serve as a source of up-to-date information on medical product safety concerns. Safety alerts are posted about drug recalls or new safety concerns, and a monthly table summarizes safety-labeling changes for all drug products. Adverse drug experience reports are entered into the FDA's Adverse Events Reporting System (AERS), which now contains more than 4 million reports.[23]

Just as thalidomide led to the creation of new safety initiatives decades ago, a watershed event for pharmacovigilance came in October 2004, when Merck abruptly withdrew its nonsteroidal anti-inflammatory agent, rofecoxib (Vioxx) after an excess risk of cardiovascular events was confirmed in a clinical trial of colon carcinoma chemoprevention.[24] Nothing was distinctive about the heart attacks and strokes that occurred in patients on rofecoxib treatment, and nothing distinguished them from millions of cardiovascular events each year in comparable individuals who had not taken COX-2 inhibitors. Medical practitioners really had no reason to associate a common cardiovascular event in an elderly patient with a commonly used pain medication. In short, this was a drug safety problem that could not be addressed by the systems designed to identify thalidomide-like effects.

Rofecoxib was a member of the class of nonsteroidal anti-inflammatory drugs called COX-2 inhibitors, which included rofecoxib, celecoxib, and valdecoxib. Of the three, only Celecoxib remains on the market as of 2009. Rofecoxib, first marketed in 1999, had been used by an estimated 84 million individuals worldwide before its removal from the market in 2004. This class of drugs had been designed as a "safer" alternative to aspirin and indeed clinical trials confirmed that the risk of gastrointestinal bleeding was lower with these drugs compared with conventional nonselective nonsteroid anti-inflammatory drugs (NSAIDs). These drugs were widely prescribed to the elderly who are known to be at higher risk of gastrointestinal bleeding with NSAID use and also are at higher risk for heart attacks and strokes.

The withdrawal of Vioxx led to fear, confusion, and distrust among patients and physicians. The integrity of the U.S. drug safety system came under scrutiny with both the pharmaceutical industry and the FDA under fire. The industry was viewed as having minimized known risks, and the FDA was viewed as having been less than decisive in taking needed action swiftly, although negotiations to strengthen the warnings about cardiovascular risks related to rofecoxib had been ongoing between the FDA and Merck for more than a year before the drug was pulled abruptly from the market. The technical adequacy of the system was under fire, but more important, the adequacy of the decision-making process was in doubt and the credibility of the FDA was jeopardized.

As a direct result of the furor that arose over the safety of COX-2 inhibitors, the Institute of Medicine (IOM) was charged by the FDA to examine the U.S. drug safety system. The IOM report, issued in September 2006, presented an unsatisfactory picture of FDA postapproval safety systems.[25] The report cited a crisis of confidence undermining the credibility of both FDA and the pharmaceutical industry. Furthermore, imbalances in resource allocation and distribution within the FDA meant that most resources were concentrated in the preapproval phase; quality and transparency of communication about drug safety issues were found wanting; and FDA resources were not at all commensurate with its mandate and responsibilities.

FOOD AND DRUG AMENDMENT ACT OF 2007

Key IOM recommendations became law with the passage of the FDA Amendment Act of 2007 (FDAAA). FDA was given additional regulatory powers with a concomitant mandate to ensure transparent agency decision making. Product label changes for safety were to be imposed and had to be executed promptly. To ensure an acceptable benefit-risk balance, the agency was empowered to require additional studies or trials and it could stipulate specific conditions limiting the market availability of a product to ensure its safe use. The agency was instructed to promptly communicate evolving product safety concerns even if the available information was limited. In addition, FDAAA mandated that virtually all clinical trials, regardless of sponsorship, be registered and that efficacy and safety results be publicly posted in a timely manner.

The FDAAA also gave the FDA the power to impose postmarketing requirements (PMRs) for additional "studies" or clinical trials, on new or already marketed products to ensure a positive benefit-risk relationship.[26] Such studies or trials must focus on drug safety, not effectiveness. For the first time, a formal distinction was made between "studies" and clinical trials. Research in which some aspect of patient experience is determined by the researchers constitutes clinical trials. By contrast, studies do not involve direct experimental manipulation of human experience and include observational research as well as laboratory studies. Under FDAAA, clinical trials testing a safety endpoint can be imposed only if studies cannot provide adequate information.

The FDAAA created a major new safety information system (the Sentinel System) that would complement spontaneous adverse event reporting. The FDAAA directed the creation of a drug safety surveillance network based on electronic medical records (such as the Veterans Administration's electronic patient records) and administrative health insurance claims data. The pilot system with 25 million "lives" was to be up and running on July 1, 2010. The FDAAA stipulates that the system ultimately must cover 100 million patients by July 1, 2012.[27] The FDA has initiated the Sentinel Initiative to fulfill this mandate.[28] By using the experience of tens of millions of people, FDA hopes to speed identification and evaluation of drug-related risks. This approach redefines the nature of drug safety activities. It moves safety surveillance from dependence on submission of spontaneous reports by alert practitioners (the thalidomide model) to a structured analytic system that can be used proactively. Harvesting electronic medical data potentially can be used

to detect a modest excess in the risk of relatively common medical problems, such as myocardial infarction, associated with specific drug exposures.

While the potential is great, the obstacles to success should not be underestimated. These data are derived from different organizations and are generated and structured differently. Commercial, nonprofit, government, or academic data owners have distinct intellectual and proprietary interests. Under these circumstances, it would be futile to try to produce a massive single data set. Instead, a distributed model is being developed whereby parallel queries could be run independently on each database. This avoids privacy concerns associated with transmission of individually identifiable medical data.

The Sentinel Initiative is important to many stakeholders. But, the rules and structures that govern how it functions will be critical to its acceptance and success. Who will set scientific standards, prioritize safety issues, determine data access, and communicate results?[29]

The FDAAA requires the FDA to work closely with public, academic, and private entities to build the Sentinel Initiative framework and to address the complex logistical, technical, and analytic hurdles as well as the ethical, legal, and governance considerations. The actual situation is fluid, and the FDA has taken pains to provide current information and solicit comments via postings on the Sentinel Initiative Web site.[30] Although the Sentinel Initiative is focused on drug safety, it should be viewed in the context of the expanded use of observational data to monitor health care and to provide information for decision making. These data provide insight into drug utilization patterns and they can be used to evaluate the effectiveness of therapies in a real-world setting, which will affect all aspects of medical care, not simply drug risks. Already comparisons of the effectiveness of different therapies may influence drug formulary choices. In a climate of change for health care delivery, pharmacoepidemiological studies should be undertaken to complement clinical trials as sources of key information for decision making.

The FDAAA provides the FDA with new powers to manage drug risks. The first line of risk minimization consists of strengthening the product label. The 2007 act unambiguously allows the agency to require inclusion of "new safety information" in an existing product label and thus avoids protracted negotiations while the product continues to be marketed. The FDAAA empowers the agency to specify content and ensure that important new safety information is presented promptly and clearly to prescribers.[31] The FDAAA also gives the FDA extensive powers to impose conditions on drug marketing and distribution. These conditions, referred to in the aggregate as REMS (Risk Evaluation and Mitigation Strategies), have the potential to profoundly change how drugs are used. In the past, the FDA had two basic tools for managing drug risks: providing information to prescribers or removing the drug from the market. These choices were frequently unsatisfactory. Prescriber behavior is resilient and safety warnings in the label are frequently ignored. On the other hand, withdrawal of the marketing license is an extreme measure that makes the drug unavailable to patients who need it. Given that the choice was between the ineffective and the draconian, the need for more discriminating risk management tools was evident. With REMS, if the FDA determines that a serious risk is associated with a medical product, it can impose restrictions on how it is marketed so that the benefit-risk relationship is acceptable. In taking such actions, the FDAAA requires the FDA to

consider several factors: size of the population exposed to the drug, seriousness of the condition to be treated, expected benefits of treatment, duration of treatment, nature and frequency of known or potential adverse events that may be related to treatment, and whether the drug is a new molecular entity. In addition to a required assessment timetable, risk mitigation efforts may include or be limited to enhanced education elements, such as a patient package insert, a medication guide, and a communication plan for health professionals. An evaluation component may be required to show that such strategies have worked.

Perhaps most important, the FDAAA gives the FDA the power under REMS to restrict commercialization of an approved drug to control a serious safety concern that otherwise would keep the drug off the market. Specific "elements to ensure safe use" might include limiting drug distribution, for example, or certification of prescribers and dispensers. Companies may be required to set up special systems, including patient registries, to collect additional information on benefits and risks. Patient eligibility to receive treatment may be carefully defined and patients may have to provide specific information, such as laboratory test results, to continue treatment.[32, 33] Regardless of the specific measures, REMS requires a reporting timetable.

Experience with the multiple sclerosis drug natalizumab (Tysabri, Elan Pharmaceuticals, Inc.) is illustrative of how such a program can salvage an effective and novel drug with a major safety problem.[34] Natalizumab had been pulled abruptly from the market in March 2005 following three reports of progressive multifocal leukoencephalopathy (PML), a rare and fatal viral infection of the brain. It was returned to the U.S. market in July of 2006 under an elaborate program, dubbed TOUCH™, intended to improve prescribing decisions and to minimize death and disability from PML in patients receiving natalizumab.[35] It incorporates an extensive education program and a controlled-drug distribution program that some regard as a benchmark for managing serious safety concerns under REMS.

How will the FDA use its REMS authority? At present, the majority of REMS programs have involved communication plans or medication guides rather than restrictive "elements to ensure safe use."[36, 37] Before requiring a REMS that involves changes to customary practice, the FDA must take into account a variety of real-world factors. The prospect of a multitude of different REMS programs for different drugs is not a happy one particularly for those involved in dispensing. Moreover, by law, a REMS cannot disproportionately reduce availability of a treatment in less-well-served communities. There should be equity, such that programs to control comparable risks have comparable restrictions. The program cannot be disproportionately costly or complex or excessively interfere with the normal activities of prescribers, dispensers, or patients, and it cannot be used to interfere with the availability of generic products.[38]

DRUG PRODUCTION IS GLOBAL, BUT DRUG REGULATION IS LOCAL

While the FDAAA addresses important drug safety problems, other issues affecting safety continue to emerge. One such challenge hit the headlines only six months after it was signed into law. In February 2008, the FDA issued a public health advisory reporting a substantial increase in the number of anaphylaxis cases, including fatalities, associated with heparin produced by Baxter Health Corporation.[39, 40, 41]

No increase in anaphylaxis cases was noted for heparin manufactured by other companies. Ultimately, the outbreak resulted in hundreds of cases and 81 deaths due to serious allergic reactions.[42] It turns out that Baxter's heparin had been produced in a factory in Changzhou City, China. After several weeks of investigation, the culprit was identified. The material supplied to Baxter had been deliberately adulterated through the substitution of oversulfated chondroitin sulfate for heparin.[43] Oversulfated chondroitin sulfate mimics heparin's activity in standard tests, but costs $9 a pound compared with $900 a pound for heparin. Because the factory produced material only for export and not for domestic Chinese use, it was not subject to inspection by any Chinese regulatory authority.

This episode brought to light a gaping hole in the system of drug oversight. Production of pharmaceuticals is a global process with raw materials now sourced from around the world. Defects in production and problems in product integrity can abruptly create safety issues in otherwise well-characterized drugs. For the FDA and other regulatory agencies, the implications are staggering. The prospect of inspecting thousands of geographically and linguistically inaccessible local factories is not a task for which the FDA or other national agencies are staffed or funded. One small positive step to protect drug quality is a mandated system to identify each drug product to the level of the package and batch. This system makes it possible to control specific drug products on the market and is a significant step in protecting against drug counterfeiting, theft, and adulteration. While by no means a panacea, it may help to locate and isolate defective products.[44]

PERSONALIZING BENEFITS AND RISKS

At present, most information about drug risks and drug benefits refers to effects on large groups of people. What if it were possible to predict who could benefit from specific types of treatment, and who is at the highest risk of possible side effects? Pharmacogenetics, a new avenue for the pharmaceutical industry, could provide a means for a personalized approach to drug benefits and risks. Over the past decade, there has been increasing awareness of the interaction between patient and drug on a genetic level.[45] Already approximately 10 percent of FDA-approved drugs contain information on pharmacogenomics in the product label. For example, one widely used anticoagulant, warfarin, has a narrow therapeutic index: too much and the patient may suffer severe bleeding, too little and no protection is provided from thromboemboli.[46, 47] Patients with specific genetic characteristics metabolize warfarin more slowly and need lower doses. Such patients are at higher risk of bleeding early in treatment. Commercially available tests can identify patients with these genes. However, numerous scientific, practical, and legal issues remain unresolved [48] and the warfarin product label in the United States refers to these tests but does not require their use.

DRUG SAFETY IN THE 21ST CENTURY: THE REMAINING CHALLENGES

Despite efforts to tighten the mechanisms to ensure drug safety, the system is not fail-safe. The goal for drug safety is to have a flexible and responsive system able to recognize potential risks early, collect information efficiently, and take

action that is appropriate in the context of both benefits and risks. Changes mandated by the FDAAA should have a great impact on transparency of drug-related information as well as the way in which drugs reach the market.

Pharmacoepidemiological techniques can utilize the massive amount of data that are generated by the health care system to identify and promptly evaluate emerging safety issues. Information technology will be a powerful and major force in the transformation of drug safety in the 21st century. Scientific progress will be able to dissect disease processes down to the molecular level. Drug treatments will become more targeted and individualized. Using such data constructively to support decision making about drug benefits and risks will translate scientific developments into individually optimized, affordable new drug treatments. Enhanced understanding of genetic characteristics of both the patients and their diseases increasingly will inform prescribing decisions, leading to an improved understanding of the benefits and risks of drug treatment for individual patients. Such changes will have a profound impact on drug development and prescribing. Ultimately, however, improving the balance between benefits and risks depends on better methods of evaluation, decision making, and communication about safety issues.

Yet, with all these promising advances, there still there will be unanticipated safety problems. Even as the FDA works to improve management of drug risks using the powers provided by the FDAAA, a whole new set of concerns emerges as a result of a system of drug development and production that is increasingly global. Harm to patients due to sloppy drug quality oversight remains a serious and important safety issue. Drug development and drug production have become global activities, but most regulatory structures are national and a tremendous gap remains between the task and the available resources for protecting and ensuring drug quality. There is an urgent need to put into place an oversight product to ensure the quality of medicines regardless of where they are produced or used. Such efforts will require a global approach, coordination, and resources.

Issues of drug-related benefits and risks are closely linked to issues of equity, transparency, affordability, and individual choice and are knit into the fabric of the health care system. Changes in the provision and financing of medical services will affect how drugs are used and inevitably will affect benefit and risk both for individuals and for the public at large.

APPENDIX: SELECTED MILESTONES IN U.S. DRUG REGULATION[1]

1820: U.S. Pharmacopeia was founded by representatives of state medical societies, who created a system of standards and quality control as well as a national formulary. Initially 217 drugs were admitted.[2]

1. US Food and Drug Administration: FDA Centennial at CDER: Milestones of Drug Regulation in the United States. May 26, 2006. http://www.fda.gov/AboutFDA/WhatWeDo/History/FOrgsHistory/CDER/CenterforDrugEvaluationandResearchBrochureandChronology/ucm114463.htm.
2. US Pharmacopeia. USP History. http://www.usp.org/aboutUSP/history.html.

1848: The Drug Importation Act of 1848 was intended to prevent the importation of adulterated drugs.[3]

1902: The Biologics Control Act was designed to ensure quality of biologic products, including vaccines, serums, and blood products. Passage followed a much publicized instance of tetanus transmitted through antidiphtheria serum derived from a horse that had died of tetanus.[4]

1906: The Food and Drugs Act of 1906 was passed to prevent adulterated food stuffs and drugs entering interstate commerce. The uproar associated with the graphic portrayal of unhygienic conditions in Chicago meat-packing plants described in Upton Sinclair's novel *The Jungle* contributed to its passage. The focus for drugs was on quality standards and accurate labeling of ingredients.[5] No demonstration of effectiveness or safety was required.

1938: The current system of drug regulation was initiated with the Food, Drug, and Cosmetic Act (FDCA), which required that evidence of drug safety be submitted before a drug could be marketed. The immediate impetus for this legislation was the massive poisoning with at least 107 fatalities, due to Elixir Sulfanilamide, an untested formulation of the antibiotic that contained toxic diethylene glycol.

1962: The Kefauver-Harris Drug Amendments to the FDCA required that safety and effectiveness of a drug be demonstrated before marketing through adequate and well-controlled clinical trials. These amendments created the U.S. drug approval system as we know it.

Although the legislation had been under consideration for years, its passage was a response to widespread publicity about thalidomide, where a near catastrophe had been averted by the Food and Drug Administration (FDA).

1983: The Orphan Drug Act was passed to encourage development of drugs for rare diseases.

1984: The Drug Price Competition and Patent Term Restoration Act (Hatch-Waxman Act) opened the way to approval of generic drugs.

1987: Investigational drug regulations were altered to allow access to experimental drugs for patients with serious and untreatable disease.

1988: Food and Drug Administration Act of 1988 placed the FDA under the Department of Health and Human Services and created the role of FDA Commissioner.

The Prescription Drug Marketing Act (1988) added measures to protect drug supply and quality.

1989: FDA guidelines were issued to increase evaluation of drugs in the elderly.

1991: In response to the emergence of the AIDS epidemic, procedures for the Accelerated Review of Drugs were established for life-threatening diseases.

---

3. Heath WJ. America's first drug regulation regime: the rise and fall of the Import Drug Act of 1848. *Food Drug Law J.* 2004;59(1):169–199.
4. Junod SW. Biologics Centennial: 100 Years of Biologics Regulation. http://www.fda.gov/AboutFDA/WhatWeDo/History/ProductRegulation/SelectionsFromFDLIUpdateSeriesonFDAHistory/ucm091754.htm.
5. Young JH. The long struggle for the law. http://www.fda.gov/AboutFDA/WhatWeDo/History/CentennialofFDA/TheLongStruggleforthe Law/default.htm.

Policies for the protection of human subjects referred to as "the Common Rule" were adopted by FDA and other federal agencies.[6] It extended and standardized practices such as informed consent and the role of Institutional Review Boards (IRBs).

1992: International Conference on Harmonization (ICH) was established to develop common regulatory requirements for drugs and biological products across the United States, European Union, and Japan. ICH standards provide the basis the global adverse reaction reporting system.

The Prescription Drug User Fee Act (PDUFA) of 1992 provided improved resources for FDA drug reviews through a requirement of drug and biologics manufacturers to pay fees in association with regulatory submissions. This was designed to address concerns that long application review times at FDA delayed availability of new drugs in the United States. The legislation was renewed and extended in 1997, 2002, and 2007.

1993: The MedWatch system was launched to enable health professional and consumer reporting of adverse drug experiences to FDA.

1997: Food and Drug Administration Modernization Act (FDAMA) established accelerated review of new medically important products, and regulated advertising for off-label uses of drugs and devices. Incentives were created for clinical studies in pediatric populations.

1998: FDA's Adverse Event Reporting System (AERS) was introduced to improve computerized tracking of adverse event reports.

1999: FDA introduced new strategies for risk evaluation, prevention, and communication in its May 1999 report "Managing the Risks from Medical Product Use: Creating a Risk Management Framework."

In September 1999, the Institute of Medicine issued its report "To Err Is Human," highlighting the public health importance of safety risks from medical treatments.

2002: The Public Health Security and Bioterrorism Preparedness and Response Act directed FDA to protect regulated products from terrorist activities.

The renewal of the Prescription Drug User Fee Act (PDUFA III) permitted the FDA to use these fees for postapproval safety activities.

The FDA issued guidance to industry on ascertaining pregnancy outcomes following maternal drug exposures.[7]

2003: The Pediatric Research Equity Act confirmed FDA's authority to require sponsors to conduct clinical research supporting use of drugs and biologics in pediatric populations.

2004: Project BioShield Act authorized FDA to provide a pathway for approval of treatments for potential terrorist attacks with chemical, biological, or nuclear agents.

---

6. Korenman SG. Teaching the Responsible Conduct of Research in Humans. Office of Research Integrity, RCR Resource Development Program. http://ori.dhhs.gov/education/products/ucla/chapter2/page04b.htm.
7. US Food and Drug Administration. Guidance on Establishing Pregnancy Exposure Registries. August 2002. http://www.fda.gov/ScienceResearch/SpecialTopics/WomensHealthResearch/ucm131182.htm.

Anabolic Steroid Control Act of 2004 introduced stronger measures to decrease misuse of steroids.

The FDA issued regulations to barcode all prescription medicines to reduce medication errors.

The Critical Path Initiative was launched by FDA to help speed the translation of scientific advances into therapeutic breakthroughs.[8]

The FDA issued a public health advisory to limit the use of COX-2 selective agents, which in clinical trials had been shown to be associated with an elevated risk of heart attacks and stroke.

2005: A Drug Safety Board was created within the FDA to provide advice on management and communication concerning drug safety issues.

Three final Guidance for Industry documents were issued concerning evaluation of on periapproval drug safety: "Premarketing Risk Assessment," "Development and Use of Risk Minimization Action Plans (RiskMAPs)," and "Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment."

2006: Requirements on Content and Format of Labeling for Human Prescription Drug and Biological Products rules were approved. Changes in label content and format were revised to make FDA-approved labeling more useful to health professionals.

The Institute of Medicine issued its report on "The Future of Drug Safety," which included recommendations for enhancement of FDA authority over marketed drugs. These recommendations were in part a response to perceptions about how safety issues with COX-2 inhibitors had been managed.

2007: Food and Drug Administration Amendments Act of 2007 (FDAAA) gave extensive new authority to FDA to mandate measures (Risk Evaluation and Mitigation Strategies) to ensure a satisfactory benefit-risk balance for marketed drugs. In addition, a large observational data set (Sentinel Network) was to be created from electronic health records. This system is intended to be used for drug safety surveillance and for evaluation of suspected serious safety risks.

2009: Guidance for Industry documents were issued by the FDA for "Post-Marketing Studies and Clinical Trials" and for "Risk Evaluation and Mitigation Strategies" as part of the process of translating FDAAA mandates into regulatory practice.

REFERENCES

1. Moynihan R, Bero L, Ross-Degnan D, Henry D, Lee K, Watkins J, et al. Coverage by the news media of the benefits and risks of medications. *N Engl J Med*. 2000;342(22): 1645–1650.
2. Poisal JA, Murray L. Growing differences between Medicare beneficiaries with and without drug coverage. *Health Aff (Millwood)*. 2001;20(2):74–85.
3. Kreling D, Mott D, Wiederholt J, Lundy J, Levitt L. *Prescription Drug Trends: A Chartbook*. Washington, DC: The Kaiser Family Foundation; 2000.

8. US Food and Drug Administration Critical Path Initiative. http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm.

4. Sweetman SC. *Martindale: The Complete Drug Reference*. 36th ed. London: Pharmaceutical Press; 2009.
5. Pray L. *Understanding the Benefits and Risks of Pharmaceuticals: Workshop Summary Forum on Drug Discovery, Development, and Translation*. Washington, DC: National Academy of Sciences Press; 2007. http://www.iom.edu/en/Reports/2007/Understanding-the-Benefits-and-Risks-of-Pharmaceuticals-Workshop-Summary.aspx. Published August 15, 2007.
6. Opinion Research Corp. Public attitudes towards risk, regulation, and patient voices in the treatment of chronic, debilitating conditions. http://www.nclnet.org/news/2006/chronic_disease_survey_report_march2006.pdf. National Consumers League Survey Report. Published March 2006.
7. Hux JE, Naylor CD. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? *Med Decis Making*. 1995;15(2):152–157.
8. Woodcock J, Drug safety and the drug approval process: Statement before the Committee on Health, Education, Labor and Pensions of the United States Senate; March 3, 2005; Washington, DC. http://www.hhs.gov/asl/testify/t050303b.html.
9. Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything all right? interpreting zero numerators. *JAMA*. 1983;249(13):1743–1745.
10. International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). Clinical Safety E1: the extent of population exposure to assess clinical safety for drugs intended for long-term treatment of non-life threatening conditions. http://www.ich.org/cache/compo/475-272-1.html.
11. Hutchins LF, Unger JM, Crowley JJ, Coltman CA Jr, Albain KS. Underrepresentation of patients 65 years of age or older in cancer-treatment trials. *N Engl J Med.* 1999;341(27):2061–2067.
12. West J, Wright J, Tuffnell D, Janowicz D, West R. Do clinical trials improve quality of care? a comparison of clinical processes and outcomes in patients in a clinical trial and similar patients outside a trial where both groups are managed according to a strict protocol. *Qual Saf Health Care*. 2005;14(3)175–178.
13. Giezen TJ, Mantel-Teeuwisse AK, Straus SM, Schellekens H, Leufkens HG, Egberts AC. Safety-related regulatory actions for biologicals approved in the United States and the European Union. *JAMA*. 2008;300(16):1887–1896.
14. Barratt A, Wyer PC, Hatala R, et al. Tips for learners of evidence-based medicine: relative risk reduction, absolute risk reduction and number needed to treat. *CMAJ*. 2004;171(4):353–358.
15. National Health Services, Health Technology Assessment. The use of economic evaluations in NHS decision-making: a review and empirical investigation. *Health Technol Assess*. 2008;12(7):1–196. http://www.nelm.nhs.uk/en/NeLM-Area/Evidence/Guidelines/The-use-of-economic-evaluations-in-NHS-decision-making-a-review-and-empirical-investigation/?id=495722.
16. Horowitz R, Brass LM, Kernan WN, Viscoli CM. Phenylpropanolamine and risk of hemorrhagic stroke: final report of the hemorrhagic stroke project. http://www.fda.gov/ohrms/dockets/ac/00/backgrd/3647b1_tab19.doc. Published May 10, 2000.
17. US Food and Drug Administration. Phenylpropanolamine (PPA) information page. http://www.fda.gov/Drugs/DrugSafety/InformationbyDrugClass/ucm150738.htm.
18. Editorial. How a statin might destroy a drug company. *Lancet*. 2003;361(9360):793
19. US Food and Drug Administration. Withdrawal of troglitazone and cisapride. *JAMA*. 2000;283(17):2228.

20. Avorn J. *Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs*. New York: Alfred A Knopf; 2004.
21. US Food and Drug Administration. Milestones of Drug Regulation in the United States: FDA Centennial at Center for Drug Evaluation and Research; May 26, 2006; Rockville, MD. http://www.fda.gov/AboutFDA/WhatWeDo/History/ForgsHistory/CDER/CenterforDrugEvaluationandResearchBrochureandChronology/ucm114463.htm.
22. US Food and Drug Administration. Medwatch: safety alerts for human medical products. http://www.fda.gov/safety/medwatch/default.htm.
23. US Food and Drug Administration. Adverse Event Reporting System (AERS) description. http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm.
24. Bresalier RS, Sandler RS, Quan H, et al. Cardiovascular events associated with rofecoxib in a colorectal adenoma chemoprevention trial. *N Engl J Med*. 2005;352(11):1092–1102.
25. Institute of Medicine. The future of drug safety: promoting and protecting the health of the public. Washington, DC: September 22, 2006. http://www.iom.edu/reports/2006/the-future-of-drug-safety-promoting-and-protecting-the-health-of-the-public.aspx.
26. US Food and Drug Administration. Guidance for industry post-marketing studies and clinical trials: implementation Section 505(o) of the Federal Food, Drug, and Cosmetic Act. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM172001.pdf. Draft July 2009.
27. eHealth Initiative Foundation to the Food and Drug Administration. Final report: developing a governance and operations structure for the sentinel initiative. http://www.regulations.gov/search/Regs/home.html#documentDetail?R=09000064809a82f0. Published April 2009.
28. US Food and Drug Administration. FDA's sentinel initiative. http://www.fda.gov/safety/fdassentinelinitiative/default.htm.
29. US Food and Drug Administration. The sentinel initiative: a national strategy for monitoring medical product safety. http://www.fda.gov/safety/fdassentinelinitiative/ucm089474.htm. Published May 2008.
30. US Food and Drug Administration. About the sentinel discussion room. http://blog.govdelivery.com/sentinelinitiative/about-the-sentinel-discussion-room.html.
31. US Food and Drug Administration. Structured product labeling resources. http://www.fda.gov/forindustry/datastandards/structuredproductlabeling/default.htm.
32. US Food and Drug Administration. Opioid drugs and Risk Evaluation and Mitigation Strategies (REMS): FDA to meet with drug companies about REMS for certain opioid drugs. http://www.fda.gov/drugs/drugsafety/informationbydrugclass/ucm163647.htm.
33. Alvir JM, Lieberman JA, Safferman A, Schwimmer JL, Schaaf JA. Clozapine-induced agranulocytosis: incidence and risk factors in the United States. *N Engl J Med*. 1993;329(3)162–167.
34. Thompson JP, Noyes K, Dorsey ER, Schwid SR, Holloway RG. Quantitative risk-benefit analysis of Natalizumab. *Neurology*. 2008;71(5):357–364.
35. US Food and Drug Administration. Tysabri® risk minimization action plan: summary of TOUCH™. http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/UCM107197.pdf.
36. US Food and Drug Administration. Identification of drug and biological products deemed to have risk evaluation and mitigation strategies for purposes of the Food and Drug Administration Amendments Act of 2007. *Federal Register*. March 27, 2008; 73(60). http://www.fda.gov/OHRMS/DOCKETS/98fr/E8-6201.htm. Docket No.FDA-2008-N-0174.

37. US Food and Drug Administration. Approved Risk Evaluation and Mitigation Strategies (REMS). http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm111350.htm.
38. Hughes B. Anticipating REMS. *Nat Rev Drug Discov*. 2008;7(12):963.
39. US Food and Drug Administration. Important warnings and instructions for heparin sodium injection (Baxter). http://www.fda.gov/drugs/drugsafety/publichealthadvisories/ucm051133.html.
40. Blossom DB, Kallen AJ, Patel PR, Elward A, Robinson L, Gao G et al. Outbreak of adverse reactions associated with contaminated heparin. *N Engl J Med*. 2008; 359(25):2674–2684.
41. Woodcock J. The heparin disaster: Chinese counterfeits and American failures. Testimony: Subcommittee on Oversight and Investigations, Committee on Energy and Commerce, House of Representatives, 110th Congress; April 29, 2009; Washington, DC. http://energycommerce.house.gov/index.php?option=com_content&view=article&id=627&catid=31&Itemid=58.
42. US Food and Drug Administration. Information on heparin. http://www.fda.gov/drugs/drugsafety/postmarketdrugsafetyinformationforpatientsandproviders/ucm112597.
43. Laurencin CT, Nair L. The FDA and safety: beyond the heparin crisis. *Nature Biotechn*. 2008;26(6)621–623.
44. US Food and Drug Administration. Guidance for industry: standards for securing the drug supply chain: standardized numerical identification for prescription drug packages. http://www.fda.gov/RegulatoryInformation/Guidances/ucm125505.htm. Draft 2009.
45. Marsh S, McLeod HL. Pharmacogenomics: from bedside to clinical practice. *Hum Mol Genet*. 2006;(Spec No 1):R89–R93.
46. Wysowski DK, Nourjah P, Swartz L. Bleeding complications with warfarin use: a prevalent adverse effect resulting in regulatory action. *Arch Intern Med*. 2007;167(13):1414–1419.
47. Rosove MH, Grody WW. Should we be applying warfarin pharmacogenetics to clinical practice? no, not now. *Ann Intern Med*. 2009;151(4):270–273
48. US Food and Drug Administration. Table of valid genomic biomarkers in the context of approved drug labels. http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm.

This page intentionally left blank

SECTION 3

PUBLIC HEALTH CONCERNS OF SPECIAL NEEDS POPULATIONS

This page intentionally left blank

CHAPTER 11

Public Health Considerations of People with Disabilities

Arun Karpur, MBBS, MPH, Melissa J. Bjelland, PhD, and Susanne M. Bruyère, PhD, CRC

INTRODUCTION

The Institute of Medicine (IOM) [1] refers to disability as "the nation's largest public health problem," one that affects not only the health of people with disabilities, but also their immediate families and the population at large. Between 40 and 50 million Americans report a disability,[2] with various studies projecting an increasing trend in prevalence of disabilities in the United States and globally.[3–6] People with disabilities account for the single largest demographic group with unique needs for maintaining health and effective daily functioning that deserve due consideration when examining challenges to the public health system.

Traditionally, the function of a public health agency is described across three areas: (1) *assessment* (that is, systematically collecting and compiling health-related information); (2) *policy development* (that is, developing policies that are designed to improve the health and well-being of people); and (3) *assurance* (that is, assuring people that services necessary to maintain their health and well-being are provided through regulations, encouraging individual action, or through direct service delivery).[7] With respect to people with disabilities, public health has not performed adequately across these three central roles because federal databases have been inconsistent in the use of disability identifiers and in recording public health issues of concern to persons with disabilities; policy development has tended to marginalize the needs and rights to social participation of persons with disabilities; and assurances have been minimal.[2]

The evolving definition of disability poses a key hurdle in understanding what is needed for people with disabilities to maintain good health and functioning. The medical model of disability adopted within public health defines disability as a condition resulting from an underlying pathology. This model proposes only individual-level etiological factors (for example, injuries, accidents, and congenital abnormalities) and equates disability with a state of being "unhealthy." Consequently, much of the focus in public health has remained in the area of "disability

prevention," with a lack of attention to the health needs of people living with disabilities.

The social model recognizes disability as a dynamic state resulting from the interplay between individual functioning and societal and environmental reactions to a person with disability. This model emphasizes the idea that people with disabilities can lead a healthy and productive life with necessary supports. The U.S. surgeon general's call to improve the health and well-being of people with disabilities and the inclusion of several health and wellness indicators in national surveys are indications that the broader social model of disability is gradually being adopted within the field of public health.[5]

Health care policies grounded in the medical model, however, continue to pose barriers to accessing health care, prevention, and rehabilitation-related services for people with disabilities. Although advances in assistive technologies have, in general, enhanced access for people with disabilities to most public facilities and transportation, they have not immensely increased their access to health care facilities (for example, many health care facilities and clinics remain physically inaccessible and many lack adaptive medical equipment to adequately serve the health care needs of people with disabilities). Contributing to this could also be the wariness of people with disabilities to be serviced by a system so medically model driven.

Longer life spans and the aging of the baby boomers will combine to double the population of Americans ages 65 and older during the next 25 years. By 2030, the projected 72 million older adults will account for nearly 20 percent of the U.S. population.[8] Older age is often accompanied by an increased risk of chronic diseases and disorders, with large proportions of older Americans reporting a variety of chronic health conditions such as hypertension and arthritis.[9] The nation's health care spending is projected to grow by 25 percent as a result of these demographic changes.[3] The public health system needs to be better equipped to effectively deal with these challenges.

This chapter highlights the role of public health in improving the health, well-being, and overall quality of life for people with disabilities through the consideration of epidemiological trends in disability prevalence, including issues related to health disparities, the legal and regulatory environment affecting access to preventive and curative health services, methods of measuring and tracking the population of people with disabilities, and specific priorities in public health. Future directions describing ongoing efforts (in the United States and globally) in public health policy, program development, and research to improve health and well-being of people with disabilities are addressed as well.

MEASURING THE DISABILITY POPULATION

The multifaceted nature of disability makes it a challenge to adequately define, which limits the ability of public health agencies to systematically respond to the health care needs of people with disabilities. Over the years, the perception of disability has evolved from a treatable condition, initially, to a dynamic process manifested through the interactions of pathologies, impairments, activity limitations, and the environment, including social expectations.[1, 10, 11, 12, 13] Despite refinements to its meaning, no concise formal definition comprehensively encompasses all individuals

affected by disability. Currently, the federal government outlines 67 disparate legal definitions across its many programs,[14] while population surveys generally tend to contain broad questionnaire items designed to identify the population with disabilities. The range of classifications results in many unique, noncomparable indicators of the aspects of disability from these myriad sources.[15]

Over the past few years, there has been a movement toward a universal language and a common framework to describe health and health-related states, including disability.[13] Several U.S. national surveys recently have implemented or proposed new survey designs to address the changing needs of disability advocates, including the incorporation of a standard set of six disability indicators.[1] While revised questionnaires may hinder longitudinal characterizations of the community of individuals with disabilities in the years immediately surrounding survey redesign, the standardization of disability questions is anticipated to enhance our understanding of the variation in measurement and outcomes across surveys for this population going forward.[16]

In addition to addressing inconsistencies in survey measures, another priority is establishing a framework to develop a global common nomenclature.[17, 18] The International Classification of Functioning, Disability, and Health (ICF) provides a consistent language to describe the components of health and health-related conditions for people with disabilities across disciplines and between countries.[19] The ICF not only includes functional assessment, but also considers environmental and contextual factors that affect a person's functioning.[13][2]

MEASURING HEALTH STATUS

Similar to disability, health is complex to describe because it embodies physical, mental, and social well-being and not merely the absence of disease.[20] Most measures of health status rely on a medical model measuring functional limitations [21] and disability is equated with ill health.[22] A widely used concept in measuring health is health-related quality of life (HRQOL). HRQOL is defined as the extent to which health affects an individual's (1) ability to function and (2) perceived physical, mental, and social well-being.[23] The functional component of HRQOL includes the Activities of Daily Living (ADL) and the Instrumental Activities of Daily Living (IADL).[3] The HRQOL measures are used to understand health disparities among disadvantaged populations, and to assess the impact of public health interventions.[24] They also are used to describe the "burden of disease," or the extent of morbidity among people with chronic illness.[25]

1. The surveys include the American Community Survey (ACS), Current Population Survey (CPS), National Health Interview Survey (NHIS), and Survey of Income and Program Participation (SIPP). The six measures cover hearing, visual, mental, physical, self-care, and go-outside-home disabilities.
2. A review of the literature regarding the ICF can be found in Bruyère, Nan Looy, and Peterson.[137]
3. ADLs are personal care activities such as bathing, dressing, eating, walking across the room, transferring, and going to toilet. IADLs are activities essential to maintain independent living such as shopping, managing money, and doing household work (for example, cooking, laundry, and cleaning).

For people with disabilities, however, the emphasis on ADLs and IADLs may inappropriately indicate diminished overall health status.[26–29]

Another measure of the health status of people with disabilities is known as Disability-Adjusted Life Years (DALYs). Developed within the Global Burden of Disease Study supported by the World Bank, the notion of DALYs was introduced in an attempt to accommodate the role played by morbidity to calculate the life expectancy for people with chronic illnesses.[30] As the DALY treats each year of life lived with a disability as a fraction of a year lived without a disability, it may devalue the life of people with disabilities.[31, 32] If DALYs were used to conduct cost-benefit analyses of public health interventions, the artificial weights implicit to its computation could undervalue resources spent on people with disabilities, thus driving public health resources away from these people and resulting in an increase in health disparities.[33]

NATIONAL DATABASES

The United States collects administrative data to support its federally funded programs and conducts more than 40 surveys that provide researchers, policy makers, and advocates with information regarding people with disabilities.[34] Key surveys containing relevant disability and health measures are outlined in table 11.1; many of these contain differing definitions. Despite the vast quantity of data, the quality is limited with respect to the frequency of collection, the size and coverage of samples, the level of detail regarding disability types, and the duration of follow-up interviews (cross-sections are favored over longitudinal schemes). For these reasons, currently existing survey and administrative data instruments are admittedly insufficient to effectively analyze minute dimensions of the various disability subpopulations.[35, 36][4]

Lacking detailed information, it is difficult to ascertain the heterogeneity in the disabled population (for example, those that might incorporate elements of workplace environment, transition into adulthood, local geography, modes of transportation, accommodations, and onset of medical conditions).[34, 37, 38] Such knowledge is necessary to adequately evaluate the impact of programs aimed to improve the health status of people with disabilities and inform key stakeholders regarding proposed changes to program initiatives. Bridging this information gap is no straightforward task, because survey forms have finite space and must include questions to address priorities of competing agencies.

Administrative data, the alternative to national survey data sources, are collected to allow the monitoring and accountability of government-sponsored programs. By construct, these data are cost-effective and contain superior data detail for elements required to track experiences of program participants and applicants. A noted disadvantage of using these data is that those who are program ineligible or who do not apply to receive services remain unobserved. Also, driven by the medical model, most data sets continue to track "deficit reduction" as opposed to "status

4. From these sources, it is challenging to obtain statistics for specific low-prevalence disabilities (for example, autism) and for broader disabilities among fine demographic cuts of data (for example, among specific ethnic subgroups cross-tabulated by gender and age categories).

Table 11.1
Examples of Surveys Containing Disability and Health Indicators

Behavioral Risk Factors Surveillance System	National Health and Nutrition Examination Survey	National Survey of Children with Special Health Care Needs
General Social Survey	National Health Interview Survey	National Survey of Children's Health
Health and Retirement Survey	National Household Travel Survey	National Survey of Drug Use and Health/ National Household Survey of Drug Abuse
Longitudinal Studies of Aging	National Long Term Care Survey	National Survey of Family Growth
Longitudinal Study of the Vocational Rehabilitation Services Program	National Longitudinal Survey of Youth	National Survey of Homeless Assistance Providers and Clients
Medical Expenditure Panel Survey	National Longitudinal Study of Adolescent Health	National Survey of Supplemental Security Income Children and Families
Medicare Current Beneficiary Survey	National Longitudinal Transition Survey	National Survey of Veterans
National Beneficiary Survey	National Long-Term Care Survey	Nursing Home Minimum Data Set
National Comorbidity Survey	National Mortality Followback Survey	Panel Study of Income Dynamics
National Crime Victimization Survey	National Nursing Home Survey	Survey of Income and Program Participation

Source: Adapted from Livermore and She.[34]

improvement" for people with disabilities. Programs targeted toward people with disabilities often collect abundant information about the type, duration, and severity of each applicant's condition. However, these records do not contain supplemental details (for example, functional codes, enhanced demographic details, or economic indicators), which in conjunction with existing administrative data would increase the value of these files. Direct survey links to merge in additional variables are less common due to privacy concerns and government restrictions on the use of confidential information.

Despite noted limitations,[5] survey and administrative data collection efforts do allow the comparison of trends over time and often permit cross-classifications by

5. Extensive discussions of disability data limitations can be found in Livermore and She.[34] Stapleton, Livermore, and Kennell,[138] Wittenburg and Stapleton,[139] and Institute of Medicine and National Research Council.[140] Livermore and She [34] also comprehensively outline U.S. national surveys.

disability status, gender, race, ethnicity, educational attainment, and employment status at the national or state levels. Many surveys also contain modules that probe into the particulars of health status, access to health care services, unmet health care needs, and behavioral risks of the population. While these special topical sections may not consistently be part of the core set of questionnaire items, their availability nonetheless provides the opportunity to understand several aspects of health disparities.

CONSIDERATIONS IN SURVEYING PEOPLE WITH DISABILITIES

People with disabilities require special considerations to ensure that they are adequately represented in surveys and are able to participate in the interview process despite possible communication, stamina, and cognitive barriers.[39] A key first step in surveying this population is the construction of a sampling frame that will include people with disabilities. Unfortunately, in many surveys, the group quarters institutional population is excluded, which severely restricts our knowledge of the substantial number of people with disabilities who reside in hospitals, nursing homes, correctional facilities, and juvenile institutions. Group quarters of noninstitutional facilities (such as college dormitories, group homes, military quarters, and shelters) are important to include in evaluating the prevalence of disability, as well, to ensure a more accurate characterization of the state of the nation. Incomplete sampling frames that target households also miss surveying the homeless, many of whom are people with disabilities.

Survey methodologies can make it difficult to include people with certain disabilities as respondents without specialized interview preparations.[29] Major surveys with the most substantial samples tend to use mail-back questionnaires, computer-assisted telephone interview (CATI), or computer-assisted personal interview (CAPI) techniques. These may be particularly problematic when respondents may have cognitive disabilities and require additional assistance, when printed materials are submitted to people with visual impairments,[40, 41] or when the telephone is used to reach those with impaired hearing.[42, 43] Additionally, deaf respondents pose a distinct challenge due to unique cultural and linguistic issues relating to the use of American Sign Language (ASL), which differs from English in its syntax.[44, 45] Interviewers need to be trained in how best to obtain complete answers to lengthy survey questionnaires, in adopting neutral feedback to elicit continued responses without influencing the answers, and in allowing for additional time or intermittent breaks when the respondent indicates fatigue.[46]

Formulating the questionnaire, deciding upon which items to include and in what order, also requires care. Subtleties such as the phrasing and interpretation of questions,[6] as well as survey layout,[47–49] can affect the accurate measurement of the disabled population, resulting in widely varying statistics on the prevalence of disability that are not easily reconciled. The subjective nature of disability

6. Consider the item that aims to identify respondents experiencing a work limitation, defined as a "health problem or a disability which prevents work or limits the kind or amount of work" they can do. Ambiguity can result in people with both temporary and long-lasting conditions answering in the affirmative.

questions,[7] as well as the influence of cultural stigmas, may result in inaccurate answers, and proxy responses (in which someone supplies answers on behalf of the respondent) can further affect the measurement of disability. These factors contribute to the range of disability prevalence rates that are apparent when comparing statistics across the major national surveys.[50–54]

EPIDEMIOLOGY OF DISABILITY

Approximately 50.4 million noninstitutionalized Americans have a disability.[2] People 65 years and older have four times the prevalence of disability compared with younger people. Recent reports on disability statistics based on the American Community Survey data indicate that the rates of disability range from 6 percent (2.7 million) for children ages 5 to 15 years to 13 percent (22.3 million) for working-age adults (ages 21 to 64 years). The prevalence rate rises to 30 percent (5.6 million) and 53 percent (9 million) for older adults, ages 65 to 74 years and 75 years and over, respectively.[2] These data do not include people living in group homes, nursing homes, and other types of long-term care facilities, many of whom are disabled.[8]

Primary conditions associated with disability vary by age-group. Among children ages five and older, cognitive, neurodevelopmental, and speech difficulties dominate the list of disabilities; asthma is also a primary health condition associated with activity limitations.[55] Mental illness is the most common primary health condition among adults ages 18 to 44. For individuals 45 to 64, arthritis and musculoskeletal disorders top the list, followed by heart or circulatory conditions, mental illness, and diabetes.[56] For people over 65, dementia is the most prevalent condition associated with disability.[4]

Trends in the prevalence and types of disability have shifted dramatically over the past half-century in a direction that raises public health concerns. Lakdawalla et al. [57] demonstrated an increasing prevalence of disability among working-age adults, specifically among those ages 30 to 39. The rates of disability for this group increased by 50 percent between 1984 and 1996, whereas the rates of disability decreased among people between 60 and 69.[57] Furthermore, the trends beyond 1997 continue to show a moderate increase in disability among people younger than 50, a result that stands even when controlling for changes in the underlying demographic characteristics of this subpopulation over the years. Interestingly, the disability prevalence for people 85 and older has remained constant or declined over the last decade.[4]

---

7. Consider the item that aims to identify respondents experiencing a work limitation, defined as a "health problem or a disability which prevents work or limits the kind or amount of work" they can do. Ambiguity can result in people with serious disabilities indicating they have no such work limitation because (1) their current employer accommodates their disability, (2) they have self-selected into a job that avoids conflict with their disability, or (3) they would not work even if they did not have a disability. Alternatively, people without long-term disabilities may answer in the affirmative because they are experiencing a brief period of ill or disabling health that will resolve itself within a few weeks (that is, they have recently undergone a medical operation from which full recovery is anticipated).
8. The Medicare Current Beneficiary Survey (MCBS) reports that more than 2 million people with disabilities reside in long-term care facilities (Center for Medicare and Medicaid Services 2005). For further information see http://www.cms.gov/MCBS.

Several explanations exist regarding the cause of these shifts: the epidemic of obesity among children and younger adults, advances in medical technology, and generous disability benefits that encourage less-skilled and unemployed people to report a disability.

Lakdawalla et al. [57] documented that obesity accounted for approximately a 50 percent surge in the prevalence of disability among those between 18 and 29 years old; a 25 percent rise among those 30 to 39; and a 10 percent increase for those between the ages of 40 and 49. Strum et al. [58] estimate that if the juggernaut of the obesity epidemic is not stemmed, the prevalence of disability among those 50 to 69 will rise on average by 1 percent each year.

Meanwhile, advances in medical technology have improved the likelihood of survival for low-birth weight and premature high-risk infants, a risk factor for neurodevelopmental, sensory, and other health impairments. With more children developing and being diagnosed with asthma, autism, and autism-spectrum disorders (1 in every 150 children has autism or an autism-spectrum disorder), disability among children continues to show an increasing trend.[3] Improved diagnostic screening also may be contributing to what we now observe as a rising trend. The introduction of the Early and Periodic Screening, Diagnostic, and Treatment (EPSDT) benefit for Medicaid-eligible children, a program initiated in the 1960s by the Social Security Administration, has enabled earlier identification of disabling conditions among children.[59]

Generous disability benefits also are responsible for observed shifts in the trends of disability prevalence and types. An approximately 60 percent increase in enrollment for disability benefits was noted after 1984 for adults, especially among low-skilled workers, as a result of the expansion of disability welfare programs.[60] Subsequent to increasing labor market trends to hire people with advanced skills, federal disability welfare program participation has risen. Concurrent relaxation in the eligibility criteria and larger financial incentives increased the overall number of Americans on disability rolls.[61, 62]

HEALTH DISPARITIES FOR PEOPLE WITH DISABILITIES

One of the primary concerns of public health is to address the existing health disparities between people with disabilities and the general population. However, the definition of health disparities is still evolving; the Health Resources and Services Administration (HRSA) defines a health disparity as the difference in rates of diseases, health outcomes, or access to health care across population subgroups.[63] In the context of this chapter, we define health disparities as differences in health status and access to health care and preventive health services between people with and without disabilities. Kilbourne et al. [64] proposed three sequential steps to advance health disparities research: (1) detection, (2) identification of causal factors and mechanisms, and (3) development of interventions to eliminate the identified disparities.

Health disparities related to disability have been shown to exist. People with disabilities often experience poorer health, have a "thinner margin of health," and are more susceptible to secondary conditions than the general population.[27, 65] They report higher health care utilization and higher health care expenditures

compared with the general population. More than two-thirds of hospitalizations and nearly 40 percent of all prescriptions and refills are accounted for by people with disabilities.[66] They are more likely to utilize emergency care services and constitute a higher proportion of in-patient admissions compared with people without disabilities.[67] Despite this high rate of health care utilization, almost one-third of people with disabilities report unmet health care needs.[68]

The mechanisms or factors leading to health disparities and effective strategies to eliminate them are not well understood. Ethnic minorities and poor people are overrepresented among people with disabilities. A greater proportion of people with disabilities are Native American, with the lowest proportion being Asian. The number of people with disabilities is highest among low-income groups.[69] Furthermore, She and Livermore [70] demonstrated that annual poverty rates among people with disabilities are two to five times higher compared with people without disabilities, and the prevalence of long-term poverty increases with the duration of disabling conditions.[70] These socioeconomic factors are correlated with poor health conditions, which further complicate issues related to health disparities for people with disabilities. Later sections of this chapter have more extensive discussions on health disparities for specific demographic subgroups of people with disabilities (for example, children, young people, and women).

LEGAL AND REGULATORY ENVIRONMENT

Over the past several decades the federal government has enacted laws and policies aimed at reducing or eliminating disparities for people with disabilities across the areas of health, employment, education, and civil society in general. Below, we describe the protections provided by the Americans with Disabilities Act of 1990 (ADA).[9]

Americans with Disabilities Act of 1990

The employment provisions (Title I) of the ADA prohibit discrimination in employment on the basis of disability by private employers with at least 15 employees. These provisions were modeled after Section 503 of the Rehabilitation Act of 1973,[10] prohibiting job-related discrimination and requiring that employers provide "reasonable accommodations." The intent of Title I of the ADA is to provide equal employment rights to people with disabilities and to improve their opportunities in the labor market.[11]

---

9. The details of all disability-related laws and regulations are out of the scope of this chapter and readers are urged to secure appropriate resources for in-depth information. See, for example, Bruyère and Nan Looy.[141]
10. Section 503 of the Rehabilitation Act of 1973, as amended, requires covered employers to take affirmative steps to hire, retain, and promote qualified individuals with disabilities. The regulations implementing Section 503 make clear that this obligation to take affirmative steps includes the duty to refrain from discrimination in employment against qualified individuals with disabilities. For further information see http://www.dol.gov/compliance/laws/comp-rehab.htm.
11. The U.S. Equal Employment Opportunity Commission (EEOC) is the federal enforcement agency for the ADA employment provisions. See the EEOC's Web site at http://www.eeoc.gov/policy/ada.html for further information.

Under the ADA, individuals with disabilities are those who have a physical or mental impairment that substantially limits one or more major life activity (for example, walking, speaking, breathing, performing manual tasks, seeing, hearing, learning, caring for oneself, and working), who have a record of such impairment, or who are regarded as having such impairment.[12] In the late 1990s, a series of Supreme Court decisions significantly narrowed the definition of persons covered under the ADA. In response to a perceived threat of exclusion of eligible people with disabilities to the original intent of the legislation due to restrictive interpretation by the courts, work began to craft language to restore the original protections. The ADA Amendments Act of 2008 corrects the previous narrowing of the definition of disability that was occurring through court interpretations, restoring the originally intended broad interpretation.[13] Specifically, Title II covers all health care facilities and services provided by state and local governments, and Title III extends the coverage to private health care providers and facilities. Though most requirements regarding accessibility are applicable to health care facilities, additional requirements could apply to medical care facilities such as hospitals. The U.S. Architectural and Transportation Compliance Board (the Access Board) establishes guidelines for construction and alteration of health care facilities under Titles II and III of the ADA. The Centers for Medicare and Medicaid Services (CMS) also promulgates the ADA accessibility requirements by making them mandatory for participating health care facilities.

The Title I employment provisions of the ADA include several distinctly health-related aspects, including prohibitions against preemployment health-related examinations, health-related inquiries, and disparate treatment in health care coverage. An employer cannot require a medical examination, ask the employee whether the employee is an individual with a disability, or inquire as to the nature or severity of a disability unless such examination or inquiry is shown to be consistent with business necessity. The employer may make inquiries into the ability of an employee to perform job-related functions.[14] The employer also may conduct voluntary medical examinations, including voluntary medical histories, when they are part of an employee health program available to employees at that work site.[15] Insurers are prohibited from writing health care plans that disparately cover people with disabilities.[16]

One of the criticisms of the ADA is that its implementation has been more reactive than proactive, since enforcement of its guidelines is dependent upon the regulating agency's response to complaints filed by individuals.[71] Many people with disabilities may not be aware of all the ADA-related regulations and may be financially unable to pursue lawsuits. The remedies for violations of the ADA in health care may be limited to only modifications made by the offending party, and existing

12. Definitions under the ADA can be found on the EEOC Web site; Part 1630–Regulations to Implement the Equal Employment Provisions of the Americans with Disabilities Act; see Section 1630.2 Definitions.
13. Further information about the ADA Amendments Act of 2008 can be found at http://www.eeoc.gov/ada/amendments_notice.html.
14. Americans with Disabilities Act, Sec. 12113 (d)(4)(A).
15. Americans with Disabilities Act, Sec. 12113 (d)(4)(B).
16. Americans with Disabilities Act, Sec. 12201 (c)(1).

regulations discourage monetary compensation to plaintiffs.[72, 73] Furthermore, the ADA does not specifically cover access to medical equipment, which creates problems in ensuring access to high-quality health care.[74] In a national survey, 69 percent of people with disabilities indicated difficulty using medical examination tables as they were too high, and many had difficulty accessing scales. They also complained of safety issues, such as "exam tables are too narrow" or "too slippery." In addition to these, the participants indicated patient transfer issues, and problems in following directions as the visual displays and markings on these instruments were not readable.[75] The ADA does not provide any such detailed specification to improve accessibility for people with disabilities.

The ADA represents an effort to protect the rights of people with disabilities to get employment, to stay employed, and to thrive in their work environments, while being afforded equal access to all the benefits and privileges of employment, including equal access to health care. Without such protections, experience has taught us that people with disabilities are marginalized from opportunities to equitably access and succeed in the workplace. Knowledge of such laws and their protections is imperative for public health professionals, so that they can also contribute to promotion of these rights on behalf of people with disabilities.[76]

PRIORITY PUBLIC HEALTH ISSUES FOR PEOPLE WITH DISABILITIES

Efforts in public health to address and eliminate health disparities for people with disabilities must consider several priorities to better channel efforts and resources. Secondary conditions continue to pose a threat to the health and well-being of people with disabilities while contributing to increasing health care costs and disparities in employment and social participation. Simultaneously, issues related to health insurance and access to health care play a crucial role in healthy living for people with disabilities.

HEALTH PROMOTION AND PREVENTION OF SECONDARY CONDITIONS

For many years, significant attention and resources in public health have been dedicated to *primary prevention*, or the prevention and elimination of risk factors associated with disabling conditions (for example, prevention of spinal cord anomalies in newborns by routine intake of folic acid during pregnancy), and *secondary prevention*, or early diagnosis and treatment to assuage the effects of disabling conditions (for example, dietary interventions for persons with phenylketonuria). However, the issues surrounding *tertiary prevention,* the prevention of secondary conditions and overarching promotion of health and well-being for people with disabilities, slowly are being made an issue of focus within the scope of the national public health agenda.[2]

Secondary conditions for purposes of this discussion are defined as being "any additional physical or mental health condition that occurs as a result of having a primary disabling condition."[1] Though population-level data are scarce on secondary conditions, recent research indicates that they are reported by 87 percent of people with disabilities.[77] These conditions may additionally limit activities, and may

contribute to the increasing mortality and health care costs of people with disabilities.[78, 79] Many community-based studies suggest that secondary conditions decrease the likelihood of employment and productive social engagement for people with disabilities,[80–82] while contributing to the rising costs of health care for people with disabilities.[83]

The emerging role of health promotion (for example, routine exercises, healthy sleeping habits, and a balanced diet) in the prevention of these secondary conditions has enhanced its potential as one of the targets for public health intervention at the tertiary level.[78, 84] Participation in health promotion programs leads to a reduction in limitations due to secondary conditions, increased life satisfaction, and reduced costs in health care for the participants.[85] Despite its observed effectiveness in programmatic settings, significant disparities remain in access to needed interventions for people with disabilities.[86, 87]

The majority of businesses in the United States with more than 50 employees offer access to health promotion programs (for example, access to fitness centers), but many people with disabilities are not workforce participants and thus are excluded from such opportunities.[73, 88–91] Most health insurance programs (especially Medicaid and Medicare) focus only on reimbursement for acute exacerbations of chronic conditions and do not provide health and wellness as benefits. Health communication materials targeting people with disabilities are scarce, further contributing to their lack of access to health promotion.

ACCESS TO HEALTH CARE: ROLE OF HEALTH INSURANCE AND QUALITY OF HEALTH CARE SERVICES

Access to health care in the United States is primarily dependent upon the availability and coverage of health insurance plans. The impact of the lack of access to health insurance varies across age-groups for people with disabilities. Though the majority of older adults with disabilities are covered by Medicare (some are dually eligible for Medicare and Medicaid), they have to rely on supplemental insurance programs (private or public) to support their health care needs. Older adults with only Medicare coverage are much more likely to have unmet health care needs compared with their peers with supplemental insurance as well as Medicare.[92] Additionally, nearly 10 million older adults with disabilities do not have insurance coverage to pay for medications.[93] The decline in employer-sponsored retiree health insurance compounds the ability of elderly people to secure additional supplemental insurance.[94]

Health insurance coverage is mostly an issue for young adults and working-age adults with disabilities. Between 5 and 15 percent of adults with disabilities lack any type of health insurance [95] and those with no health insurance are more likely to have no regular physician, to experience trouble finding health care providers who can understand their needs, and to lack access to prescription medications for maintenance and preventative services.[67] Lack of health insurance also adversely affects the health status of immediate family members, who are often the ones caring for the person with a disability.[7]

Employer-based health insurance is an important source of affordable health care coverage for working-age people with disabilities, and a lack of employment

opportunities poses a major barrier in accessing this type of health insurance. Besides the attitudinal barriers, many employers (especially those in small businesses) fear increased costs associated with the provision of health insurance to people with disabilities,[96] which results in discriminatory recruiting practices. Private nongroup insurance, in addition to being costly, often does not provide the necessary coverage for the health conditions of people with disabilities.[97]

While the Consolidated Omnibus Budget Reconciliation Act of 1985 (COBRA) allows people who lose their job as a result of disability to keep employer-sponsored coverage for 18 months by paying the premium themselves, these plans often are not affordable without the benefit of the employer's contribution. Periods of economic downturn, in which securing and retaining jobs is increasingly difficult, heightens the issue of employment and its tie to health insurance for people with disabilities. Additionally, shorter job tenure and longer gaps between jobs occur for persons with disabilities, which may compound these issues.[98, 99]

Children with disabilities, on the other hand, are more likely to have health insurance coverage compared with their peers without disabilities. Most children with disabilities are covered through Medicaid rather than through private insurance carriers.[100] The State Child Health Insurance Plan (SCHIP) is another source of health insurance coverage for children who are above the federal poverty guidelines and do not qualify for Medicaid. Only around 5 to 7 percent of children with disabilities do not have any form of health insurance.[100] However, the frequent reenrollment procedures due to changes in circumstances (that is, change in family income) and program requirements may pose barriers in securing seamless coverage of health insurance throughout childhood. Limited transition planning in pediatric and general medical practices often forestalls continued quality care as youth move into adult health services.[101]

Access to health care for people with disabilities also is limited by several provider and facility-related factors. Despite the ADA's requirement for the accessibility of health care facilities by people with disabilities, not all physician offices and buildings have complied. Although the ADA requires that no service be denied at any health care facility, many physicians refer persons with disabilities to another physician, especially if they do not feel adequately prepared to address their health conditions.[97] The fact that educational programs training for health professionals, including physicians, rarely include considerations in caring for persons with disabilities in their curricula further contributes to these long-term barriers in accessing health care for people with disabilities.[102]

PUBLIC HEALTH CONSIDERATIONS FOR SUBPOPULATIONS

Young people, women, and elderly people with disabilities deserve unique consideration due to their distinct health-related needs and specific health policies influencing their access to health care services. A strong relationship exists between adolescent health status and adult employment, education engagement, and independent living for people with disabilities.[82, 103] Addressing health disparities for youth with disabilities is paramount to improving these outcomes for young people, complemented by their participation in evidence-based transition programs.[104] Additionally, a public health focus on reducing high-risk health-related behaviors (for example, substance

use/alcohol abuse, high-risk sexual behavior, and smoking) that have detrimental health and developmental consequences on young people with disabilities will improve their situations, as this population, on average, exhibits an increased propensity for such behaviors compared with their nondisabled peers.[105–109]

The health care needs for women with disabilities vary over their lifetime with respect to reproductive health, preventive screening, and access to health care facilities. Nearly 20 percent of American women (that is, 26 million women) have a disability, with the prevalence ranging from 7 percent among women between the ages of 16 and 64 to 30 percent among women 65 years or older.[110] Women are twice as likely to report disabilities compared with men in general.[111] Back disorders, orthopedic impairments of lower extremities, paralysis, multiple sclerosis, and mood disorders are some of the common disabling conditions affecting women.

Women with disabilities are overrepresented among poor people in America,[112] with a larger fraction living in families below the federal poverty line compared with their male counterparts.[17] A greater proportion of women with disabilities do not have access to health care because of a lack of health insurance and the high cost of health care, compared to women without disabilities.[113] Being single or unmarried and unemployed contributes to their higher likelihood of lack of access to health care, when compared to women without disabilities.[114] In addition to predisposing these women to secondary conditions, lack of access to health care contributes to strained social relationships, changes in social roles, and strained financial situations.[114]

The issue of access to preventive health care deserves unique consideration with respect to women with disabilities. Women with mobility impairments generally are less likely to receive preventive health services.[97] Less than two-thirds of disabled women receive Pap smears and less than half receive mammograms, disproportionately lower percentages than their nondisabled peers. Perhaps due to the prejudiced view that women with disabilities may not be sexually active and thus may be less likely to acquire cervical cancer, some physicians or health care providers overlook the importance of the Pap test.[115] The mammography exam requires women to stand up and many screening facilities lack adjustable mammography machines, which poses a major barrier in access to such screening, especially for women with physical disabilities.[116] Anxiety and lack of knowledge regarding such preventive services also contribute to their lower participation rates.[117]

In addition to increased attention to youth and women with disabilities, the U.S. public health agenda must increasingly focus on the needs of older people with disabilities. Americans are living longer than ever before [9] as improved medical care and prevention efforts have contributed to dramatic increases in life expectancy in the United States over the past century.[118] Research shows that the incidence and prevalence of disability increases with age.[119, 120] Platt [121] found that by the age of 50, a person's first serious medical problem will occur, with a 25 percent chance that it will be a life-long condition. Age combines with factors such as levels of functioning, marital status, gender, and socioeconomic status to affect health care access for older adults with disabilities.[113] Additionally, this population often

17. Women with severe disabilities have the highest rate of poverty among all U.S. demographic subgroups.[111]

does not engage in physical activity or follow a healthy diet, behaviors that predispose them to various disease conditions of note to public health.[122]

HEALTH AND EMPLOYMENT FOR PEOPLE WITH DISABILITIES: A PUBLIC HEALTH ISSUE

Employment for people with disabilities is a vital route to obtaining affordable health care [123] and providing access to health promotion programs necessary to maintain health and well-being. As figure 11.1 illustrates, the gap in the employment rate, or employment-to-population ratio, between working-age people with disabilities and those without has widened over the years.[124] People with disabilities are more unemployed and underemployed[18] than their nondisabled peers, which means that people with disabilities have comparatively limited access to employer-based health insurance.[52] Though most depend on public health insurance (that is, Medicaid and Medicare) for their health care needs, not all become eligible. Additionally, to obtain and remain eligible for public health insurance, individuals must demonstrate both a disability and limited economic resources. The latter criterion has been implicated as a major disincentive to employment for people with disabilities, as most fear losing access to Medicaid and not receiving adequate health insurance from a desirable job.[125]

A noted gap in public insurance is that neither Medicaid nor Medicare cover health promotion-related services (some of which are covered by employer-supported health insurance) to prevent secondary conditions for people with disabilities.[126] These services are especially important for unemployed people with

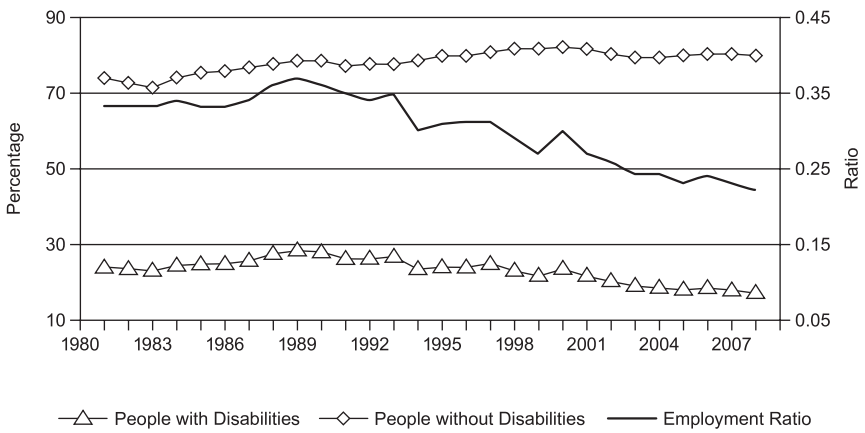

Figure 11.1 Trend in Employment Rates of the Working-Age Population, March 1981–2008. (Source: Bjelland, Burkhauser and Houtenville [124].)

18. Unemployment occurs when an individual is a labor force participant (that is, available and seeking work), but does not have a job; underemployment occurs when an individual has secured employment, but in a position that provides fewer hours than desired or that does not fully utilize one's skills or experience qualifications.

disabilities, as they are more likely than their employed peers to develop secondary conditions.[87] Not discounting the beneficial effect employment has on the lives of people with disabilities, a comprehensive public health approach to employment for people with disabilities must include workplace health and safety considerations, as is done for workers without disabilities. Research suggests that people with disabilities who are employed may have a higher incidence of work-related accidents and illnesses with resulting long-term health and economic consequences, compared with their peers without disabilities.[127] Also, exposure to chronic employment-related stress poses significant health problems for people with disabilities.[128] Thus, although workforce participation has many beneficial implications for people with disabilities and continues to be a desired outcome, the possible attendant workplace health and safety implications of employment must be a part of the public health agenda for people with disabilities.

LOOKING TO THE FUTURE: ISSUES AND CHALLENGES

Now more than ever, America faces a significant challenge with the increasing prevalence of disability. Thousands of injured veterans returning from two wars abroad, aging baby boomers, and an uncontrolled epidemic of obesity have each begun to stress the public support systems.[2] Public health plays a critical role in shaping a response to these trends, specifically around such issues as establishing needed epidemiological data and developing programmatic interventions for prevention of secondary conditions.[129]

The IOM [2] noted modest progress on its recommendations for improving the health and well-being of people with disabilities in the United States. Innovations in assistive technologies and public policies, with due credit to the ADA, have contributed to an increase in societal participation of people with disabilities. Advances in medicine and public health has contributed significantly to reducing the incidence of activity limitations, especially among the elderly population, an achievement that could be dwarfed by increasing trends of disability prevalence among younger adults and children.[57]

The IOM recommends developing a data collection system that not only measures functional limitations, disabilities, specific underlying diseases, and secondary conditions, but also collects information at the state and national levels related to barriers in social participation, risk factors for secondary conditions, and quality of life for people with disabilities. From the measurement perspective, the increasing movement among the major national surveys and the U.S. Census Bureau is to include questions that embrace the ICF definition of disability. The data, however, are nonoverlapping across time and system, and can only contribute to the establishment of a broad picture of disability in America.

Many national surveys do not include the institutionalized population of people with disabilities and those with significant disabilities. The Department of Education's National Institute on Disability and Rehabilitation Research (NIDRR) has funded disability statistics centers to generate national and regional data on disability statistics and encourage policy-relevant analyses. Efforts at these centers have resulted not only in the development of several resources guiding disability researchers in using the national survey and census data sets, but also have

engendered discussion regarding methods for collecting high-quality data on this population.[130] Efforts are ongoing to develop tools to assess function-neutral HRQOL measures for people with disabilities to eliminate the biases resulting from an intermixing of function and health.[28] Tools measuring environmental barriers in access to health care facilities have furthered the understanding of the interactive and dynamic nature of disability.[131]

Several models of health promotion programs aimed at preventing secondary conditions among people with disabilities have been successfully tested. The NIDRR has funded the National Center on Physical Activity and Disability (NCPAD) to develop community-based models for health promotion programs. These programs are especially important to address the threat of rising obesity among people with disabilities. Models to eliminate barriers to accessing such health promotion programs need to be developed for increasing effectiveness in real-life settings. Strategies of strengthening the skill and knowledge of fitness professionals in health promotion and disability, improving collaboration between rehabilitation professionals (physical therapists) and fitness personnel, and increasing financial support from public health insurance programs will sustain community-based health promotion and improve the overall well-being of people with disabilities.[87]

Participation in health promotion can be supported by funding from the State/Federal Vocational Rehabilitation Agencies or through the Medicaid Waiver programs.[85] Efforts to identify health promotion practices that are affordable, community based, and easily accessible for people with disabilities are especially needed. Population-level cost-benefit analyses of health promotion programs must be done to impress policy makers to consider including health promotion as a benefit within the public health insurance programs.

Currently, public health insurance does not universally cover people with disabilities. Existing policies call for innovations to reduce the waiting time for eligibility determination and the lengthy list of unmet service needs.[97] Recent revisions to federal laws have encouraged states to extend Medicaid coverage to working people with disabilities through the Medicaid Buy-In program as a means of encouraging workforce participation of people with disabilities. Additionally, through Medicaid partnerships with various state welfare and disability programs, the employability of people with disabilities is being further enhanced. These changes are important, as Medicaid determines the eligibility and pays for programs supporting employment-related services. Effecting an increase in interagency cooperation has not been a facile process. Several pieces of legislation over the past 10 years such as the Balanced Budget Act of 1997, the Workforce Investment Act of 1998, and the Ticket to Work and Work Incentives Improvement Act of 1999 have emphasized collaboration and participation of multiple agencies in working toward better employment outcomes for people with disabilities.[132] Even so, the partnership between Medicaid and the One Stop systems[19] (one of the key agencies

19. One Stop systems are responsible for providing employment-related services to job seekers. Mostly their clients are composed of people with disabilities, people eligible to receive public financial assistance, people involved with criminal justice system, and recent immigrants and refugees.

responsible for providing employment services to people with disabilities) is still in its infancy.[133]

Despite almost two decades of ADA enforcement, many health care facilities, hospitals, and diagnostic centers remain physically inaccessible and cannot accommodate the special needs of people with disabilities. Increased advocacy and education through public health research focusing on issues related to accessibility are necessary to advance the implementation of ADA regulation in health care facilities. The Joint Commission on Accreditation of Healthcare Organizations, CARF (formerly known as The Commission on Accreditation of Rehabilitation Facilities), and other public health accrediting agencies must develop standards for new facilities to comply with the ADA's recommended standards for accessibility. Besides improving physical accessibility, medical and public health programs should consider providing greater hands-on exposure and training to health care professionals for providing high-quality services to people with disabilities. The IOM recommends enhancing funding for supporting research to examine and eliminate health disparities for people with disabilities in the United States.

The focus of public health has shifted gradually from the prevention of mortality to the prevention of morbidity by improvements to the health and well-being of the population. The Healthy People 2010 was the first Healthy People to establish specific goals for prevention of secondary conditions, increasing employment and access to health care for people with disabilities. The workgroup on Healthy People 2020 is currently developing new indicators aimed at capturing the participation of people with disabilities in programs promoting health and well-being, and improving access to health care services. The overall focus of this new plan is to outline an action-oriented set of indicators that provides policy makers and stakeholders with necessary information to facilitate the development of new programs and policies aimed at eliminating health disparities for people with disabilities.[134]

GLOBAL CONSIDERATIONS

Key issues for people with disabilities requiring attention in the U.S. public health system also need to be addressed at the global level. It is estimated that there are approximately 650 million people with disabilities in the world, with about 80 percent living in developing countries.[135] Two of the challenges in the United States have been unique definitions of disability and the many different indexes of successful outcomes in measuring the effectiveness of services for people with disabilities. Moving our public health model for people with disabilities into a classification paradigm (that is, the ICF) that allows comparisons across national and international systems will afford us an opportunity to draw further from health, medical, and rehabilitation service advances globally in a much more timely way than previously possible.

Another global movement evident in the past few years worthy of attention is the increasing recognition of the rights of people with disabilities, as evidenced in the adoption of the United Nation's Convention on the Rights of Persons with Disabilities (UNCRPD) and its Optional Protocol[20] by the General Assembly in

20. For further information see the UN Enable Web site at http://www.un.org/disabilities/.

December 2006. The UNCRPD is one of nine core international human rights conventions. The tenets of effective public health are aligned with the principles of the UN Convention, the purpose of which is "to promote, protect and ensure the full and equal enjoyment of all human rights and fundamental freedoms by all persons with disabilities, and to promote respect for their inherent dignity."[136] Many other UNCRPD articles specifically support a public health agenda for people with disabilities globally (that is, Article 10–Right to life; Article 11–Situations of risk and life emergencies; and Article 25–Health).

Making these principles for persons with disabilities integral to the delivery of public health services globally will improve health and well-being for people with disabilities. Many of the issues discussed in this chapter that are relevant to U.S. concerns—such as health disparities for people with disabilities; unique considerations for youth, women and the aged with disabilities; equitable access to health care, health care insurance coverage, health promotion, and prevention of secondary conditions; and the need for an overarching approach that comes from a social rather than a medical model of disability—also are universal concerns that public health systems in all countries should be considering in the development of their national strategy. Collective international and multidisciplinary efforts on this front are needed to ensure that we continue to address these poignant and universal issues in the most effective way possible.

REFERENCES

1. Institute of Medicine. *Disability in America: Toward a National Agenda for Prevention*. Washington, DC: National Academies Press; 1991.
2. Institute of Medicine. *The Future of Disability in America*. Washington, DC: National Academies Press; 2007.
3. Centers for Disease Control and Prevention. *Prevalence of Autism Spectrum Disorders.* Atlanta, GA: Centers for Disease Control and Prevention; 2007. Surveillance Summaries No. 56(SS–1).
4. Freedman VA. Late-life disability trends: an overview of current evidence. In: Field MJ, Jette AM, Martin L, eds. *Workshop on Disability in America: A New Look*. Washington, DC: National Academies Press; 2006.
5. US Department of Health and Human Services. *The Surgeon General's Call to Action to Improve the Health and Wellness of Persons with Disabilities*. Washington, DC: US Department of Health and Human Services, Office of the Surgeon General; 2005.
6. Waidmann TA, Liu K. Disability trends among elderly persons and implications for the future. *J Gerontol*. 2000;55(5):S298–S307.
7. Institute of Medicine, ed. *Coverage Matters: Insurance and Health Care*. Washington, DC: National Academies Press; 2001.
8. U.S. Census Bureau. National population projections: 2008 summary tables. http://www.census.gov/population/www/projections/summarytables.html. Accessed May 6, 2009.
9. Federal Interagency Forum on Aging-Related Statistics. Older Americans 2008: key indicators of well-being. http://www.agingstats.gov/agingstatsdotnet/main_site/default.aspx. Accessed June 23, 2009.
10. Nagi S. Some conceptual issues in disability and rehabilitation. In: Sussman MB, ed. *Sociology and Rehabilitation*. Washington, DC: American Sociological Association; 1965.

11. World Health Organization. *International Classification of Impairments, Disability, and Handicaps*. Geneva, Switzerland: World Health Organization; 1980.
12. Verbrugge L, Jette A. The disablement process. *Soc Sci Med*. 1994;38(1):1–14.
13. World Health Organization. *International Classification of Functioning, Disability and Health*. Geneva, Switzerland: World Health Organization; 2001.
14. Cherry Engineering Support Services for Interagency Committee on Disability Research. Federal statutory definitions of disability. http://www.icdr.us/documents/definitions.htm. Accessed June 20, 2003.
15. U.S. General Accounting Office. *Public Housing: Impact of Designated Public Housing on Persons with Disabilities*. Washington, DC: US Government Accountability Office; 1998. No.USGAO/RCED-98-160.
16. Rietschlin J, MacKenzie A. *Variation in Disability Rates in Statistics Canada National Surveys: Building Policy on a Slippery Foundation: Proceedings of 2004 Statistics Canada Symposium on Innovative Methods for Surveying Difficult to Reach Populations*. Ottowa, Canada: Canada National Surveys; 2005. http://www.statcan.gc.ca/pub/11-522-x/11-522-x2004001-eng.htm.
17. Iezzzoni L, Greenberg M. Capturing and classifying functional status information in administrative databases. *Health Care Fin Rev*. 2003;24(3):61–76.
18. Ustun T, Chatterji S, Kostansjek N, Bickenbach J. WHO's ICF and functional status information in health records. *Health Care Fin Rev*. 2003;24(3):77–88.
19. Gray D, Hendershot G. The ICF: developments for a new era of outcomes research. *Arch Phys Med Rehabil*. 2000;8(Suppl 2):S10–S14.
20. World Health Organization. *The World Health Report 1997: Conquering Suffering, Enriching Humanity*. Geneva, Switzerland: World Health Organization; 1997.
21. Hall T. Review of existing measures. State of the Science Conference on Health, Wellness, and Disability; May 9, 2008; Portland, OR.
22. Hahn H, Hegamin AP. Assessing scientific meaning of disability. *J Disabil Policy Stud*. 2001;12:114–121.
23. Steward AL, Sherbourne CD, Hays RD. Summary and discussion of MOS measures. In: Stewart AL, Ware JE, eds. *Measuring Function and Well-Being: The Medical Outcomes Study Approach*. Durham, NC: Duke University Press; 1992:345–371.
24. Ravesloot C, Seekins T, White G. Living well with a disability health promotion intervention: improved health status for consumers and lower costs for health-care policy makers. *Rehabil Psychol*. 2005;50:239–245.
25. Hays R, Cunningham W, Sherbourne C. Health-related quality of life in patients with human immunodeficiency virus infection in the United States: results from the HIV cost and services utilization study. *Am J Med*. 2000;108:714–722.
26. Dudly-Javorski S, Sheilds RK. Assessment of physical function and secondary complication after complete spinal cord injury. *Disabil Rehabil*. 2006;28:103–110.
27. Krahn GL, Putnam M, Drum CE, Powers L. Disabilities and health: towards a national agenda for research. *J Disabil Policy Stud*. 2006;17(1):18–27.
28. Krahn GL, Fujiura GT, Drum CE, Cardinal BJ, Nosek M, Rehabilitation Research and Training Center Expert Panel on Health Measurement. The dilemma of measuring perceived health status in the context of disability. *Disabil Health J*. 2009;2:49–56.
29. Meyers AR, Andresen E. Enabling our instruments: accommodation, universal design, and access to participation in research. *Arch Phys Med Rehabil*. 2000;81(Suppl 2):S5–S9.
30. Murray C, Lopez A, eds. *The Global Burden of Disease*. Cambridge, MA: Harvard University Press; 1996.
31. Mont D. Measuring health and disability. *Lancet*. 2007;369:1658–1663.

32. Anand S, Hanson K. Disability adjusted life years: a critical review. *J Health Econ*. 1997;16:685–702.
33. Hays RD, Hahn H, Marshall G. Use of the SF-36 and other health-related quality of life measures to assess persons with disabilities. *Arch Phys Med Rehabil*. 2002; 83(Suppl 2):S4–S9.
34. Livermore G, She P. *Limitations of the National Disability Data System*. Ithaca, NY: Cornell University, Rehabilitation Research and Training Center on Employment Policy for Persons with Disabilities; 2007.
35. Cherry Engineering Support Services Inc. *Research Methods for Low-Incidence Populations. Prepared for the Interagency Committee on Disability Research*. McLean, VA: CESSI; 2005.
36. Altman B, Brown SC, Hendershot G, Larson S, Weathers B, Chevarley FM. *From Theory to Questions: Considerations of Disability Definition and Survey Purpose in Survey Implementation*: State of the Art Conference on Developing Improved Disability Data; July 12–13, 2006. Washington, DC: Interagency Subcommittee on Disability Statistics; 2006.
37. US Department of Health and Human Services. *Disability and Secondary Conditions: Healthy People 2010*. Washington, DC: US Government Printing Office; 2000.
38. Mathiowetz N, Wunderlich G. *Survey Measurement of Work Disability: Summary of a Workshop*. Washington, DC: National Academies Press; 2000.
39. Mitchell S, Ciemnecki A, CyBulski K, Markesich J. *Removing Barriers to Survey Participation for Persons with Disabilities*. Ithaca, NY: Cornell University ILR School; 2006.
40. Murray P. A multi-mode approach for surveying visually impaired populations. In: Kroll T, Keer D, Placek P, Cyril J, Hendershot G, eds. *Towards Best Practices for Surveying People with Disabilities*. Hauppauge, NY: Nova Publishers; 2007: 165–179.
41. *Assuring Inclusiveness in Data Collection among People with Visual Impairments: Proceedings in the Society for Disabilities Conference Portland University of Southern Maine*. Portland, ME: Society for Disability Studies/Edmund S. Muskie Institute of Public Affairs; 1995.
42. Wilson BF, Benaiisa S, Whitaker K, Beatty P, Hendershot G. Improving the feasibility of including deaf respondents in telephone surveys: Paper presented at the annual conference of the American Association for Public Opinion Research; St. Louis, MO. May 15–18, 1998.
43. CyBulski K, Ciemnecki A. Interviewing populations with disabilities by telephone: survey design and operations. Paper presented at the 55th Annual Conference of the American Association for Public Opinion Research and World Association for Public Opinion Research. Proceedings of the American Statistical Association. *Survey Research Methods Section.* Portland, OR. May 17–21, 2000.
44. Goldstein M, Eckhardt E, Joyner P. The inclusion of deaf individuals in survey research: Paper presented at the Federal Interagency Committee on Disability Research Conference on Best Practices for Surveying People with Disabilities; April, 27 2004; Washington, DC.
45. Margellos-Anast H, Hedding T, Perlman T, Miller L, Rodger R, Kivland L, et al. Developing a standardized comprehensive health survey for use with deaf adults. *Am Ann Deaf*. 2005;150:388–396.
46. Ciemnecki A, CyBulski K. Removing the barriers: modifying telephone survey methodology to increase self response among people with disabilities. In: Kroll

T, Keer D, Placek P, Cyril J, Hendershot G, eds. *Towards Best Practices for Surveying People with Disabilities*. Hauppauge, NY: Nova Publishers; 2007:13–31.
47. Stern SM. *Counting people with disabilities: How survey methodology influences estimates in Census 2000 and the Census 2000 Supplementary Survey*. Washington, DC: US Census Bureau; 2003.
48. Stern S, Brault M. Disability data from the American Community Survey: a brief examination of the effects of a question redesign in 2003. Washington, DC: US Census Bureau; 2005. http://www.census.gov/hhes/www/disability/ACS_disability.pdf. Census Bureau Staff Paper.
49. Houtenville AJ, Erickson WE, Bjelland MJ. *Complex Survey Questions and the Impact of Different Enumeration Procedures: Census/American Community Survey Disability Questions*. Ithaca, NY: Cornell University, Employment and Disability Institute; 2008.
50. Weathers R. *A Guide to Disability Statistics from the American Community Survey*. Ithaca, NY: Cornell University, Rehabilitation Research and Training Center on Disability Demographics and Statistics; 2005. Disability Statistics User Guide Series.
51. Erickson WE, Houtenville AJ. *A Guide to Disability Statistics from the 2000 Decennial Census.* Ithaca, NY: Cornell University, Rehabilitation Research and Training Center on Disability Demographics and Statistics; 2005. Disability Statistics User Guide Series.
52. Burkhauser RV, Houtenville AJ. *Employment among Working-Age People with Disabilities: What the Latest Data Can Tell Us*. Ithaca, NY: Cornell University, Employment and Disability Institute; 2008.
53. Harris BH, Hendershot G, Stapleton DC. *A Guide to Disability Statistics from the National Health Interview Survey*. Ithaca, NY: Cornell University, Rehabilitation Research and Training Center on Disability Demographics and Statistics; 2005. Disability Statistics User Guide Series.
54. Wittenburg DC, Nelson S. *A Guide to Disability Statistics from the Survey of Income and Program Participation*. Ithaca, NY: Cornell University, Rehabilitation Research and Training Center on Disability Demographics and Statistics; 2005. Disability Statistics User Guide Series.
55. Stein REK. Trends in disability in early life. In: Field MJ, Jette AM, Martin L, eds. *Workshop on Disability in America: A New Look*. Washington, DC: National Academies Press; 2006:143–156.
56. Bhattacharya J, Choudhry K, Lakdawalla D. Chronic disease and trends in severe disability in working-age populations. In: Field MJ, Jette AM, Martin L, eds. *Workshop on Disability in America: A New Look*. Washington, DC: The National Academies Press; 2006:113–142.
57. Lakadwalla DN, Bhattacharya A, Goldman DP. Are the young becoming more disabled? *Health Aff*. 2004;23(1):168–176.
58. Strum R, Ringel JS, Andreyeva T. Increasing obesity rates and disability trends. *Health Aff*. 2004;23(2):199–204.
59. Hertz E, Chawla A, Gavin N. Preventative services for children under Medicaid, 1989 and 1992. *Health Care Fin Rev*. 1998;19(4):25–44.
60. Autor D, Duggan M. The rise in disability recipiency and the decline in unemployment. *Q J Econ*. 2003;118(1):157–205.
61. Goodman N, Waidman T. Social Security disability insurance and the recent decline in the employment rate of people with disabilities. In: Stapleton D, Burkhauser R, eds. *The Decline in Employment of People with Disabilities: A Policy Puzzle*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research; 2003.

62. Burkhauser R, Butler J, Weathers R II. How policy variables influence the timing of social security disability insurance applications. *Soc Sec Bull*. 2002;64(1):52–83.
63. Health Resources and Services Administration. *Eliminating Health Disparities in the United States*. Washington, DC: US Department of Health and Human Services; 2000.
64. Kilbourne AM, Switzer G, Hyman K, Crowly-Matoka M, Fin MJ. Advancing health disparities research within the health care system. *Am J Public Health*. 2006; 96(12):2113–2121.
65. Pitetti K, Campbell KD. Mentally retarded individuals: a population at risk. *Med Sci Sports Exerc*. 1991;23:586–593.
66. DeJong G, Palsbo SE, Beatty PW, Jones GC, Kroll T, Neri MT. The organization and financing of health services of persons with disabilities. *Milbank Q*. 2002;80(2):261–301.
67. Hanson KW, Neuman P, Dutwin D, Kasper JD. Uncovering the health challenges facing people with disabilities: the role of health insurance. *Health Aff*. 2003; Suppl Web Exclusive:W3-552-565.
68. Harris Interactive. *2000 National Organization on Disability/Harris Survey of Americans with Disabilities*. New York, NY: Harris Interactive, Inc.; 2000.
69. Fujiura GT, Yamaki K, Czechowicz S. Disability among ethnic and racial minorities in the United States: a summary of economic status and family structure. *J Disabil Policy Stud*. 1998;9(2):111–130.
70. She P, Livermore G. Long-term poverty and disability among working-age adults. *J Disabil Policy Stud*. 2009;19(4):244–256.
71. National Council on Disability. *Promises to Keep: A Decade of Federal Enforcement of the Americans with Disabilities Act*. Washington, DC: National Council on Disability; 2000.
72. US Equal Employment Opportunity Commission. *Federal Laws Prohibiting Job Discrimination: Questions and Answers*. Washington, DC: US Equal Employment Opportunity Commission; 2002.
73. Jones NL. *Report to Congress: The Americans with Disabilities Act: Statutory Language and Recent Issues*. Washington, DC: Congressional Research Service; 2003. http://wuinfo.state.gov/usa/infousa/society/rights/ada.pdf.
74. Rosenbaum S. The Americans with Disabilities Act in a health care context. In: Field M, Jette A, eds. *The Future of Disability in America*. Washington, DC: National Academies Press; 2007:426–452.
75. Winters JM, Story MF, Barnekow K, Kailes JI, Premo B, Schweir E. Results of a national survey on accessibility of medical instrumentation for consumers. In: Winters JM, Story MF, eds. *Medical Instrumentation Accessibility and Usability Considerations*. Boca Raton, FL: CRC Press; 2007:13–28.
76. Bruyere SM. Civil rights and employment issues of disability policy. *J Disabil Policy Stud*. 2000;11(1):18–28.
77. Kinne S, Patrick DL, Doyle DL. Population prevalence of 'secondary conditions' among people with disabilities. *Am J Public Health*. 2004;94:443–445.
78. Marge M. Health Promotion for persons with disabilities: moving beyond rehabilitation. *Am J Health Prom*. 1988;2:29–35.
79. Kinne S. Distribution of secondary medical problems, impairments, and participation limitations among adults with disabilities and their relationship to health and other outcomes. *Disabil Health J*. 2008;1(1):42–50.
80. Jones GC, Bell K. *Date Brief: Health and Employment among Adults with Disabilities*. Washington, DC: NRH Center for Health and Disability Research; 2003.

81. Dooley D, Fielding J, Levi L. Health and unemployment. *Ann Rev Public Health*. 1996;17:449–465.
82. Ipsen C. Health, secondary conditions, and employment outcomes for adults with disabilities. *J Disabil Policy Stud*. 2006;17(2):77–87.
83. Morgan RO, Byrne MM, Hughes RB, Petersen NJ, Taylor HB, Robinson-Whelen S. Do secondary conditions explain the relationship between depression and health care cost in women with physical disabilities? *Arch Phys Med Rehabil*. 2008;89:1880–1886.
84. Lollar D. Public health and disability: emerging opportunities. *Public Health Rep*. 2002;117:131–137.
85. Ipsen C, Roosevelt C, Seekins T, Seninger S. A financial cost-benefit analysis of a health promotion program for individuals with mobility impairments. *J Disabil Policy Stud*. 2006;17(4):220–228.
86. Ravesloot C. *Marketing Health Promotion for People with Disabilities*. Missoula, MT: University of Montana Rural Institute; 2003. Rural Disability and Rehabilitation Research Progress Reports No.18.
87. Rimmer JH. Health promotion for people with disabilities: the emerging paradigm shift from disability prevention to prevention of secondary conditions. *Phys Ther*. 1999;79:495–502.
88. Cardinal BJ, Spaziani MD. ADA compliance and the accessibility of physical activity facilities in western Oregon. *Am J Health Promot*. 2003;17:197–201.
89. Rimmer JH, Braunscweig C, Hedman G, Heller T. *Final report: Center on Health Promotion Research for Persons with Disabilities.* Chicago, IL: National Center on Physical Activity and Disability; 2002.
90. Association for Workplace Health Promotion, Mercer W.M. Inc., US Department of Health and Human Services. *Report of 1999 National Worksite Health Promotion Survey Findings*. Northbrook, IL: Association for Workplace Health Promotion/ Mercer W.M. Inc./US Department of Health and Human Services; 2000.
91. US Department of Health and Human Services. *Healthy People 2010: Understanding and Improving Health*. 2nd ed. Washington, DC: US Government Printing Office; 2000.
92. Cohen RA, Bloom B, Simpson G, Parsons PE. Access to health care, part 3: older adults. *Vital Health Stat*. 1997;198:1–31.
93. Altman DE. The new Medicare prescription-drug legislation. *N Engl J Med*. 2004;350:9–10.
94. Frostin P. *The Impact of the Erosion of Retiree Health Benefits on Workers and Retirees.* Washington, DC: Employee Benefit Research Institute; 2005.
95. Olin G, Dougherty DD. *Characteristics and Medical Expenses of Adults 18 to 64 Years Old with Functional Limitations, Combined Years 1997–2002*. Rockville, MD: Agency for Healthcare Research and Quality; 2006. Working Papers No.06002.
96. Batavia MJ. Ten years later: the ADA and the future of disability policy. In: Francis LP, Silvers A, eds. *Americans with Disabilities: Exploring Implications of the Law for Individuals and Institutions* New York: Taylor and Francis; 2000:283–292.
97. Iezzoni LI, O'Day B. *More than Ramps: A Guide to Improving Health Care Quality and Access for People with Disability*. New York: Oxford University Press; 2006.
98. Hayward BJ, Tashjian MD. *Longitudinal Study of the Vocational Rehabilitation Service Program Second Interim Report: Characteristics and Perspectives of Vocational Rehabilitation Consumers*. Research Triangle Park, NC: Research Triangle Institute; 1996. http://www.rti.org/publications/pubabs.cfm?PUB_ID=246. Online Paper No.2000. Accessed March 15, 2000.

99. Virtanen M, Kivimaki M, Joensuu M, Virtanen P, Elovianio M, Vahetra J. Temporary employment and health: a review. *Int J Epidemiol*. 2005;34(63):610–622.
100. Chevarley FM. *Utilization and Expenditures for Children with Special Health Care Needs*. Rockville, MD: Agency for Healthcare Research and Quality; 2006.
101. Lotstein DS, McPherson M, Strickland B, Newacheck PW. Transition planning for youth with special care needs: results from the national survey of children with special health care needs. *Pediatrics*. 2005;115(6):1562–1568.
102. Tanenhaus RH, Meyers AR, Harbison LA. Disability and the curriculum in US graduate schools of public health. *Am J Public Health*. 2000;90(8):1315–1316.
103. Blohmquist KB. Health, education, work, and independence of young adults with disabilities. *Orthop Nurs*. 2006;25(3):168–187.
104. Currie J. Healthy, wealthy, and wise: socioeconomic status, poor health in childhood, and human capital development. Cambridge, MA: National Bureau of Economic Research; 2008. http://www.nber.org/papers/w13987. NBER Working Paper No.13987.
105. Case A, Fertig A, Paxson C. The lasting impact of childhood health and circumstances. *J Health Econ*. 2005;24:365–389.
106. Hollar D. Risk behaviors for varying categories of disability in NELS:88. *J Sch Health*. 2005;75(9):350–358.
107. Hollar D, Moore D. Relationship of substance use by students with disabilities to long-term educational and social outcomes. *Subst Use Misuse*. 2004;39(6): 929–960.
108. Jones SE, Lollar DJ. Relationship between physical disabilities or long-term health problems and health risk behaviors or conditions among US high school students. *J Sch Health*. 2008;78(5):252–257.
109. Moore D, Li L. Alcohol use and drinking-related consequences among consumers of disability services. *Rehabil Couns Bull*. 1994;38(2):123–133.
110. Waldrop J, Stern S. *Disability Status 2000: Census 2000 Brief*. Washington, DC: US Census Bureau; 2003.
111. Jans L, Stoddard S. *Chartbook on Women and Disability in the United States: An InfoUse Report*. Washington, DC: US Department of Education/National Institute on Disability and Rehabilitation Research; 1999. http://www.infouse.com/disabilitydata/womendisability/.
112. Holden R, Holden K, Wolfe B, Smith P, Wilson K. *The Changing Economic Status of Disabled Women, 1982–1991: Trends and Their Determinants*. Madison, WI: University of Wisconsin, Institute for Research on Poverty; 1997.
113. Smith DL. Disparities in health care access for women with disabilities in the United States from the 2006 National Health Interview Survey. *Disabil Health J*. 2008;1(1):79–88.
114. Piotrowski K, Snell L. Health needs of women with disabilities across the lifespan. *JOGN Nurs*. 2007;36(1):79–87.
115. Nosek M, Howland CA. Breast and cervical screening among women with physical disabilities. *Arch Phys Med Rehabil*. 1997;78(Suppl 5):S39–S44.
116. Iezzoni LI, McArthy EP, Davis RB, Sebens H. Mobility impairments and use of screening and preventive services. *Am J Public Health*. 2000;90(6):955–961.
117. Sullivan SG, Slack-Smith LM, Hussain R. Understanding the use of breast cancer screening services by women with intellectual disabilities. *Soc Prev Med*. 2004;49:398–405.
118. Centers for Disease Control and Prevention, Merck Company Foundation. *The State of Aging and Health in America: 2007 Report*. Whitehouse Stations, NJ: The Merck Company Foundation; 2009. http://www.cdc.gov/aging/saha.htm.

119. Ho P. *Disability and Employment Status among Older Workers*. Washington, DC: National Rehabilitation Hospital, Center for Health and Disability Research; 2002.
120. He W, Sengupta M, Velkoff VA, DeBarros KA. *65+ in the United States: 2005 US Department of the Census Current Population Reports*. Washington, DC: US Government Printing Office; 2005.
121. Platt RK. The aging work force: angst for organizations, baby boomers. *Workspan*. 2001;44(1):26–29.
122. McGuire LC, Strine TW, Okoro CA, Ahluwalia IB, Ford ES. Healthy lifestyle behaviors among older US adults with and without disabilities, behavioral risk factor surveillance system. *Prev Chronic Dis*. 2007;4(1):1–11.
123. Holahan J, Cook A. Changes in economic conditions and health insurance coverage, 2000–2004. *Health Aff*. 2005;W5:498–508.
124. Bjelland MJ, Burkhauser RV, Houtenville AJ. *2008 Progress Report on the Economic Well-Being of Working Age People with Disabilities*. Ithaca, NY: Cornell University, Rehabilitation Research and Training Center on Employment Policy for Persons with Disabilities; 2008. http://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1255&context=edicollect.
125. Stapleton D, Livermore G, Scrivner S, Tucker A. *Literature Review: Exploratory study of health care coverage and employment of people with disabilities*. Washington, DC: The Lewin Group; 1997.
126. Gordon C, Lapin P. A government perspective: if there is so much proof, why is Medicare not rapidly adopting health promotion and disease prevention programs? *Am J Health Promot*. 2001;15:383–387.
127. Blum RW, Kelly A, Ireland M. Health-risk behaviors and protective factors among adolescents with mobility impairments and learning and emotional disabilities. *J Adolesc Health*. 2001;28:481–490.
128. Koletsi M, Niersman A, van Busschbach JT, Catty J, Becker T, Burns T. Working with mental health problems: clients experiences of IPS, vocational rehabilitation, and employment. *Soc Psychiatry Psychiatr Epidimiol*. 2009;Epub ahead of print.
129. Lollar D, Crews J. Redefining the role of public health in disability. *Ann Rev Public Health*. 2003;24:195–208.
130. Houtenville AJ, Stapleton DC, Weathers R, Burkhauser RV. *Counting Working Age People with Disabilities: What Current Data Tell us and Options for Improvement*. Kalamazoo, MI: WE Upjohn Institute; 2009.
131. Craig Hospital Research Department. *Craig Hospital Inventory of Environmental Factors Manual*. Englewood, CO: Craig Hospital; 2001.
132. Sulewski J, Gilmore D, Foley S. *Medicaid Involvement in Employment Related Programs: Findings from the National Survey of State Systems and Employment for People with Disabilities*. Boston, MA: Institute for Community Inclusion; 2003. http://www.communityinclusion.org/article.php?article_id=122. Research to Practice No.33.
133. Sulewski J, Gilmore D, Foley S. Medicaid and employment of people with disabilities: findings from the national survey of state systems and employment for people with disabilities. *J Disabil Policy Stud*. 2006;17(3):158–165.
134. The Secretary's Advisory Committee. *National Health Promotion and Disease Prevention Objectives 2020: Recommendations for the Framework of Healthy People 2020*. Washington, DC: US Department of Health and Human Services; 2008.
135. World Health Organization. Disability, including prevention, management, and rehabilitation. http:www.who.int/nmh/a5817/en/. Published 2005. Accessed June 13, 2009.

136. United Nations. Convention on the rights of persons with disabilities. http://www.un.org/disabilities/default.asp?navid=12&pid=150. Published 2006. Accessed April 19, 2009.
137. Bruyere S, Nan Looy SA, Peterson D. The International Classification of Functioning, Disability, and Health (ICF): contemporary literature overview. *Rehabil Psychol*. 2005;50(2):113–121.
138. Stapleton DC, Livermore GA, Kennell D. *Final Report: Design Options for Modeling Policy Impacts on Working-Age Adults and Children with Disabilities*. Washington, DC: US Department of Education, Office of the Assistant Secretary for Planning and Evaluation. 2003.
139. Wittenburg DC, Stapleton DC, National Institute for Disability and Rehabilitation Research. *Review of Data Sources for School to Work Transitions by People with Disabilities*. Ithaca, NY: Cornell University, Rehabilitation Research and Training Center for Economic Research on Employment Policy for People with Disabilities; 1999.
140. Institute of Medicine, National Research Council. *Survey Measurement of Work Disability*. Washington, DC: National Academies Press; 2000.
141. Bruyere S, Nan Looy SA. Legislation and rehabilitation service delivery. In: Eisenberg M, Glueckauf R, Zaretsky H, eds. *Medical Aspects of Disability: A Handbook for the Rehabilitation Professional*. 3rd ed. New York: Springer Publishing; 2005:827–850.

This page intentionally left blank

CHAPTER 12

Health Care Expenditures of Immigrants in the United States: A Nationally Representative Analysis

Sarita A. Mohanty, MD, MPH, Steffie Woolhandler, MD, MPH, David U. Himmelstein, MD, MPH, Susmita Pati, MD, MPH, Olveen Carrasquillo, MD, MPH, and David H. Bor, MD

The United States is a nation of immigrants. In 2000, the immigrant population of the United States was 28.4 million, 10.4 percent of the total population.[1] In one of the most comprehensive analyses to date on the costs and benefits of immigrants to the U.S. economy, the National Research Council concluded that immigrants add as much as $10 billion to the economy each year and that immigrants will pay on average $80,000 per capita more in taxes than they use in government services over their lifetimes.[2] The Social Security Administration estimates that workers without valid social security numbers contribute $8.5 billion annually to Social Security and Medicare. Such workers, most of them immigrants, usually receive no eligibility credits for their contributions.[3] Taxpayers and politicians in states such as New York, California, Texas, Arizona, and Florida have expressed concern about the potential extra burden immigrants place on their states' health care systems,[4–8] particularly state welfare and Medicaid programs.[9]

Researchers from the Center for Immigration Studies have concluded that because immigrant labor has "limited value . . . in an economy that increasingly demands educated workers," providing insurance to immigrants is "at the taxpayer expense."[10] These views have resulted in legislative initiatives such as California's Proposition 181, which attempted (before it was ultimately overturned in court) to bar undocumented immigrants from receiving nonemergency health services.[11] Similarly, the 1996 Personal Responsibility and Work Opportunity Reconciliation Act made most legal immigrants who entered the United States after 1996 ineligible for Medicaid for five years after entry.[12]

Although more recent surveys suggest that public attitudes toward immigrants' contributions, particularly with regard to economic impact, are becoming more positive,[13] public fears that arose after September 2001 may reverse this trend.

In this study, we used nationally representative data to compare the health care expenditures of immigrants and U.S.-born individuals.

This chapter was originally published as Sarita A. Mohanty et al, "Health Care Expenditures of Immigrants in the United States: A Nationally Representative Analysis," American Journal of Public Health, Vol. 95: 1431–1438. Reprinted, with permission, from The American Journal of Public Health, Volume 95

METHODS

Survey Instrument

We analyzed data from the Agency for Healthcare Research and Quality's 1998 Medical Expenditure Panel Survey (MEPS). This survey is designed to provide nationally representative estimates of expenditures and health services for the U.S. civilian noninstitutionalized population.[14] To provide estimates for specific priority populations, MEPS oversamples low-income families and ethnic minorities. The MEPS data are compiled through information obtained from the Household Component, the Medical Provider Component, and the Insurance Component of MEPS. In the MEPS Household Component, respondents use a computer-assisted program to report sociodemographic characteristics, health and functional status, use of medical care services, health insurance coverage, income, and employment. The MEPS Medical Provider Component supplements and validates information on medical care events reported in the Household Component by contacting providers and facilities identified by household respondents. The Medical Provider Component includes expenditure data from hospitals, outpatient medical providers, home health agencies, and pharmacies.

We analyzed total health expenditures during 1998, including expenditures for several specific population subgroups and categories of health care. MEPS defines expenditures as the sum of payments for care provided during 1998. This figure includes payments such as out-of-pocket payments, insurers' payments, and imputed payments for free care received in public hospitals or clinics. The Agency for Healthcare Research and Quality uses weighted sequential hot-deck imputation [15] for any missing values (for a respondent with missing data, values are imputed from the nearest preceding respondent in the sequence who has similar characteristics and complete information).[16] MEPS combines facility and physician expenses when tabulating emergency department, hospital-based outpatient, and inpatient expenditures. Payments for over-the-counter drugs and for alternative medicine (for example, acupuncture, chiropractic care) are not included in MEPS. MEPS expenditure estimates exclude costs for health care administration and institutionalized care. After adjustment for these omissions, however, MEPS estimates of national health expenditures substantially agree with those of the U.S. Department of Health and Human Services' National Health Accounts.[17]

MEPS expenditure data include estimates of free care and bad debt in public hospitals or clinics. These imputed expenditure data are designed to account for payments made from government budgets that are not tied to specific patients. MEPS expenditure data do not cover uncollected liabilities, negotiated discounts, bad debt, and free care associated with private providers.[15] By some estimates, U.S. hospitals (public and private) write off as much as $2 billion a year in unpaid medical bills to treat illegal immigrants.[18] Therefore, we performed a separate confirmatory analysis of MEPS total charges (rather than expenditures) for health care, which include free care delivered at any site. Charge variables should be interpreted with caution, because they do not represent actual dollars exchanged for services or the resource costs of those services.[15]

To obtain data on the immigration status of respondents, we combined the Household Component file of the 1998 MEPS with the 1996–1997 National Health Interview Survey (NHIS), which asked respondents about their place of birth. Each year, MEPS draws a new panel from the previous year's NHIS sample. The NHIS includes self-reported data on place of birth as well as on a variety of other sociodemographic and household characteristics not included in the MEPS. As described elsewhere,[19] NHIS and MEPS data sets can be linked. In 1998, MEPS sampled 24,072 individuals and assigned positive person-level weights for 22,953 individuals. We are able to link 21,241 individuals in the MEPS sample (18,398 U.S.-born persons and 2,843 immigrants) with the NHIS sample. Individuals sampled in MEPS were not linked with the NHIS sample (or did not receive a person-level weight) if they were not a member of an NHIS household at the time of the 1996–1997 NHIS interview but had entered the household by the time of the MEPS interview (for example, newborns; those returning from military service, college, or travel; those newly married or moving into a new household).

We found that when these files were linked, 7.4 percent of the MEPS sample was omitted. This factor remains a limitation of the MEPS-NHIS merge, because no weighting adjustment was made for these missing individuals. Despite this limitation, the merging of these two national data sets is an accepted methodology.[20] An individual was defined as U.S. born if he or she was born in one of the 50 states or the District of Columbia. All others were classified as foreign born. Foreign-born persons included naturalized citizens, permanent residents, visa holders, refugees, and undocumented immigrants. Data on specific resident categories were not provided in the NHIS. For the purpose of this study, the terms "foreign born" and "immigrant" were considered to be synonymous.

Statistical Analysis

To obtain nationally representative estimates, we used person-level weights (which reflect population distributions and account for each household's probability of selection), ratio adjustment to national population estimates at the household level, and adjustment for nonresponse. Because population estimates may be unstable if cells have fewer than 100 respondents, we combined such small cells with other subgroups for our analyses.[21] To obtain estimates of variability, we used a Taylor Series estimation approach with the SUDAAN software package.[22] We performed χ^2 analyses to examine the distribution of categorical variables among immigrants and U.S.-born persons. We used t tests to compare mean per capita health expenditures among groups.

To obtain estimates of health expenditures adjusted for potential covariates, we used the Rand Health Insurance Experiment two-part regression model.[23–25] This model is used to analyze heteroscedastic and highly skewed data, such as health care expenditures (many people report no health care expenditures). The model uses an initial multivariate logistic regression to predict the probability of having any expenditure. This probability is multiplied by the predicted log-transformed expenditure of any individual with nonzero expenditures (as determined from a multivariate linear regression model of individuals with nonzero expenditures). For this two-part

model, we used SUDAAN statistical software, which allows adjustment for complex survey design.

Covariates in the two-part model included the following: age (analyzed as both a continuous and a categorical variable), gender, race or ethnicity, family income (dichotomized as either <200 percent or ≥ 200 percent of the federal poverty level [FPL]), education, insurance status, self-reported health status, residence in a metropolitan statistical area, and geographic region. In preliminary models, we found that after adjustment for other covariates, gender, education, geographic region, and metropolitan statistical area were no longer significant predictors of health care expenditures, nor did they improve the model fit. They therefore were excluded, leaving the following covariates in the final regression models to predict expenditures: age (as a continuous variable), race or ethnicity, insurance status, family income, and self-reported health status. Additionally, we explored the possibility of interactions of the covariates with immigrant status. We found a significant interaction between immigrant status and race or ethnicity, and therefore included an interaction term in the multivariate regression analyses.

As in other studies,[26, 27] we used smearing factors to retransform the final estimates [28, 29] and calculated standard errors for predicted expenditures, using bootstrapping with 2,000 iterations.[30] We conducted a stratified regression analysis of health care expenditures by insurance status and income, again controlling for the other covariates in the model. We opted to perform these stratified analyses because income and insurance status are important predictors of health service use.

We preformed a subgroup analysis of government payments (Medicare, Civilian Health and Medical Program of the Uniformed Services of the United States [CHAMPUS], Civilian Health and Medical Program of the Veterans Administration [CHAMPVA], Tricare, Medicaid, and other public hospital/physician coverage) by using a two-part multivariate regression model similar to that described in this section.

Because children's health care use differs from that of adults and is of particular policy interest,[26, 31, 32] we performed separate analyses comparing immigrant children (n = 276) with U.S.-born children (n = 5,657) younger than 18 years. For children, we also used a two-part model regression analysis similar to that described in this section, controlling for age, race or ethnicity (including a term capturing the interaction of race or ethnicity with immigrant status), poverty level, insurance status, and functional status. In our model for children, we included two variables that have been used as surrogates for a child's functional status:[20, 26] (1) whether a child resists illness well (reported by a parent) and (2) whether a child performs age-appropriate tasks (also reported by a parent).

RESULTS

In 1998, immigrant health care expenditures were \$39.5 billion (SE = \$4.0 billion), or 7.9 percent of the U.S. total. This figure included \$25 billion (SE = \$3.4 billion) in payments made by private insurers on behalf of immigrants, \$2.8 billion (SE = \$0.4 billion) paid directly by immigrants, and \$11.7 billion (SE = \$1.7 billion) paid by government sources. U.S.-born individuals (90 percent of the

population) accounted for 93 percent of private insurer expenditures and 92 percent of both government and out-of-pocket payments.

We found that immigrants differ from U.S.-born persons in demographics, unadjusted per capita health expenditures, and adjusted health expenditures. Immigrants overall were younger, although the immigrant population contained a lower proportion of children than did the U.S.-born population. In addition, compared with U.S.-born persons, immigrants had lower incomes and education attainment and lower self-reported health status, and were more likely to live in the West, the Northeast, and urban regions.

Unadjusted per capita total health care expenditures were lower for immigrants than for the U.S.-born across all age-groups (the difference for those 65 years and older was not statistically significant). For example, per capita expenditures of immigrant children younger than 12 years were 49 percent lower than those of U.S.-born children, and expenditures of immigrant children age 12 to 17 years were 76 percent lower than those of U.S.-born adolescents. The differences in expenditures between immigrants and nonimmigrants were substantially greater for men than for women. Poorer immigrants and immigrants with government insurance had lower expenditures than did the poorer U.S.-born and the U.S.-born publicly insured. Total health care expenditures for both U.S.-born persons and immigrants were highly skewed. The median total expenditure for health care was $1,563 for U.S.-born persons versus $1,163 for immigrants ($P < .0001$). In particular, data showed that immigrants have a lower probability of expenditures and a lower probability of expenditures for emergency care, office-based visits, and prescription medications than U.S.-born persons.

We also performed a multivariate analysis of health care expenditures stratified by insurance status and income. Per capita total expenditures of insured immigrants (those with any private or public insurance) were 52 percent lower than those of insured U.S.-born individuals; expenditures for uninsured immigrants were 61 percent lower than those for the U.S.-born uninsured. In a subgroup analysis limited to persons with public coverage, per capita expenditures of publicly insured immigrants were 44 percent lower than those of U.S.-born persons who were publicly insured ($2,774 [SE = $231] versus $4,963 [SE = $189]; $P < 0.0001$). Expenditures of higher income immigrants (those with incomes $\geq$200 percent of the FPL) were 53 percent lower than those of higher-income U.S.-born persons; health care expenditures of lower-income immigrants (those with incomes $<$200 percent of the FPL) were 60 percent lower than those of lower-income U.S.-born individuals. Similar patterns were seen in analyses of expenditures for emergency care, office-based visits, outpatient visits, inpatient visits, and prescription drugs stratified by insurance and income status.

Immigrant children were much more likely than U.S.-born children to be uninsured (29 percent versus 9 percent, $P < .0001$) or publicly insured (31 percent versus 20 percent, $P < .0001$). However, immigrant children's rates of public coverage were disproportionately low compared with the same children's poverty rates; 43 percent of immigrants children lived in low-income families, compared with 23 percent of U.S.-born children ($P < .0001$).

Expenditures for total health care, office-based visits, outpatient visits, inpatient visits, and prescription drugs were markedly lower for immigrant children than for U.S.-born children. Per capita emergency department expenditures, however, were

more than three times higher among immigrant children than among U.S.-born children. We performed a stratified analysis by insurance status and income of children's health care expenditures. Health care expenditures for insured immigrant children were 60 percent lower than those for insured U.S.-born children. Health care expenditures for uninsured immigrant children were 86 percent lower than those for uninsured U.S.-born children. Expenditures among higher-income immigrant children were 53 percent lower than those among higher-income U.S.-born children. Expenditures of immigrant children in lower-income brackets were 84 percent lower than those of lower-income U.S.-born children.

We also estimated health care expenditures among all U.S.-born persons and immigrants according to race or ethnicity. As shown in table 12.1, after multivariate adjustment, non-Hispanic whites had the highest per capita expenditures, whereas Hispanics and Asians had the lowest per capita expenditures. Health care expenditures were similar for U.S.-born and immigrant Asians. In contrast, adjusted health expenditures for immigrant non-Hispanic whites, non-Hispanic blacks, and Hispanics were lower than those for U.S.-born individuals from these groups.

DISCUSSION

Immigrants have less access to health care and less health care use than do U.S.-born individuals, as reflected in their lower health care expenditures. Studies have shown that insurance coverage increases access to care and thus utilization of care, as well as improving health outcomes.[33, 34, 35] In our study, we found that per capita health care expenditures for immigrants in 1998 were far lower than expenditures for the U.S. born. In addition, among adults and children enrolled in publicly financed insurance programs, immigrants had lower per capita publicly financed health care expenditures than did the U.S. born. We also found grave disparities in expenditures among racial and ethnic groups, particularly among immigrants who were non-Hispanic white, non-Hispanic black, or Hispanic.

When stratified by age, immigrants in every age-group but 65 years and older had health care expenditures that were 30 percent to 75 percent lower than those for

Table 12.1
Adjusted per Capita Health Care Expenditures among U.S.–Born Persons and Immigrants of All Ages, by Race or Ethnicity

	Per Capita Expenditures (U.S. dollars)	
Race/Ethnicity	U.S.-Born (SE)	Immigrants (SE)
Non-Hispanic White	3117 (40)	1747 (115)***
Non-Hispanic Black	2524 (80)	1030 (123)***
Hispanic	1870 (60)	962 (53)***
Asian/Pacific Islander	1460 (198)	1324 (82)

Note: Mean per capita expenditures were predicted by a two-part model with adjustments for age, poverty level, insurance status, and patient-reported health status.

Source: Data are from the 1998 Medical Expenditure Panel Survey and the 1996–1997 National Health Interview Survey.

*** $p < 0.001$ (for comparison with U.S.-born).

U.S.-born persons. Disparities among children were greatest, particularly among adolescents 12 to 17 years old. Combined with our finding of higher per capita emergency department expenditures for immigrant children, our data suggest that access to routine and ongoing care may be especially problematic for immigrant children. These findings are consistent with those of a 1999 study using NHIS data [36] that showed foreign-born children were five times more likely than U.S.-born children to lack a usual source of health care.

Ku and Matani [37] found that noncitizen children were less likely than citizen children to have made both ambulatory and emergency department visits. Like Ku and Matani, we found a significantly lower mean number of emergency department visits among immigrant children than among U.S.-born children (data not shown); however, per capita emergency department expenditures for immigrant children were significantly higher because immigrant children's costs per visit were much higher. This finding suggests that immigrant children may be sicker when they arrive at the emergency department. The higher emergency department expenditures we found for immigrant children probably reflect poor access to primary care (as evidenced by such children's low outpatient, office-based-visit health expenditures).

Some of our findings may be explained by the limits that the 1996 welfare reform legislation [38, 39] imposed on immigrants' eligibility for government health services. The Personal Responsibility and Work Opportunity Reconciliation Act [12] and the Illegal Immigration Reform and Immigrant Responsibility Act [40] substantially restricted recent immigrants' eligibility for Medicaid and other public benefits.

Before 1996, all legal permanent residents and other legal immigrants had the same access to public benefits, including Medicaid, as did U.S. citizens. Welfare reform and other policies established a five-year ban on Medicaid eligibility for non-refugee immigrants entering the United States after August 1996. The reform also stated that the income of immigrants' sponsors would be counted in determining eligibility and that sponsors could be held financially liable for public benefits used by immigrants. These policies created confusion about eligibility and appeared to lead even eligible immigrants to believe that they should avoid public programs. Even in states that have attempted to continue public insurance for immigrants, lack of awareness of eligibility for these programs remains a problem.[41]

Our findings remained robust even after adjustment for health insurance status, suggesting that immigrants compared with the U.S. born face additional unmeasured access barriers, including cultural and linguistic barriers.[42–44] As an example, one study at an inner-city clinic found that one in nine immigrant parents reported that they had not brought their children in for care because they felt that the medical staff did not understand Latino culture.[45] Additionally, among the 5 to 10 million immigrants residing in the United States who are undocumented, fear of deportation is a barrier.[46]

Our finding of lower health care expenditures among immigrants cannot be explained by free care. The MEPS captures free care (and bad debt) in public (but not private) institutions as expenditures; the MEPS captures free care at any site as a charge. Our charge-based analysis indicates that adjustment of expenditure data for free care at private institutions would not change our results. This conclusion is supported by a recent study that found no relationship between a state's uncompensated

care burden and its percentage of noncitizen immigrants.[47] The deficit of care among immigrants is probably not because of less need; immigrants in our study had slightly worse self-reported health than U.S.-born persons.

Several limitations of this study should be noted. First, because the 1998 MEPS, like the 2000 U.S. Census,[1] did not ask about immigration or citizenship status, we could not distinguish between naturalized citizens and other immigrant groups. Thus, our immigrant category included many European-born persons who resided in the United States for decades, had already become U.S. citizens, and had fully assimilated into U.S. culture and the U.S. economy and health care system. Had we been able to exclude such immigrants, we probably would have found greater disparities. Similarly, we could not specifically identify undocumented persons, whom we suspect have the lowest health care expenditures.

Our study could not capture health care expenditures outside the United States, where some immigrants may travel to obtain care or prescription drugs. For example, immigrants near the Mexican border may obtain medications from pharmacies in Mexico. These omitted out-of-country expenditures could not be viewed as a burden on the U.S. health care system. MEPS also omits expenditures for medical care received by institutionalized persons (including nursing home residents) and for nonprescription drugs. Studies have consistently found that racial or ethnic minority populations reside in nursing homes less often than do non-Hispanic whites.[48]

Our findings show that widely held assumptions that immigrants are consuming large amounts of scarce health care resources are invalid; these findings support calls to repeal legislation proposed on the basis of such assumptions. The low expenditures of publicly insured immigrants also suggest that policy efforts to terminate immigrants' coverage would result in little savings. In addition, lower health care expenditures by immigrants suggest important disparities in health care use, especially for children. Immigrant children will grow up to become a major segment of the U.S. workforce in the coming years. Ensuring access to health services needed for proper growth and development should be a national priority. Policies that may improve immigrants' access to care include providing interpreter services, ending restrictions on Medicaid and State Children's Health Insurance Program eligibility, improving employer-provided coverage for immigrant workers, and implementing universal national health insurance.[49] Our study lends support to these and other initiatives aimed at reducing and ultimately eliminating disparities in access to and use of health services.

REFERENCES

1. Schmidley AD. *Profile of the Foreign-Born Population in the United States: 2000*. Washington, DC: US Census Bureau; 2001. Current Population Reports, series P23–206.
2. Smith JP, Edmonston B. *The New Americans: Economic, Demographic, and Fiscal Effects of Immigration*. Washington, DC: National Academies Press: 1997.
3. Porter E. Illegal immigrants are bolstering social security with billions. *New York Times*. April 5, 2005;A1,C6.
4. Carrasquillo O, Carrasquillo AI, Shea S. Health insurance coverage of immigrants living in the United States: differences by citizenship status and country of origin. *Am J Public Health*. 2000;90(6):917–923.

5. House Research Organization. *Health Care for Undocumented Immigrants: Who Pays?* Austin, TX: Texas House of Representatives; October 19, 2001. Report 77–13.
6. Federation for American Immigration Reform. The costs of illegal immigration to Arizonians: a report by the Federation for American Immigration Reform. http://www.fairus.org/news/NewsPrint.cfm?ID=2442&c=55. Published June 2004. Accessed May 3, 2005.
7. American Medical Association. Hawryluk M. States might get help with immigrant care. http://www.ama-assn.org/sci-pubs/amnews/pick_02/gvsa1021.htm. Published October 21, 2002. Accessed March 6, 2003.
8. Levin M. Must Texas taxpayers subsidize their own dispossession? http://www.vdare.com/misc/levin_illegals_in_er.htm. Published July 31, 2001. Accessed March 6, 2003.
9. Allen C. Immigration and the budget. *The Austin Review*, April 29, 2003. http://www.austinreview.com. Accessed May 15, 2003.
10. Center for Immigration. *Without Coverage: Immigration's Impact of the Size and Growth of the Population Lacking Health Insurance*. Washington, DC: Center for Immigration; July 2000.
11. *Proposition 187: Illegal Aliens. Ineligibility for Public Services. Verification and Reporting. California Ballot Pamphlet: General Election*. Sacramento: Secretary of State; November 8, 1994:54–55.
12. Personal Responsibility and Work Opportunity Reconciliation Act, Pub L 104–193, 110 Stat 2105. http://wdr.doleta.gov/readroom/legislation/pdf/104-193.pdf. Published August 22, 1996. Accessed May 3, 2005.
13. The Pew Research Center for the People and the Press. Demographic shifts divide races: no consensus on the census. http://people-press.org/reports/display.php3?PageID=104. Published May 13, 2001. Accessed March 15, 2003.
14. Cohen JW, Monheit AC, Meauregard KM et al. The Medical Expenditure Panel Survey: a national health information resource. *Inquiry*. 1996;33(4):373–389.
15. Agency for Healthcare Research and Quality. MEPS HC-028: 1998 full year consolidated data file. http://www.meps.ahrq.gov/Pubdoc/HC028/H28DOC.pdf. Published December 2001. Accessed April 2003.
16. Monheit, AC, Wilson R, Arnett RH. *Informing American Health Care Policy: The Dynamics of Medical Expenditure and Insurance Surveys, 1977–1996*. San Francisco, CA: Jossey-Bass Publishers; 1999.
17. Selden TM, Levit KR, Cohen JW, et al. Reconciling medical expenditure estimates from the MEPS and the NHA, 1996. *Health Care Finance Rev*. 2001;23(1):161–178.
18. Canady D. Hospitals feeling strain from illegal immigrants. *New York Times*. August 25, 2002;sect 1:16.
19. Agency for Healthcare Research and Quality. Public use file main data results. http://www.meps.ahrq.gov/Puf/PufDetail.asp?ID=69. Published April 2002. Accessed April 4, 2003.
20. Newacheck PW, Inkelas M, Kim SE. Health services use and health care expenditures for children with disabilities. *Pediatrics*. 2004;114(1):79–85.
21. Cochran WG. *Sampling Techniques*. 3rd ed. New York: John Wiley and Sons, Inc.; 1977.
22. Shah BV, Barnwell BG, Beiler GS. *SUDAAN User's Manual*. Research Triangle Park, NC: Research Triangle Institute; 1997.
23. Manning WG, Bailit HL, Benjamin B. Newhouse JP. The demand for dental care: evidence from a randomized trial in health insurance. *J Am Dent Assoc*. 1985;100(6):895–902.

24. Brook RH, Ware JE Jr, Rogers WH, et al. Does free care improve adults' health? results from a randomized controlled trial. *N Engl J Med*. 1983;309(23):1426–1434.
25. Duan N, Manning W, Morris C, Newhouse J. A comparison of alternative models for the demand for medical care. *J Bus Econ Stat*. 1983;1:115–126.
26. Pati S, Shea S, Rabinowitz D, Carrasuillo O. Does gatekeeping control costs for privately insured children? findings from the 1996 Medical Expenditure Panel Survey. *Pediatrics*. 2003;111(3):456–460.
27. Chen AY, Chang RK. Factors associated with prescription drug expenditures among children: an analysis of the Medical Expenditure Panel Survey. *Pediatrics*. 2002;109(5):728–732.
28. Duan N. Smearing estimate: a nonparametric retransformation method. *JAMA*. 1983;78:605–610.
29. Manning WG. The logged dependant variable, heteroscedasticity, and the retransformation problem. *J Health Econ*. 1998;17(3):283–295.
30. Mooney CZ, Duval R. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. Newbury Park, CA: Sage Publications; 1993.
31. Agency for Healthcare Research and Quality. HCUP fact book no. 4: care of children and adolescents in US Hospitals. http://www.ahrq.gov/data/hcup/factbk4/factbk4.htm. Accessed April 8, 2003.
32. Pati S, Shea S, Rabinowitz D, Carrasquillo O. Health expenditures among privately insured adults enrolled in managed care gatekeeping versus indemnity plans. *Am J Public Health*. 2005;95:286–291.
33. Kaiser Family Foundation. Immigrants' health care: coverage and access. http://www.kff.org. Published August 2000. Accessed November 4, 2002.
34. Eisert S, Gabow P. Effect of Child Health Insurance Plan enrollment on the utilization of health care services by children using a public safety net system. *Pediatrics*. 2002;110(5):940–945.
35. Salganicoff A, Wyn R. Access to care for low-income women: the impact of Medicaid. *J Health Care Poor Underserved*. 1999;10(4):453–467.
36. Brown ER, Wyn R, Yu H, Valenzuela A, Dong L. The health and nutritional status of analyses of the Hispanic Health and Nutrition Examination Survey. In: Hernandez DJ, ed. *Children of Immigrants: Health, Adjustment, and Public Assistance*. Washington, DC: National Academy Press; 1999:126–186.
37. Ku L, Matani S. Left out: immigrants' access to health care and insurance. *Health Aff (Millwood)*. 2001;20(1):247–256.
38. Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services. Summary of immigrant eligibility restrictions under current law. http://aspe.hhs.gov/hsp/immigration/restrictions-sum.htm#top. Published October 2002. Accessed November 8, 2002.
39. Zimmerman W, Tumlin K. *Patchwork Policies: State Assistance for Immigrants under Welfare Reform*. Washington, DC: The Urban Institute; 1999. Assessing the New Federalism Occasional Paper, report 24.
40. Illegal Immigration Reform and Immigrant Responsibility Act, Pub L 104–208. 110 Stat 3009. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=104_cong_public_ laws&docid=f:publ208.104.pdf. Published September 30, 1996. Accessed May 3, 2005.
41. Massachusetts Immigrant and Refugee Advocacy Coalition. Unequal access: a report on barriers children in low–income immigrant families face in accessing food stamps and health insurance. http://www.miracoalition.org. Published 2000. Accessed March 6, 2002.

42. The Commonwealth Fund. Insuring the children of New York City's low-income families: focus group findings on barriers to enrollment in Medicaid and Child Health Plus. http://www.cmwf.org/programs/newyork/feld_insuring_children_nyc_305.asp. Published December 1998. Accessed April 10, 2002.
43. Woloshin S, Bickell NA, Schwartz LM, Gany F, Welch HG. Language barriers in medicine in the United States. *JAMA*. 1995;273(9):724–728.
44. Jacobs EA, Lauderdale DS, Meltzer D, Shorey JM, Levinson E, Thisted RA. Impact of interpreters services on delivery of health care to limited-English-proficient patients. *J Gen Intern Med*. 2001;16(7):468–474.
45. Flores G, Abreu M, Olivar MA, Kastner B. Access barriers to health care for Latino children. *Arch Pediatr Adolesc Med*. 1998;152(11):1119–1125.
46. Asch S, Leake B, Gelberg L. Does fear of immigration authorities deter tuberculosis patients from seeking care? *West J Med*. 1994;161(4):373–376.
47. Castel LD, Timbie JW, Sendersky V, Curtis LH, Feather KA, Schulman KA. Toward estimating the impact of changes in immigrants' insurance eligibility on hospital expenditures for uncompensated care. *BMC Health Serv Res*. 2003;3(1):1.
48. Dey AN. Characteristics of elderly home health care users: data from the 1994 National Home and Hospice Care Survey. *Adv Data*. 1996;(279):1–12.
49. Schur CL, Feldman J. *Running in Place: Characteristics, Immigrant Status and Family Structure Keep Hispanics Uninsured*. Bethesda, MD: Project HOPE Center for Health Affairs; 2001.

This page intentionally left blank

CHAPTER 13

Prison Health Care

Madelon L. Finkel, PhD

INTRODUCTION

The provision of health care in prisons has mushroomed into a large, often ignored, global issue of concern. While prisoners are secured behind high fences wired with electronic sensors and cameras or brick or stone barriers, locked doors and barred windows, and cells encased in thick concrete, the probability of developing an illness or disease while incarcerated is quite high. Prison overcrowding, a universal problem, is an instant recipe for spread of disease both within the confines of the prison as well as to the outside population. For example, in New York City in 1989, an outbreak of multidrug-resistant tuberculosis (MDR-TB) was later linked to prisons that had given inadequate treatment. Indeed, the public health implications of the provision of inadequate health and medical care in prisons are potentially serious; inmates released back to the community more often than not have communicable diseases such as TB, HIV/AIDS, and hepatitis C that could be spread to the general population.

The spectrum of health problems in correctional facilities is significant. Inmate health and medical conditions range the gamut from minor (colds or viruses) to the significant (HIV/AIDS and TB). In addition to the communicable diseases, the prevalence of mental health and psychiatric diseases and substance abuse is higher among the prison population than the general population. The scope of this chapter is to provide an overview of the state of health among prisoners, to assess the provision of health care to those incarcerated, to examine the policies regulating care of prisoners including the challenges governments face in their ability to provide health and medical care to inmates, and to discuss the pros and cons of having the private sector (privatization) involved in prison health care delivery.

THE PURPOSE OF PRISONS

Prisons have long been the place of confinement for the punishment and rehabilitation of criminals. Indeed, incarcerating individuals has a long, sordid history.

As long ago as Old Testament times, prisons in Jerusalem were used for detention to confine individuals until corporal or capital punishment could be administered. Prisons and dungeons were used over centuries to hold prisoners who were then transported out of the country, killed, or left to die. Special prisons were built to house specific prisoners. Debtor's prisons, for example, were places where debtors were held until they paid their debts. Penal colonies were formed to exile prisoners from the general population by locating them in a remote location, often an island or distant colonial territory where wardens or governors had absolute authority. Historically, the British and French colonial empires used North America, Australia, and other parts of the world as penal colonies. Regardless of the type of prison, the purpose was to isolate the prisoner from society and mete out punishment through a regimen of strict discipline, including but not limited to forced labor.

Britain holds the distinction of leading the way for prison reform. British prison reformer Howard John (1726–1790) had great influence in improving sanitary conditions and securing humane treatment in prisons not just in England but also throughout Europe. He was responsible for persuading the House of Commons in 1774 to enact a set of penal reforms and improvement in conditions. At the time, prisons were known by the term "penitentiary," suggesting that the purpose was penance by the prisoner, including rehabilitation through education and skilled labor. The modern age of prison reform continued in the 19th and 20th centuries. Incarceration of individuals as a form of punishment, not simply as a holding state until trial or death, was a true break with the past. In North America, two notable women—Elizabeth Fry and Dorothea Dix—advocated for humane treatment of the prisoners and spearheaded prison reforms, including the individualization of treatment, psychiatric assistance, professionalization of correctional officers, the introduction of work release programs, and constructive labor and vocational training. Whereas in the past labor was introduced in prisons chiefly as punishment and to keep discipline among the inmates, such work is now considered a necessary part of the rehabilitation of the criminal.

The chief types of prisons in the United States (with similar institutions in other countries) are the local jail, for pretrial detention and short sentences, and the state and federal penitentiaries for convicts with longer sentences. Special penal institutions house juveniles, the criminally insane, and the military. Modern prison systems include different types of security levels (maximum, medium, minimum). Yet, even in the 21st century, prisons are used as a tool of political repression to detain individuals considered "enemies of the state."

PRISON POPULATION STATISTICS

So many factors have contributed to the burgeoning prison population around the world. As of 2006, it is estimated that at least 9.25 million individuals are imprisoned worldwide. This number is widely believed to be an underestimate as underreporting and lack of accurate data from various countries is acknowledged. In absolute terms, the United States has the largest inmate population in the world with more than 2.5 million prisoners being held in federal or state prisons or local jails.[1] Stated another way, over one-quarter of those incarcerated around the world are housed in the American prison system. Thus, the United States has the dubious

distinction of housing the world's largest prison population and many reasons for this are unique to the United States. From 1980 to 2000, the prison population in the United States increased exponentially. Zero tolerance policies, "three strikes" legislation, and longer sentences each contribute to the situation. Russia and China also have a significant number of individuals imprisoned, and the United Kingdom had 80,000 inmates in its facilities as of 2007, which is one of the highest rates among the western members of the European Union (EU). Poland, however, has the highest imprisonment rate among EU countries (234 prisoners per 100,000 population). [2, 3]

All things being equal, high rates of incarceration in a given nation should correlate with a high crime rate. It is difficult, however, to explain trends in national crime rates and in the type or severity of legal punishment because of differences in definitions of various offenses and reporting methodologies. The higher incidence of violent crimes and use of firearms as contributing factors to the high murder rate in the United States, for example, explain, in part, the differential in the U.S. rate of incarceration compared with other countries. Almost half of the inmates in American prisons and jails are incarcerated for a violent offense, a proportion much higher than that seen in other developed countries. Also, the dramatic increase in the number of individuals incarcerated for a drug offense in the United States is not seen in other countries. Moreover, American sentencing practices seem to be harsher for many offenses, not just for violent crimes. Stated bluntly, the U.S. incarcerates more individuals and for longer periods of time than other countries.

U.S. PRISON STATISTICS

In the United States in 2008, 739 per 100,000 population were either serving time, awaiting trial, or otherwise detained. Local jails held 785,556 persons awaiting trial or serving a sentence at midyear 2008. An additional 72,852 persons under jail supervision were serving their sentence in the community.[4] Ninety percent of jail inmates were adult males, although the number of adult females in jail increased faster than adult males. Differences by ethnic group are quite apparent in the U.S. prison population: 4,777 black male inmates per 100,000 black males compared with 1,760 Hispanic male inmates per 100,000 Hispanic males and 727 white male inmates per 100,000 white males.[5]

Although these statistics refer to the adult prison population, juvenile crime has increased substantially over the past decade. According to the Federal Bureau of Investigation (FBI), juveniles accounted for 17 percent of all arrests, while 15 percent of those arrests were considered violent crimes, and juvenile arrest for drug abuse violations increased considerably. Females made up 28 percent of the total of juvenile arrests. The arrest rates are substantially different for young males compared to females for simple assault charges; young males were accountable for 8 percent of the statistics, whereas young females were arrested in 58 percent of assault cases. There appeared to be a significant increase among young female simple assault arrests as well as for aggravated assaults.

If recent incarceration rates in America remain unchanged, an estimated 1 of every 15 persons (6.6 percent) will serve time in a prison during their life. Lifetime chances of a person going to prison are higher for men (11.3 percent) than for

women (1.8 percent), and for blacks (18.6 percent) and Hispanics (10 percent) than for whites (3.4 percent). Based on current rates of first incarceration, an estimated 32 percent of black males will enter state or federal prison during their lifetime, compared with 17 percent of Hispanic males and 5.9 percent of white males.[6] These data are based on prisoners under the jurisdiction of federal or state correctional authorities on June 30, 2008, collected from the National Prisoner Statistics series.

MEDICAL PROBLEMS OF PRISONERS

In the United States, the Bureau of Justice surveys inmates in state and federal correctional facilities to quantify the incidence and prevalence of medical problems and conditions. These statistics clearly quantify the substantial burden of illness among those incarcerated.[7] It is no secret that prison inmates face health threats behind bars that equal anything they face in the streets. Violent assault, rape, or the outbreak of highly infectious diseases are much more common in correctional facilities than in the general population. Prison conditions can easily fan the spread of disease through overcrowding, poor ventilation, and late or inadequate medical care.

Not surprisingly, communicable diseases among prison inmates, including TB, HIV/AIDS, and hepatitis C, are an international problem. Studies document that prisons are conducive to the spread of TB; it is well known that TB in prisons poses a threat to the general population.[8] For example, globally, the level of TB among those incarcerated is significantly higher than that of the civilian population and may account for up to 25 percent of a country's burden of TB.[9] TB among foreign-born inmates is particularly alarming in the United States. Furthermore, the World Health Organization (WHO) reports that high levels of MDR-TB have been reported from some prisons with up to 24 percent of TB cases being classified as MDR-TB.[9] Clearly, the unregulated, untreated, or erratic treatment of TB in prisons is a huge issue not only for those incarcerated, but also for the general population.

That being said, in the United States, most state and federal inmates receive some type of medical service (medical exam, blood test, TB test) either at admission or while incarcerated. Perhaps as an acknowledgment of the high prevalence of HIV and AIDS among prisoners, the Bureau of Justice survey data show that almost three-quarters of inmates get tested for HIV at or after admission. In California, for example, within two weeks after incarceration in a state prison, an inmate will receive full dental and physical examinations, which include testing for TB and HIV.[10] Given California's serious financial crisis, however, cuts in prison health care could be expected.

Inmate health and medial problems vary by nature of disease, age, gender, and ethnicity of the prisoner. In 2004, an estimated 44 percent of state inmates and 39 percent of federal inmates reported a current medical problem other than a cold or virus. More than one-third of state inmates and nearly a quarter of federal inmates reported having an impairment (learning was the most commonly reported impairment, with many reported to have multiple impairments including speech, hearing, vision, mobility, or mental impairment). Impairments were most commonly reported among inmates who were homeless or who injected drugs in the year before arrest.[7]

The specific needs of female prisoners have received little attention. Women tend to serve shorter sentences and are imprisoned for nonviolent, property, or

drug-related offences compared with men; but many female inmates enter prison with substance abuse dependencies, and reproductive diseases, as well as histories of physical and sexual abuse. Pregnancy is a reality in prison with many inmates being incarcerated while pregnant. Slightly less than 5 percent of state and federal female inmates said that they were pregnant at the time of incarceration, yet only half of these inmates received some type of pregnancy care.[7]

Worldwide, depression, anxiety, and tendency to self-harm also are more prevalent among women prisoners compared with male prisoners.[11] Yet, widespread gender insensitivity and the failure to meet the basic needs of women offenders dominate criminal justice systems. Although overall there are fewer female inmates, their health needs are such that they tend to utilize the prison medical system more frequently. In the United States, females were more than 1.5 times more likely to report two or more current medical problems than male inmates.[7]

Despite the fact that young people commit the majority of crimes, the prison population of older people is growing. Older prisoners have greater health needs than younger prisoners; multipathology is common. Among the adult prison population, arthritis and hypertension were the two most commonly reported chronic medical problems. It has been estimated that 85 percent of older prisoners have more than one major illness, including psychiatric illness.[12] The issue of the older prisoner has been highlighted in the United Kingdom. Older prisoners constitute a comparatively low percentage of the overall prison population; however, between 1990 and 2000, the prison population over the age of 60 tripled, giving clear indication of an aging prison population.[13] The United Kingdom does not have specialist facilities for older prisoners, unlike some U.S. states that have designed "nursing home prisons" and "geriatric prisons."

SEXUAL VICTIMIZATION

Despite professed zero tolerance for sexual victimization in prison, such activity continues. Public awareness of prison rape and sexual assault (male and female) is a relatively recent development although its prevalence has occurred for decades. In 2001, the Human Rights Watch released a paper, "No Escape: Male Rape in U.S. Prisons," that was the seminal event leading up to the passage of the Prison Rape Elimination Act of 2003 (PREA).[14, 15] The Act called for a zero-tolerance policy regarding rape in prisons; called for the development of national standards to prevent and detect sexual violence in state and federal prisons; and provided a framework to make corrections facilities more accountable for incidents of prison rape. PREA also applies to all federal immigration detention centers. Also inherent in the act is the stipulation that failure to comply with the provisions of PREA would result in a 5 percent reduction of federal funding to a facility for each year that it was not compliant.

A major part of the PREA was the establishment of a National Prison Rape Reduction Commission designed to undertake a study on the comprehensive effects and occurrences of prison rape. PREA pertained to adult and juvenile detention facilities. Of note is that young first-time offenders are at an increased risk for sexually motivated crimes. Although juveniles held in adult facilities are five times more likely to be sexually assaulted than juveniles held in juvenile facilities,[16] almost

one-third of all allegations of sexual violence reported in state juvenile facilities were youth-on-youth nonconsensual sexual acts, such as rape and forcible sodomy, and one-fifth were youth-on-youth abusive sexual contact, such as unwanted touching or grabbing with the intention to exploit sexually. During 2005 and 2006, 4,000 allegations of sexual violence involved youth held in juvenile facilities.[17]

One-third of all allegations of sexual violence reported in state juvenile systems and local or private juvenile facilities involved staff sexual misconduct. Females were more likely to be victims of staff sexual violence than victims of youth-on-youth sexual violence. Force or threat of force was more common among male victims in youth-on-youth incidents. The majority of incidents of sexual violence took place outside the victim's room or dormitory. In all substantiated incidents of youth-on-youth sexual violence, one in eight victims were injured and two-thirds required medial attention or counseling.[17]

As of 2007, an estimated 4.5 percent of state and federal prisoners reported sexual assault. Approximately 2 percent of the reported incidents involved another inmate while 3 percent involved facility staff members. These statistics do not take into account whether the act was consensual or not. A small proportion (less than 1 percent) reported being injured by the sexual victimization; injuries included anal or vaginal tearing, knife or stab punctures, broken bones, bruises, black eyes, and other less serious injuries. These statistics are based on a survey that consisted of an audio computer-assisted self-interview in which inmates used a touch screen laptop and interacted with a computer-assisted questionnaire.[18] Inmate participation was voluntary and the results represent inmate allegations. Hence, some inmates may not have felt comfortable completing the survey resulting in an undercount or underestimate of the actual number of incidents, and others might have made false allegations or unfounded allegations.

HIV/AIDS IN PRISONS

HIV/AIDS in state and federal prisoners was more than 2.5 times higher than that in the U.S. population. Approximately 46 per 10,000 prison inmates were estimated to have confirmed AIDS compared with 17 per 10,000 persons in the general population.[19] At the end of 2006, approximately 22,000 state and federal prisoners were HIV-positive or had confirmed AIDS; 1.6 percent of male inmates and 2.4 percent of female inmates in state and federal prisons were known to be HIV positive or have confirmed AIDS. New York, Florida, and Texas housed nearly half of the inmates known to be infected.[19] Twenty-one states reported testing all inmates for HIV at admission or at some time while in custody. Forty-seven states and the federal system reported testing inmates if they have HIV-related symptoms or if they requested an HIV test. Trend data indicate that the overall number of HIV-positive state and federal inmates has decreased from 22,676 to 21,980. Whereas 16 states reported a decrease in the number of HIV-infected prisoners, 25 states reported an increase. Texas reported the largest increase and New York posted the largest decline in HIV-positive cases.[19] HIV/AIDS statistics from other countries also show that HIV/AIDS in prisoners is much higher than in the general population.[20]

MENTAL HEALTH OF PRISONERS

Mental health problems are more prevalent among prisoners than the general population both in the United States and in other countries.[21] Whether or not being in prison exacerbates the mental health problems of a prisoner is not known, but it is clear that having mental health problems is one of the causative factors in imprisonment. Moreover, prisoners with mental health problems pose a greater risk to correctional staff and to other prisoners than inmates without mental health problems.[22] Prisoners who have mental health problems are at higher risk of suicide, a fact recognized by the WHO in its resource document for prison officers.[23]

The U.S. Department of Justice reports that more than half of all prison and jail inmates were found to have a mental health problem defined as a recent history of a mental health problem or recent history of mental health problems including a clinical diagnosis or treatment by a mental health professional. Findings represent inmates' reporting symptoms rather than an official diagnosis of a mental illness. Mental health problems primarily were associated with violence and past criminal activity and inmates with a mental health problem also had high rates of substance abuse or dependence in the year before their admission. Furthermore, among inmates who had mental health problems, approximately 15 percent said that they were homeless in the year before their incarceration. The study found that half of all prison and jail inmates had symptoms of mania, one-third had major depression, and one-fifth had psychotic disorders such as delusions or hallucinations. Female inmates had higher rates of mental health problems than male inmates. Only one in three had received health treatment for their mental illness since admission.[24]

In the early 1980s, suicides had been the leading cause of inmate deaths, but, over time, prison suicide rates have dropped sharply. Yet, small jails (fewer than 50 inmates) had a suicide rate five times higher than the largest jails (2,000 or more inmates).[25] Almost half of the jail suicides occurred during the inmate's first week in custody with 80 percent of suicides occurring in the inmate's cell. In some instances, inmates attempt suicide with no intention of ever completing the act. In other cases, individuals use more lethal methods until they are successful. Of note is that in the United States, the suicide rate among juvenile offenders who are placed in adult detention facilities is almost eight times greater than the rate in juveniles housed in juvenile detention facilities. Most who attempt suicide slash their wrists (as opposed to hanging or overdosing on medication, which are common methods used by completers). Both suicide attempters and completers are generally younger than 25, previously have attempted suicide, have a history of psychiatric treatment, and are likely to be addicted to opiates or other substances.[26]

A national survey conducted in England and Wales in 1999 to 2000 focused on self-inflicted deaths within the prison system.[27] Findings showed that one-third of all suicides occurred within seven days of reception into prison with 11 percent occurring during the first day. Hanging was the most common method of death. Fully trained mental health and correctional staff in prisons are rare because of the lack of qualified professionals working in the system as well as because of budgetary constraints. In fact, prison staff are rarely given adequate training in recognizing, dealing with, and understanding the motivations behind suicidal behavior. A U.K. study on mental health problems in prisons found that no prison doctors in their

study had specialist training and less than a quarter of nurses had mental health training. The authors concluded that the mental health services offered to prisoners fell below the standard of the National Health Service. Given the high incidence of mental health problems among prisoners, this is a matter of great concern.[28]

Assessment is an important factor in addressing mental health problems in prisoners. Calls for action in this regard date from the mid-1990s when Birmingham et al. called for assessing prisoners for mental health problems as soon as they enter the prison system and Metzner et al. called for the use of standardized assessment procedures to screen for mental illness.[29, 30]

SUBSTANCE ABUSE

The United States now incarcerates people at a rate nearly five times the world average primarily because of drug offenses. More than two-thirds of jail inmates were found to be dependent on or to abuse alcohol or drugs. Furthermore, half of all convicted jail inmates were under the influence of drugs or alcohol at the time of their offense.[31] Substance abuse and the enforcement of antidrug laws have fundamentally affected the growth of America's prisons and jails over the past 20 years and the types of inmates they house. This rise corresponds to the focus on the War on Drugs now entering its 40th year in the United States. Enforcement of harsh drug laws has served to fill the prisons with minor players in this war (possession of small amounts of illegal substances) and has done little to abate the buying and selling of narcotics and other illegal substances.

In the United States, an overwhelming majority of state and federal inmates as well as jail inmates were either convicted of substance-related crimes such as drug selling or driving while intoxicated; were under the influence of drugs or alcohol at the time of their crime; committed a crime to get money to buy drugs; or had histories of regular illegal drug use or alcohol abuse.[32] More than 277,000 individuals, which accounts for 20 percent of state prisoners, and more than 60 percent of federal prisoners have been put in jail as a result of state and federal drug laws. Statistics from the United States show that half of state and federal inmates were abusing or were dependent on drugs in the year before their incarceration. Nearly half of violent offenders in state prisons met the criteria for recent drug dependence or abuse and more than a quarter of these individuals committed their current offense while under the influence of drugs. Ten percent said that the need for money for drugs was a motive for their crimes.[33]

Women were more likely than men to have used drugs before their incarceration. In fact, some studies have shown that women appear to have greater problems with mental health and substance abuse than men. One study identified that women in prisons used drugs more frequently and that they used harder drugs compared with male prisoners.[34] Acknowledging this situation, state and federal prisons have implemented drug treatment programs. Among drug-dependent or -abusing prisoners, almost half of state and federal inmates took part in drug abuse treatment programs since their admission to prison.

The use of illegal drugs was recognized as a major category for health promotion among prisoners by the National Institute of Corrections in the USA in 1991

[35] and the WHO in 1999.[36] Clearly, a strong relationship exists between substance dependency and crime and mental health problems among prisoners.

HOW SHOULD PRISON HEALTH CARE BE PROVIDED AND FUNDED?

The delivery of prison health care ranges from that delivered by prison service employees to care delivered by on-site nurses or doctors. The differences in the quality of care are substantial and, globally, health promotion as well as illness prevention are almost nonexistent. The focus tends to be on immediate curative care. Regardless of the mode of delivery, providing health care to the prison population is expensive and, as the above indicates, many medical problems among inmates are not being addressed well. Poorly trained correctional officers and poor quality of health care are just a few of the problems the correctional system must address. Furthermore, the cost of delivering health care to prisoners is increasing rapidly. The annual median cost of incarcerating a prisoner in 2003–2004 was $28,000 per state prisoner in the United States, $45,000 in Australia, and $53,000 in Great Britain.[37] States spend 10 percent of their corrections budgets on average to cover the cost of inmate health care—a total of more than $3 billion annually. One could argue that prison health care services are expensive primarily because the focus is on curative care rather than on preventing disease and disease surveillance.

Clearly something needs to be done as the costs of providing health care are substantial and many governments (federal and state) are having a difficult time keeping a lid on prison health service budgets. California now spends about two and a half times as much per prison inmate as it does per student in the University of California system. Texas has the highest rate of incarceration in the country, and, as such, has particular reason to worry about increasing health costs. The state pays nearly $270 million for inmate health care per year, which translates to approximately $2,150 per inmate.[38]

The need for governments to contain prison health care costs has economic as well as ethical implications. A nationwide public debate erupted in the United States when in 2002 a twice-convicted felon received a $1 million heart transplant. In another case, $5.1 million was spent in five months to treat a 49-year-old hemophiliac and diabetic who was serving a 15-month jail sentence.[39] How then should care be provided? What are prisons obligated to provide? Who should pay for the services? How can quality be assured?

The notion of universal access to health care for those incarcerated has its roots in the 1929 Geneva Convention, and more recently through the efforts of Amnesty International and the International Red Cross. The Council of Europe in 1989 issued a directive stating that prisoners should have the same access to health care as the nonincarcerated population, and that the health care provided to prisoners should be equivalent to that provided to the nonincarcerated population. In the United States, prison inmates are the only citizens with a constitutional right to health care. A prisoner's constitutional right to health care dates back to a 1976 U.S. Supreme Court decision in the Texas case, *Estelle v. Gamble*. In this case, judges ruled that deliberate indifference to a prisoner's serious illness or injury constitutes cruel and unusual punishment under the Eighth Amendment. (A precursor to this dates from the English Bill of Rights of 1689 in which Parliament stipulated among other things that

cruel and unusual punishment ought not to be inflicted). Regardless of the country, the big issue in providing medical care for prisoners is cost and availability of trained personnel.

At the moment, more than half of the U.S. states are trying to recoup money by collecting a copayment from prisoners. All federal prisons and 70 percent of state prisons require inmates to pay a copayment for their medical care.[40] Chronically elderly prisoners, a group that generally has greater health care needs, are least able to afford a copayment and generally suffer more as a result. In some states, the copayment is waived for indigent prisoners. In an effort to reduce prison health care costs, eight states—Connecticut, Louisiana, Michigan, Montana, New York, South Carolina, Tennessee, and Texas—have established medical parole laws to get aging and seriously ill prisoners who no longer pose a risk to society out of prison.

In the United States, most jail and prison systems pay for and maintain an in-house infirmary. Since the 1970s, however, many states have started contracting with privately run groups to manage the health of their inmates. The extent to which the private sector should be involved in prison health care delivery is being hotly debated in the United States and in Europe. Some individuals view privatization of prison health care as a better means of managing health care costs and improving quality of health care delivered. Opponents of prison privatization posit that cost savings from privatization are usually the result of cuts in services. Although prison health care remains largely in the public sector, movement toward privatization has been seen in both the United States and in Europe. The United Kingdom has the most privatized prison system in Europe with 10 percent of prisoners housed in 11 private prisons, many of which have not been performing well with regard to providing health care to the inmates.[41]

Thirty-four U.S. states have some privatized health care, and in 24 states, inmate health care systems were run completely by private contractors. In the United States, the largest correctional health care firm, St. Louis–based Correctional Medical Services, Inc. (CMS), insures one out of every seven inmates nationally. The company provides health services to more than 273,000 inmates at 334 correctional facilities in 29 states.[38] For nearly 20 years, CMS and a few other managed care companies have been taking over prison and jail health care from government agencies coast to coast. Concerns of the quality of care delivered, however, has been the subject of intense scrutiny. A big private correctional health care provider, Prison Health Services, Inc., has been the subject of a growing list of litigation from prisoners who allege deliberate disregard for their health care needs.[42] Because prisoners are not paying customers, there is little incentive to provide good quality care. In the United States, the National Commission on Correctional Health Care is responsible for accrediting prison health facilities, but they do not serve as a watchdog organization monitoring health care abuses in prisons and jails.

In the United Kingdom, the provision monitoring of prison health care had been under the purview of the Prison Health Service (PHS), established in 1877. When the National Health Service (NHS) was formed in 1948, the PHS was incorporated into the NHS. But in 2003, funding for prison health services was transferred from the Home Office (the government department responsible for Her Majesty's Prison Service) to the Department of Health (the government department responsible for running the NHS). Prison health care was now delivered as part of the primary

health care system, which had profound implications for general practitioners, many of whom provided medical care for the local prison population. Like the United States, the United Kingdom spends a substantial amount of money on providing health care to prisoners.

One approach for improving the efficiency of prison health services is public-private health partnerships (PPHP) supported by the WHO. In 1993, the World Health Assembly called on the WHO to encourage the support of PPHPs as a strategy for bringing about health for all.[43] The WHO PPHPs bring together industry associations and government organizations to work toward a health goal that is based on a mutually agreed-on and clearly defined division of labor. The WHO has been involved in more than 70 global health partnerships to achieve health goals on the basis of a mutually agreed-upon and well-defined division of labor.[44] The PPHP initiative has a strong potential to capitalize on the presumed advantages of public and private sectors jointly providing prison health services. Time will tell how successful this initiative actually is.

CONCLUSION

A high incarceration rate comes with both an economic and social price. Housing prisoners is expensive regardless of whether the public or private sector is responsible for providing prison health care. The social costs to the prisoner as well as to his or her family is also substantial. New research focusing on the long-term harm to the children of the incarcerated parent shows that these children are more likely to be adversely affected psychologically. Physically aggressive behavior, social isolation, depression, and problems in school are frequently seen among children with an incarcerated parent. Among five-year-old urban boys, almost half of those who had a father in prison within the previous 30 months were found to have physically aggressive behaviors compared with 38 percent of those who did not have a father imprisoned.[45]

Christopher Wildeman, who has been studying the "incarceration generation," has sounded the alarm that parental imprisonment is a new U.S. childhood risk, particularly so among minority children and children of parents with low education.

It is estimated at more than 1.5 million children have a parent in prison (usually a father, but it not uncommon for the mother to be in prison). A child born in 1990, for example, was more likely to have had a parent in prison by age 14 than one born in 1978. This is particular1y so for black children compared with white children. One in four black children compared with 1 in 25 white children had a father in prison by the time they were 14 years old.[45]

Another equally disturbing trend is the record number of inmates serving life sentences. The Sentencing Project, a group that calls for the elimination of life sentences without parole, looked at increases in life sentences from 1974 to 2008. In 2008, 140,610 of the 2.3 million inmates incarcerated in the United States were serving life terms. This translates into one in every six inmates in the United States serving a life sentence; and in California, the number is one in every five prisoners. In Alabama, Massachusetts, Nevada, and New York, the number is one in six prisoners serving a life sentence.[46] The rising number of inmates serving life terms is placing a strain on the corrections' budgets. As prisoners serve longer sentences, the

number of aging inmates has risen sharply, and these older inmates cost more than younger prisoners because the former have more health needs. This trend seems to indicate a retreat or even an abandonment of the rehabilitative or "corrective" function of prison.

Addressing the serious issues surrounding the provision of prison health care will take economic as well as political will. The public health case for action is clear, but the political commitment seems weak or nonexistent. In 1995, the WHO Health in Prisons Project (HIPP) tried to address the gap between public health and prison health. HIPP's main objective is to reduce the public health hazards associated with prisons along with protecting and promoting health in prisons by facilitating links between prison health and public health systems at the national and international level. In 2003, a Steering Group for Prisons and Health was established to act as an international steering committee for HIPP and to enhance international cooperation within the network. While laudable, much still remains to be done.

In 2005, WHO distributed what an editorial in the *Lancet* called one of the most important documents on prison health ever published.[47] The report, "Status Paper on Prisons, Drugs, and Harm Reduction," showed that infectious disease transmission in prisons can be prevented by simple and inexpensive harm-reduction strategies. Harm-reduction efforts in prisons aim to prevent or reduce the negative health effects associated with prison (overcrowding, spread of infectious diseases, mental health problems). Although intuitively simple and logical, sadly there is little political pressure to at least try these initiatives. Out of sight, out of mind seems to be the prevailing thinking when it comes to prison health reform. Certainly the provision of health care to inmates has improved, but much more still needs to be done. Whether privatization of prison health care will be successful is yet to be shown. Whether harm-reduction efforts will be introduced in prisons also remains to be seen. Meanwhile, infectious diseases transmitted or exacerbated in prisons have the potential to become full-blown public health problems when prisoners return to their communities.

REFERENCES

1. Walmsley R. World prison population list. 7th ed. http://www.prisonstudies.org.
2. Wikipedia. Prisoner population rate 2007–2008. http://en.wikipedia.org/wiki/File:Prisoner_population_rate_UN_HDR_2007_2008.PNG.
3. Mauer M. Comparative international rates of incarceration: an examination of causes and trends: Presented to the US Commission on Civil Rights; June 20, 2003; Washington, DC. http://www.sentencingproject.org.
4. US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. Correction statistics. http://www.ojp.usdoj.gov/bjs.
5. Ibid.
6. Federal Bureau of Investigation. FBI uniform crime report. http://www.fbi.gov/ucr.htm.
7. Maruschak LM for US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. Medical problems of prisoners. http://www.ojp.usdoj.gov/bjs.
8. Griefinger RB, Heywood NJ, Glaser JB. Tuberculosis in prison: balancing justice and public health. *J Law Med Ethics*. 1993;21:332–341.

9. World Health Organization. Tuberculosis in prisons. http://www.who.int/tb.
10. Shalit M, Lewin MR. Medical care of prisoners in the USA. *Lancet*. 2004;364:34–35.
11. Van den Bergh BJ, et al. Women's health in prison: urgent need for improvement in gender equity and social justice. *Bull World Health Organ*. 2009;87:406.
12. Fazel S, et al. Health of elderly male prisoners: worse than the general population, worse than younger prisoners. *Age Ageing*. 2001;30:403–407.
13. Howse K. *Growing Old in Prison*. London: Prison Reform Trust; 2003.
14. Smith BV. The Prison Rape Elimination Act: implementation and unresolved issues criminal law brief. http://papers.ssrn.com/sol3/papers.cmf?abstract_id=1129810. Washington College of Law Research Paper No.2008-49OCLC 63521701.
15. Wikipedia. Prison Rape Elimination Act of 2003. http://en.wikipedia.org/wiki/Prison_Rape_Elimination_Act_of_2003.
16. US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. Juvenile justice and delinquency programs. http://www.ojp.usdoj.gov/bjs.
17. US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. Sexual violence reported by juvenile correctional authorities: 2005–2006. http://www.ojp.usdoj.gov/bjs.
18. US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. 4.5 percent of inmates report sexual victimization in the nation's state and federal prisons. http://www.ojp.usdoj.gov/bjs.
19. Maruschak LM for US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. HIV in prisons: 2006. http://www.ojp.usdoj.gov/bjs.
20. World Health Organization. *Health in Prisons Project*. Geneva, Switzerland: World Health Organization; 1999.
21. Watson R, Stimpson A, Hostick T. Prison health care: a review of the literature. *Int J Nurs Stud*. 2004;41:119–128.
22. Hilton NZ, Simmons JL. The influence of actuarial risk assessment in clinical judgments and tribunal decisions about mentally disordered offenders in maximum security. *Law Hum Behav*. 2001;25:393–408.
23. World Health Organization. *Preventing Suicide: A Resource for Prison Officers*. Geneva, Switzerland: World Health Organization; 2000.
24. US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. Mental health problems of prison and jail inmates. http://www.ojp.usdoj.gov/bjs.
25. US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. Suicide and homicide in state prisons and local jails. http://www.ojp.usdoj.gov/bjs.
26. Daniel AE. Preventing suicide in prison: a collaborative responsibility of administrative, custodial, and clinical staff. *J Am Acad Psychiatry Law*. 2006;34:165–175.
27. Shaw J. et al. Suicide by prisoners. *Br J Psychiatry*. 2004;184:263–267.
28. Reed JL, Lynn M. Inpatient care of mentally ill people in prison: Results of a year's program of semi-structured inspections. *Br Med J*. 2000;320:1031–1034.
29. Birmingham L, Mason D, Grubin D. Prevalence of mental disorder in remand prisoners: consecutive case study. *Br Med J*. 1996;313:1521–1524.
30. Metzner JL, Grande TP, Chambers MR. The incarceration of individuals with severe mental disorders. *Community Ment Health J*. 2001;37:361–372.
31. Karberg JC, James DJ for US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. Substance dependence, abuse, and treatment of jail inmates: 2002. http://www.ojp.usdoj.gov/bjs.
32. National Center on Addiction and Substance Abuse at Columbia University. *Trends in Substance Abuse and Treatment Needs among Inmates Fund Report*. New York: Columbia University; 2002.

33. US Department of Justice, Bureau of Justice Statistics, Office of Justice Programs. Drug use and dependence, state and federal prisoners: 2004. http://www.ojp.usdoj.gov/bjs.
34. Langan NP, Pelissier BMM. Gender differences among prisoners in drug treatment. *J Subst Abuse*. 2001;12:291–301.
35. National Institute of Corrections. *Intervening with Substance Abusing Offenders: A Framework for Action.* Washington, DC: US Department of Justice; 1991.
36. World Health Organization. Health in prisons project. Geneva, Switzerland: World Health Organization; 1999.
37. Awofeso N. Making Prison health care more efficient. *BMJ*. 2005;331:248–249.
38. Kaplan S. Health care costs rising as prison population grows and ages. http://www.stateline.org.
39. Shalit M, Lewin MR. Medical care of prisoners in the USA. *Lancet*. 2004;364:34.
40. Federal Prisoner Health Care Co-payment Act of 2000. Washington Public Law 106–294. October 12, 2000.
41. Awofeso N. Debate: public-private health partnerships in prison health care-the case for a public health focus. *Public Money Manage*. 2007;305–306.
42. von Zielbauer P. As health care in jails go private, 10 days can be a death sentence. *New York Times*. February 27, 2007.
43. Buse K, Waxman A. Public private health partnerships: a strategy for WHO. *Bull World Health Organ*. 2001;79:748–754.
44. World Health Organization. The WHO health in prisons project. http://www.euro.who.int/prisons.
45. Wildeman C. Parental imprisonment, the prison boom, and the concentration of childhood disadvantage. *Demography*. 2009;46:265–280.
46. http://www.sentencingproject.org.
47. Prison health: a threat or an opportunity? *Lancet*. 2005;366(9479):1.

CHAPTER 14

Lesbian, Gay, Bisexual, and Transgender Health: A Neglected Issue

Anita Radix, MD, MPhil, MPH, and Gal Mayer, MD, MS

Lesbian, Gay, Bisexual, and Transgender (LGBT) communities have existed in all cultures throughout recorded history; however, it is only in the last two decades that the health of sexual and gender minorities has emerged as an important discipline within public health practice. Before an accurate assessment can be made of the health issues facing the LGBT population, it is important to first determine what is meant when we refer to these groups. Sexual orientation refers to the inclination or capacity to develop emotionally and sexually intimate relationships with people of the same gender, other gender, or either gender. A woman who is primarily oriented to other women is lesbian; a man primarily oriented to other men is gay, while bisexuals are oriented toward both women and men.

Determination of gender is complex. While a person may have a combination of genetic and anatomic traits that define them as male, female, or intersexed at birth, gender is the self-identification with the social, psychological, and emotional traits associated with masculinity or femininity. Transgender individuals are those with gender identities that do not fully conform to their assigned sex at birth. Individuals determined to be female at birth (natal females) may identify as men (transgender men, transmen), while natal males may identify as women (transgender women, transwomen). Some transgender individuals do not fully embrace either male or female roles, preferring to define themselves outside of traditional definitions. It is important to recognize that there is considerable fluidity in how people define their sexual orientation and gender identity depending on their life experience, ethnicity, culture, and stage in the "coming out" process.

It became clear in the early years of the AIDS epidemic that both HIV and other sexually transmitted infections (STIs) were clustered within the communities of men who have sex with men (MSM). It was evident that public health officials did not have the necessary expertise to undertake effective health promotion activities among MSM or other members of the LGBT community. The reasons for this were many and included a lack of population-based research conducted in the LGBT community, lack of political will to target a socially stigmatized group, and

the difficulty of reaching members of a community often reticent of declaring their sexual identity to government officials.

Amid the AIDS-hysteria of the early 1980s, and refusal of some clinicians to provide care to people with HIV or perceived to be at risk for HIV, several clinics led the way in providing sensitive and nonjudgmental care to gay and bisexual men. The Community Health Project in New York, New York (now Callen Lorde Community Health Center), the Fenway Community Health Center in Boston, Massachusetts, and the Howard Brown Clinic in Chicago, Illinois, are a few examples. The inclusion of HIV/AIDS as a disability under the Americans with Disabilities Act in 1998 was an important milestone in securing rights for people with HIV, many of whom were also gay or bisexual men.

An important outcome of the devastating effect that HIV/AIDS had on the LGBT community was the recognition that disparities exist in other areas not related to HIV or sexual health issues. The reasons for these disparities are complex and related to both historic and contemporary factors. Social stigma and systemic discrimination based on sexual orientation and gender identity have resulted in multilevel barriers to health care for the LGBT population. Although population-based research has been limited, the Health and Human Services Secretary's Advisory Committee on Healthy People 2020 has recognized the LGBT community as a "special" population. The Healthy People 2010 companion document highlighted the fact that the community suffers disproportionately from higher rates of health disorders such as substance and tobacco use, eating disorders, obesity, depression, infectious diseases, and reduced uptake of preventive health services.[1]

CLINICAL RESEARCH INTO LGBT POPULATIONS

For public health departments and providers to plan appropriate services for a population, especially a vulnerable one, it is essential to have reliable data on population demographics and health status. Until recently, many of the research studies conducted in the LGBT community were community-based studies using nonprobability sampling techniques.[2, 3] The studies, often using convenience sampling, were prone to selection bias. For example, studies conducted at bars and clubs were known to have overestimated certain health issues such as the prevalence of tobacco and substance use. In addition, the use of small nonprobability samples to examine the health of LGBT communities may not be representative of the population's racial, ethnic, and socioeconomic composition. To adequately describe the LGBT population and assess health indicators and risks, it is therefore necessary to conduct well-designed population-based studies. An additional issue with investigating the LGBT community is that no single question adequately captures the diversity and complexity of the lives of LGBT individuals. Self-disclosed sexual orientation has been shown to not predict sexual behavior. A person who identifies as lesbian or gay may still have sexual relations with members of the opposite sex, for example.

The first national survey on sexual behaviors among U.S. adults examined sexual orientation, attraction, and identity. This study highlighted the difficulty with defining sexual identity in this population. Findings showed that 9.9 percent of men and 4.3 percent of women reported having same-sex partners; 7.7 percent of

men and 7.5 percent of women have same-sex desire/attraction, however, only 2.8 percent of men and 1.4 percent of women reported self-identification as homosexuals. Homosexual identity was found to be less likely among African Americans, those living in rural areas, and those with less formal education.[4] Another study of men in New York City found considerable discordance between sexual orientation and behavior with 9.4 percent of straight-identified men reporting sexual intercourse with at least one man in the previous year.[5] The 2002 National Survey of Family Growth (NSFG) also measured indicators of sexual behavior, orientation, and identity among males and females, 15–44 years of age, using face-to-face interviews. While 4.1 percent of both men and women identified as homosexual or bisexual, a greater number reported same-sex activity: 6 percent of men reported having either oral or anal sexual contact with another man, while 11 percent of women answered yes when asked, "Have you *ever* had any sexual experience of any kind with another *female*?" Four percent of women and 2.9 percent of men reported same-sex partners in the last 12 months. The survey also asked about sexual attraction, with 3.2 percent of men and 3.4 percent women reporting same-sex attraction to males and females.[6] Studies conducted among adolescents have also shown a discordance of self-reported sexual orientation and the gender of sexual partners, with lesbian youth more likely than heterosexually identified women to report high-risk sexual activity with opposite-sex partners.[7, 8]

Sexual orientation and gender identity have not been included routinely as demographic markers in population-based research. One key source of demographic data, the U.S. Census, does not include a question on sexual orientation. In 1990, the category of "unmarried partner" appeared on the census form for the first time, allowing enumeration of unmarried, same-sex partner households. In the 2000 census, of 4.9 million unmarried-partner households, approximately one in nine (594,000) had partners of the same sex.[9] Census 2010 will again gather information on same-sex partner households, and in a reversal of a previous decision, will allow married same-sex couples to indicate their status as "husband or wife" instead of "unmarried partner." Transgender individuals are not separately identified on the U.S. Census. Several LGBT organizations, most notably the National Coalition for LGBT Health, based in Washington, D.C., have been advocating for the addition of sexual orientation and gender identity questions in the National Health Interview Survey (NHIS), one of the major data collection programs of the National Center for Health Statistics (NCHS). The inclusion of these demographic markers in national random-sample population-based studies will greatly increase the knowledge of health conditions and risk factors affecting the LGBT community.

In addition to the paucity of population-based studies that address LGBT individuals, there remains the difficulty of increasing participation in studies by members of the community. Despite liberalization of societal norms, LGBT participants may be uncomfortable answering sensitive questions about their sexual orientation, gender identity, and sexual practices on surveys. There is currently no federal-level protection against housing or employment discrimination on the basis of sexual orientation or gender identity, which could affect participation of LGBT in research studies, especially if they fear public disclosure of confidential information. For certain individuals (for example, those in the military), disclosure of sexual orientation can have disastrous consequences such as loss of employment, dishonorable

discharge, and loss of pensions and other benefits. For men who identify as gay or bisexual, there is an added concern of stigmatization for their perceived increased risk of HIV.

BARRIERS TO ACCESS TO CARE

The LGBT community faces several barriers to access to care that qualifies them as a special or vulnerable population. In particular, some barriers include structural (for example, lack of LGBT inclusive demographic intake forms), financial (for example, insurance benefits), and personal (for example, fear of disclosure and lack of provider cultural competency). In 1998, the American Medical Association stated that "unrecognized homosexuality by the physician or the patient's reluctance to report his or her sexual orientation can lead to failure to screen, diagnose, or treat important medical problems" [10] and noted that barriers to care exist within the patient-clinician relationship. Patients are often not willing to disclose their sexual orientation to medical providers due to past negative experiences or perceived threat of discrimination and substandard care.

Studies on the utilization of health care services have revealed that a significant number of lesbians and bisexual women surveyed concealed their sexual orientation or behavior when seeking medical care and up to 40 percent feel that disclosure would hinder the quality of medical care.[11, 12] Fears of discrimination are not unfounded with documented reports of discriminatory attitudes held by medical providers and nurses to the sexual orientation of their patients.[13, 14]

In addition to negative attitudes, LGBT persons often encounter medical providers who lack knowledge of the health issues and counseling needs relevant to this cohort.[15, 16] Medical education in the United States, both during medical school and in residency, is often unlikely to include adequate cultural competency related to the care of sexual orientation and gender identity minorities. A survey conducted to assess curricula in U.S. medical schools found that less than 3.5 hours were dedicated to teaching about health issues related to homosexuality.[17]

Financial barriers also exist, which further reduce access to health care. The most recent national survey conducted by Harris Interactive revealed that nearly one in four gay and lesbian adults (22 percent) lack health insurance compared with 12 percent of their heterosexual counterparts.[18] The lack of domestic partner benefits and tax penalties in instances where benefits are offered create a situation in which members of the community are more likely to go uninsured or underinsured.

HEALTH ISSUES AFFECTING LGBT COMMUNITY

The Healthy People 2010 Companion Document for LGBT Health highlighted several areas of concern that affect all subpopulations within this community, including mental health, tobacco, and substance use. In 1973, the American Psychiatric Association removed homosexuality from its official diagnostic manual, the *Diagnostic and Statistical Manual of Mental Disorders, Second Edition* (DSM-II). With this act, the association no longer characterized homosexuality as a mental illness. In the years that followed this major change, several studies set out to evaluate the prevalence of mental health disorders in the community. Despite homosexuality

being depathologized, there have been concerns that social stressors could lead to a higher prevalence of mental health disorders among sexual orientation and gender identity minorities. LGBT people face multiple stressors in their daily lives, such as rejection by peers or family, social stigma, sexual orientation discrimination, transphobia, and violence.[19]

For many adolescents, there also is the added burden of acknowledging one's lesbian or gay identity, often without family or social support, while also struggling to develop an adult identity.[20] The population-based studies that have included questions on sexual orientation have revealed that rates of depression, anxiety, and suicide attempts are higher among men and women with same-sex partners.[21, 22, 23, 24, 25] The National Survey of Midlife Development in the United States (MIDUS) found that 2.5 percent of participants reported being homosexual. Both major depression and panic disorders were higher in homosexual men, while generalized anxiety disorders were higher in lesbians. Both groups had higher rates than heterosexuals of two or more mental health disorders, and they used mental health services more than their heterosexual counterparts.[26]

A recent study designed to evaluate the prevalence of psychiatric disorders among lesbian, gay, and bisexual black, Latino, and white individuals revealed that white lesbian, gay, and bisexual individuals had a higher prevalence of anxiety, mood, and substance use disorders compared with black and Latino participants; however, black and Latino lesbians, bisexuals, and gays had higher rates of serious suicide attempts (statistically significant for Latino participants) that mainly occurred in adolescence and young adulthood.[27] Supporting the theory that social stressors are the main cause for mood and anxiety disorders, rather than sexual orientation minority identity, a recent study showed that rates of depression and present suicidal ideation among lesbian, gay, and bisexual youth was no higher than among heterosexual youth when adjustment was made for markers of stress, such as the number of positive events and negative events, satisfaction with social support, and acceptance coping.[28]

TOBACCO USE

Tobacco use is the single most preventable cause of morbidity and mortality in the United States. Current estimates are that between 2000 and 2004, more than 400,000 persons died prematurely from tobacco-related diseases such as lung cancer, cardiovascular disease, and chronic obstructive pulmonary disease.[29] In 2007, the median prevalence of adults currently smoking in the 50 states and Washington, D.C., was 19.8 percent and higher among men (21.3 percent) than women (18.4 percent).[30] The rates of smoking have been estimated to be higher among lesbian, gay, and bisexual persons, with estimated smoking rates up to 59 percent for gay and lesbian youth and up to 50 percent for adults.[31, 32] The prevalence of smoking among transgender persons has not been studied adequately; however, small studies suggest that the prevalence rates are higher than the general population.[33, 34] Many factors may have contributed to the higher rates, including the daily stresses of societal homophobia and also the fact that gay and lesbian bars, places of high smoking prevalence, have traditionally been sites of refuge and important components of gay and lesbian social life.

The tobacco industry has directly targeted the LGBT community with cigarette advertising and pro-tobacco imagery frequently found in the lesbian and gay press.[35] Unfortunately, many LGBT individuals do not view tobacco use as a "gay health issue" and may view the targeting as positive in that it legitimizes the community by including them in mainstream marketing campaigns.[36] Further research is needed to investigate ways in which sexual orientation, gender identity, and social factors influence smoking rates, and how this can then create culturally appropriate tobacco-cessation programs for the community.

LESBIAN HEALTH

In 1997, the Institute of Medicine (IOM) Committee on Lesbian Health Research Priorities was convened to discuss challenges involved in conducting health research on lesbian health and summarizing the known health issues affecting this population. The lack of well-conducted longitudinal studies was cited as a key issue in the paucity of data on lesbian health along with the difficulty in defining a population characterized by sexual, socioeconomic, racial, and ethnic heterogeneity.[2] Health issues most relevant to lesbians were stated as smoking, obesity, cancer, sexual health, and lack of engagement in primary preventive care and necessary screenings.

The Centers for Disease Control and Prevention (CDC) does not include women who have sex with women (WSW) as a separate risk category for the surveillance of sexually transmitted diseases (STDs), as they appear to be at lower risk for many STDs, such as HIV, gonorrhea, chlamydia, and syphilis. It is important to recognize, however, that lesbian sexual practices may include oral-genital and oral-anal sex, vaginal and anal sex using fingers, or shared sex toys, all activities that can facilitate transmission of STDs, such as human papillomavirus (HPV) and herpes simplex virus (HSV) between partners. Another fact is that many lesbians have had and continue to have male sexual partners, some of whom may be gay and bisexual men.[37] Approximately 17 percent of lesbian-identified women in one cross-sectional survey had been diagnosed with an STI, and an equal number had an abnormal Papanicolaou test. Almost two-thirds had engaged in penile-vaginal and 15.8 percent in anal-receptive intercourse without use of condoms, placing them at high risk for STIs, including HIV.[38] Bacterial vaginosis, a frequent cause of vaginal discharge, has been found to occur 2.5 times more frequently among lesbians, with high concordance among their female partners, suggesting this should be considered a sexually transmitted condition for women who have sex with women.[39, 40]

The U.S. Preventive Services Task Force (USPSTF) recommends routine screening for early detection of breast and cervical cancer in women; however, studies, including two recent population-based studies, have shown that lesbians and WSW are less likely to undergo clinical breast exams, mammography, and Pap tests.[41, 42, 43, 44] In Canada, using data from a large Canadian Community Health Survey, only 64 percent of lesbians had received a Pap test within three years compared with 77.1 percent of heterosexual women. Lesbians were significantly less likely to have visited their family physician and more likely to have an unmet medical need in the last 12 months compared with their heterosexual counterparts.[44]

Since Canadians have universal health access, these data suggest that factors other than insurance affect the uptake of preventive health by lesbians. Although lesbians are less likely to undergo screening, they have similar behavior risk factors as heterosexual women for cervical cancer, such as history of male sexual partners. Lesbians also may be less likely to use safer sex practices, have an earlier age of initiation of sexual activity, and have a higher number of sexual partners than heterosexual women.[45]

Despite these facts, primary care providers may not advise lesbian-identified women to engage in cervical cancer screening due to the perception that lesbians are not at risk for cervical cancer. Some Lesbians themselves may feel that they are not at risk for cervical cancer and therefore defer screening. They may avoid gynecological care due to discomfort with medical providers or reluctance to disclose their sexual orientation. Although no studies have shown an elevated risk for breast cancer, lesbians may be at higher risk due to the presence of risk factors, such as increased rates of nulliparity or older age at first childbirth. There is a lack of data available to determine whether colorectal cancer risk or screening rates differ between lesbians and heterosexual women.

Obesity and overweight affect a steadily increasing proportion of the U.S. population. In 2005, approximately 24 percent of U.S. adults were obese and more than 65 percent were overweight.[46] There is a paucity of research into the prevalence of obesity and overweight among sexual orientation and gender identity minorities; however, a few studies have shown that lesbians are more likely to be overweight or obese,[47, 48] which may place them at increased risk for obesity-related conditions such as hypertension, diabetes, coronary heart disease, cerebrovascular disease, and dyslipidemias. Further investigation is needed into the factors related to obesity among lesbians as well as for the development of targeted interventions to address the epidemic of obesity in this community.

TRANSGENDER HEALTH

Few data are available on the number of transgender individuals in the United States; however, studies from the Netherlands estimate a prevalence of 1 in 11,900 males and 1 in 30,400 females.[49] Although no population-based studies have addressed transgender health care issues, small nonrandom surveys suggest that this group experiences significant health disparities especially in the areas of violence, HIV disease, and other STDs. Like other members of the LGBT community, people of transgender experience also face barriers to health care access due to stigma and discrimination, fewer options for health insurance and lack of knowledge by medical providers about their specific health concerns.

The National Coalition of Anti-Violence Programs (NCAVP) has reported a steady increase in the number of transgender individuals experiencing gender-identity-related bias crimes, including sexual assault and murder between 2007 and 2008.[50] Transgender individuals experience stigma in other areas of daily life and are not protected against work, housing, or education discrimination in most locales of the United States. This social marginalization forces many transgender individuals to engage in survival sex work. In Texas, more than 50 percent of transgender women in one study had engaged in transactional sex

for either money or drugs.[51] Similar high rates have been found among transgender youth.[52]

The CDC does not collect separate statistics on transgender individuals and classifies male-to-female transgender individuals as MWM for data-collection purposes. Several prevalence surveys and an incidence study have revealed HIV rates among transgender women that are among the highest in the United States. A cross-sectional survey conducted among 392 male-to-female transgender persons in California revealed that 35 percent were HIV-positive.[53] The prevalence among African American transgender women was even higher at 63 percent. Similarly high rates among transgender women have been found by other investigators, in particular those who are African American.[54, 55]

In 1997 the San Francisco Department of Health added male-to-female transgender as a gender category enabling further investigation of HIV outcomes in this community. The incidence among transgender women was found to be 7.8 per 100 person-years (95 percent Confidence Interval-CI, 4.6–12.3) and 18.1 per 100 person-years (95 percent CI, 8.5–34.1) among African Americans.[54] The reasons for the high rates of HIV among transgender women has not been fully investigated, but so far appears to be related to substance use and transactional sex work.

Although many of the studies have been limited by small size and nonprobability sampling, all suggest that an HIV epidemic is currently under way in the male-to-female transgender community, which requires urgent attention by public health officials. Unfortunately, few HIV prevention programs in existence target this community.

Many, but not all, people of transgender experience use cross-gender hormone therapy or surgery to acquire the secondary sexual characteristics of the preferred gender. For transgender women, medical transition includes estrogen therapy and androgen blockers. Surgical treatment may include breast implants, facial feminization surgery, and genital surgery, such as orchiectomy, vaginoplasty, and labiaplasty. Transgender men may initiate testosterone therapy. Surgeries may involve chest masculinization (mastectomy) as well as genital surgery such as phalloplasty or metoidioplasty. Most public and private health insurance plans do not cover the costs of medical or surgical transition and many transgender individuals opt not to undergo expensive surgical interventions.

Transgender women may decide to use more easily available and less costly interventions such as subcutaneous injection of silicone (often industrial grade), mineral oil, or petroleum jelly into the face, chest, hips, thighs, and buttocks to achieve a more feminine appearance. The silicone injections are often provided by nonlicensed practitioners, using nonsterile technique during "pumping" parties, where many people often share available materials. Adverse effects of subcutaneous silicone injection include local reactions, skin infections, fibrosis, and deformity. Serious outcomes include respiratory failure and death due to pulmonary embolism. Complications also may include infections related to blood-borne pathogens such as hepatitis B, hepatitis C, and HIV.

Few studies have been undertaken to study the long-term effects of cross-gender hormone therapy. A study conducted in the Netherlands investigating the outcomes of hormonal therapy in 2,236 male-to-female and 876 female-to-male transsexuals did not reveal an increase in cardiovascular or cancer-related mortality.[56] Further

studies, especially prospective studies, need to be conducted to further assess health issues related to medical transition. The USPSTF does not include specific guidelines for persons of transgender experience. Most clinicians who treat transgender people advise that preventive screenings continue depending on a person's hormonal and surgical status. A comprehensive summary of best practice guidelines is available from the *Fenway Guide to Lesbian, Gay, Bisexual and Transgender Health*.[57]

GAY MEN

Gay men's health has for many years predominantly focused on HIV prevention and treatment. Despite being over two decades into the AIDS epidemic, it is clear that MSM, especially MSM of color, are still disproportionately affected by the disease. In 2008 the CDC released new data on HIV incidence using a new method of calculation.[58] Subpopulation analyses revealed that in 2006, of new HIV infections in men, the majority occurred in MSM of whom 46 percent were white, 35 percent were black, and 19 percent were Hispanic. African American youth are especially vulnerable with 5,520 new HIV infections in the 13 to 29 age group, placing them at 1.6 times the rate of white MSM.[58, 59]

Millet at al. investigated the racial disparity in HIV rates among African American MSM by conducting a meta-analysis of published studies. This revealed that African American MSM reported lower risk behaviors than white MSM, such as less substance use and fewer sexual partners. There was no increase in unprotected sex compared with whites, but the rates of STD among African Americans were higher.[60] The greater risk of HIV acquisition among African Americans appears, in part, due to the baseline higher prevalence in the community.

Other factors that may contribute are socioeconomic and insurance factors that are both related to the proportion of HIV-infected individuals who receive appropriate treatment for their condition, which can affect the reproductive rate of HIV, especially in closed communities. Currently, public health efforts to reduce HIV focus mainly on individual-level risk behavior reduction. The dynamics of HIV in the black MSM community are clearly more complex and will require a community-based approach that takes into account social networks as well as improved access to treatment of STIs and HIV. The CDC and the Center on AIDS and Community Health (COACH) at the Academy for Educational Development (AED) have developed a series of evidence-based HIV prevention strategies, called the Diffusion of Effective Behavioral Interventions (DEBIs) to be introduced both an individual and community level. Several of these focus on the MSM community, including MSM of color, such as "Mpowerment," "Many Men, Many Voices," and "d-up: Defend Yourself!" These interventions have been shown to be effective in reducing risk behaviors among MSM.[61, 62]

There are several STIs besides HIV that gay men may be at increased risk for, depending on individual sexual behavior. Anal sex has been a part of human culture for as long as sex in general, but the public health attention to the specific health risks and consequences engendered by this common practice have been scant. It has been well-described that some STIs, such as gonorrhea, chlamydia, lymphogranuloma venereum, and syphilis, may occur as asymptomatic or with

atypical infections, thus making it more difficult for clinicians to recognize, treat, and control spread of these infections.[63, 64] The prevalence rates of syphilis, gonorrhea, chlamydia, hepatitis B, and hepatitis A are all higher in MSM.[65, 66, 67, 68, 69, 70] Oral-anal sexual contact among MSM has been linked to outbreaks of giardia and other intestinal parasites.[71, 72]

Focused vaccination campaigns in gay male communities have been effective at reducing the prevalence of hepatitis B.[73] In the past few years, there have been well-documented outbreaks of sexual transmission of hepatitis C among MSM, especially MSM infected with HIV.[74, 75] Prior to this era, hepatitis C was thought to be transmitted primarily through exposure to blood (sharing needles, transfusions, and so on). In these clusters of sexual transmission, risk of hepatitis C acquisition was increased by rough sex, sex with multiple partners, specific practices like fisting, and coinfection with other STIs.[76]

Rates of substance use and abuse, including alcohol, tobacco, and recreational drugs, are high in gay men.[77] Although tobacco and alcohol are the two most widely used substances, increased attention has been given to crystal methamphetamine over the past few years because of its potential role in facilitating the spread of HIV, syphilis, hepatitis C, and other STIs.[78] Crystal methamphetamine, colloquially known as "crystal," "meth," "Tina," "speed," or "ice," is a powerful stimulant with strong addictive potential that can be smoked, snorted (intranasal), injected, or administered transrectally (as a suppository known as a "booty bump").

Anal cancer is now recognized as a disease disproportionately affecting MSM, particularly MSM who are infected with HIV.[79, 80] Anal cancer is biologically similar to cervical cancer. Both cancers are primarily caused by the sexual transmission of HPV and both are preceded by cellular changes that can be detected.[81]

In the past decade, anal cancer screening and prevention modalities have been developed, based on the time-proven approaches to cervical cancer screening. Although data are not available on the impact of these interventions on the prevalence of anal cancer, evidence suggests that these interventions can find and remove anal cancer precursors.[79] Some organizations that publish guidelines for the care of people infected with HIV, most notably New York State's AIDS Institute, have started to incorporate anal cancer screening into their recommendations.

A recombinant HPV vaccine (GARDASIL®) for use in women has been available since 2006. In late 2009, the Food and Drug Administration approved the use of the vaccine for men and boys, though the CDC declined to recommend its routine use because of the high cost-benefit analysis. Data on the vaccine's efficacy in preventing anal cancer are not yet available.

SUMMARY

LGBT individuals have specific health needs that require targeted and culturally appropriate interventions, yet they encounter multiple barriers when attempting to access health care services. Eliminating health disparities within this community will only succeed with a multilevel approach that includes input from community members, health care workers, and policy makers. First and foremost there needs to be improved identification of the determinants of health through greater inclusion in population-based research. Behavioral health interventions, such as the DEBIs, are

best-practice examples of culturally appropriate and sensitive programs; however, public health interventions still need to be developed that move beyond sexual health topics and address the other critical issues faced by the LGBT community.

REFERENCES

1. The Gay and Lesbian Medical Association. Healthy People 2010 companion document for LGBT Health. http://www.lgbthealth.net/downloads/hp2010doc.pdf. Published 2001. Accessed June 1, 2009.
2. Solarz AL, Institute of Medicine, eds. *Lesbian Health: Current Assessment and Directions for the Future*. 1st ed. Washington, DC: National Academies Press; 1999.
3. Dean L, Meyer I, Robinson K, Sell R, et al. Lesbian, gay, bisexual, and transgender health: findings and concerns. *J Gay Lesbian Med Assoc*. 2000;4(3):101–151.
4. Laumann E, Gagnon J, Michael R, Michaels S. *The Social Organization of Sex: Sexual Practices in the United States*. Chicago, IL: University of Chicago Press; 1994.
5. Pathela P, Hajat A, Schillinger J, Blank S, Sell R, Mostashari F. Discordance between sexual behavior and self-reported sexual identity: a population-based survey of New York City. *Men Ann Intern Med*. 2006;145:416–425.
6. Mosher W, Chandra A, Jones J. Sexual behavior and selected health measures: men and women 15–44 years of age, United States, 2002. *Vital Health Stat*. 2005;362:21–26. http://www.cdc.gov/nchs/data/ad/ad362.pdf. Accessed June 10, 2009.
7. Goodenow C, Szalacha L, Rubin L, Westheimer K. Dimensions of sexual orientation and HIV-related risk among adolescent females: evidence from a statewide survey. *Am J Pub Health*. 2008;98(6):1051–1058.
8. Saewyc EM, Bearinger LH, Blum RW, Resnick MD. Sexual intercourse, abuse and pregnancy among adolescent women: does sexual orientation make a difference? *Fam Plann Perspect*. 1999;31(3):127–131.
9. Simmons T, O'Connell M. *Married-Couple and Unmarried-Partner Households: 2000*. Washington, DC: US Census Bureau; 2003:14. http://www.census.gov/prod/2003pubs/censr-5.pdf. Census 2000 Special Reports, CENSR-5. Accessed June 2, 2009.
10. American Medical Association, Council on Scientific Affairs. Health care needs of gay men and lesbians in the United States. *JAMA*. 1996;275:1354–1359.
11. Smith EM, Johnson SR, Guenther SM. Health care attitudes and experiences during gynecologic care among lesbians and bisexuals. *Am J Public Health*. 1985;75:1085–1087.
12. Cochran SD, Mays VM. Disclosure of sexual preference to physicians by black lesbian and bisexual women. *West J Med*. 1988;149:616–618.
13. O'Hanlan KA. Lesbian health and homophobia: perspectives for the treating obstetrician/gynecologist. *Curr Probl Obstet Gynecol Fertil*. 1995;18:99–133.
14. Eliason MJ. Correlates of prejudice in nursing students. *J Nurs Educ*. 1998; 37(1): 27–29.
15. White J, Dull V. Room for improvement: communication between lesbians and primary care providers. *J Lesbian Stud*. 1998;2:95–110.
16. O'Hanlan KA. Lesbian health and homophobia: perspectives for the treating obstetrician/gynecologist. *Curr Probl Obstet Gynecol Fertil*. 1995;18:93–136.
17. Wallick MM, Cambre KM, Townsend MH. How the topic of homosexuality is taught at US medical schools. *Acad Med*. 1992;67:601–603.

18. Harris Interactive. Nearly one in four gay and lesbian adults lack health insurance. http://www.harrisinteractive.com/NEWS/allnewsbydate.asp?NewsID=1307. Accessed July 1, 2009.
19. Meyer I. Minority stress and mental health in gay men. *J Health Soc Behav*. 1995; 36(1):38–56.
20. Ryan C, Futterman D. *Lesbian and Gay Youth: Care and Counseling*. 1st ed. New York: Columbia University Press; 1998.
21. Cochran SD. Emerging issues in research on lesbians' and gay men's mental health: does sexual orientation really matter? *Am Psychol*. 2001;56:931–947.
22. Cochran SD, Keenan C, Schober C, Mays VM. Estimates of alcohol use and clinical treatment needs among homosexually active men and women in the US population. *J Consult Clin Psychol*. 2000;68:1062–1071.
23. Cochran SD, Mays VM. Lifetime prevalence of suicidal symptoms and affective disorders among men reporting same-sex sexual partners: results from the NHANES III. *Am J Public Health*. 2000;90:573–578.
24. Cochran SD, Mays VM. Relation between psychiatric syndromes and behaviorally defined sexual orientation in a sample of the US population. *Am J Epidemiol*. 2000;151:516–523.
25. Gilman SE, Cochran SD, Mays VM, Hughes M, Ostrow D, Kessler RC. Prevalences of DSM–III–R disorders among individuals reporting same-gender sexual partners in the National Comorbidity Survey. *Am J Public Health*. 2001;91:933–939.
26. Cochran S, Sullivan J, Mays V. Prevalence of mental disorders, psychological distress, and mental health services use among lesbian, gay, and bisexual adults in the United States. *J Consult Clin Psychol*. 2003;71(1):53–61.
27. Meyer I, Dietrich J, Schwartz S. Lifetime prevalence of mental disorders and suicide attempts in diverse lesbian, gay and bisexual population. *Am J Public Health*. 2008;98(6):1004–1006.
28. Safren S, Heimberg R. Depression, hopelessness, suicidality, and related factors in sexual minority and heterosexual adolescents. *J Consult Clin Psychol*. 1999;67:859–866.
29. Smoking-attributable mortality, years of potential life lost, and productivity losses: United States, 2000–2004. *MMWR*. 2008;57(45):1226–1228.
30. State-specific prevalence and trends in adult cigarette smoking: United States, 1998–2007. *JAMA*. 2009;302(3):250–252.
31. Ryan H, Wortley P, Easton A, Pederson L, Greenword G. Smoking among lesbians, gays, and bisexuals: a review of the literature. *Am J Prev Med*. 2001;21:142–149.
32. Gruskin E, Greenwood G, Matevia M, Pollack L, Bye L. Disparities in smoking between the lesbian, gay, and bisexual population and the general population in California. *Am J Public Health*. 2007;97(8):1496–1502.
33. Cohan DL, Lutnick A, Davidson P, et al. Sex worker health: San Francisco style. *Sex Transm Inf*. 2006;82(5):418–422. http://sti.bmjjournals.com/cgi/content/abstract/sti.2006.020628v1. doi:10.1136/sti.2006.020628.
34. Sanchez NF, Sanchez JP, Danoff A. Health care utilization, barriers to care, and hormone usage among male-to-female transgender persons in New York City. *Am J Public Health*. 2009;99(4):713–719. Epub January 15, 2009.
35. Smith E, Offen N, Malone R. What makes an ad a cigarette ad? commercial tobacco imagery in the lesbian, gay, and bisexual press. *J Epidemiol Community Health*. 2005;59:1086–1091.
36. Smith E, Thompson K, Offen N, Malone R. If you know you exist, it's just marketing poison: meanings of tobacco industry targeting in the lesbian, gay, bisexual, and transgender community. *Am J Pub Health*. 2008;98:996–1003.

37. Lemp G, Jones M, Kellogg TA, et al. HIV seroprevalence and risk behaviors among lesbians and bisexual women in San Francisco and Berkeley, California. *Am J Public Health*. 1995;85:1549–1552.
38. Diamant AL, Schuster M, McGuigan K, Lever J. Lesbians' sexual history with men: implications for taking a sexual history. *Arch Intern Med*. 1999;159:2730–2736.
39. Marrazzo JM. Elusive aetiology of bacterial vaginosis: do lesbians have a clue? *Sex Transm Infect*. 2007;83:424–425.
40. Evans AL, Scally AJ, Wellard SJ, Wilson JD. Prevalence of bacterial vaginosis in lesbians and heterosexual women in a community setting. *Sex Transm Infect*. 2007;83:470–475.
41. Diamant A, Wold C, Spritzer K, Gelberg L. Health behaviors, health status, and access to and use of health care: a population-based study of lesbian, bisexual, and heterosexual women. *Arch Fam Med*. 2000;9(10):1043–1051.
42. Kerker B, Mostashari F, Thorpe L. Health care access and utilization among women who have sex with women: sexual behavior and identity. *J Urban Health*. 2006;83(5):970–979.
43. Cochran SD, Mays VM, Bowen D, et al. Cancer-related risk indicators and preventive screening behaviors among lesbians and bisexual women. *Am J Public Health*. 2001;91:591–597.
44. Health Information and Research Division. Statistics Canada Catalogue no. 82-003-X. *Health Reports*. 2009;19(1):53–64. http://www.statcan.gc.ca/pub/82-003-x/82-003-x2005004-eng.pdf. Accessed July 15, 2009.
45. Matthews AK, Brandenburg DL, Johnson TP, Hughes TL. Correlates of underutilization of gynecological cancer screening among lesbian and heterosexual women. *Prev Med*. 2004;38(1):105–113.
46. State-specific prevalence of obesity among adults: United States, 2005. *MMWR*. 2006;55(36):985–988.
47. Boehmer U, Bowen D, Bauer G. Overweight and obesity in sexual minority women: evidence from population based data. *Am J Public Health*. 2007;97:1134–1140.
48. Aaron D, Markovic N, Danielson M. Cancer related risk indicators and preventive screening behaviors among lesbians and bisexual women. *Am J Public Health*. 2001;91:591–597.
49. Van Kesteren PJ, Gooren LJG, Megens JA. An epidemiological and demographic study of transsexuals in the Netherlands. *Arch Sex Behav*. 1996;25(6):589–600.
50. National Coalition of Anti-Violence Programs. 2008 Report on anti-LGBT violence: http://www.ncavp.org/common/document_files/Reports/2008%20HV%20Report%20smaller%20file.pdf. Accessed July 2, 2009.
51. Risser J, Shelton A, McCurdy S, et al. Sex, drugs, violence, and HIV status among male-to-female transgender persons in Houston, Texas. *Int J Transgenderism*. 2005;8:67–74.
52. Garofolo R, Deleon J, Osmar E, et al. Overlooked, misunderstood and at risk: exploring the lives and HIV risk of ethnic minority male-to-female transgender youth. *J Adolesc Health*. 2006;38:230–236.
53. Clements-Nolle K, Marx R, Guzman R, Katz M. HIV prevalence, risk behaviors, health care use, and mental health status of transgender persons: implications for public health intervention. *Am J Public Health*. 2001;91(6):915–921.
54. Nemoto T, Operario D, Keatley J, Han L, Soma T. HIV risk behaviors among male-to-female transgender persons of color in San Francisco. *Am J Public Health*. 2004;94(7):1193–1199.

55. Kellogg T, Clements-Nolle K, Dilley J, Katz MH, Mcfarland W. Incidence of Human Immunodeficiency Virus among male-to-female transgendered persons in San Francisco. *J Acquir Immune Defic Syndr*. 2001;28(4):380–384.
56. Gooren LJ, Giltay EJ, Bunck MC. Long-term treatment of transsexuals with cross-sex hormones: extensive personal experience. *J Clin Endocrinol Metab*. 2008;93: 19–25.
57. Makadon H, Mayer K, Potter J, Goldhammer H. *The Fenway Guide to Lesbian, Gay, Bisexual, and Transgender Health*. Philadelphia, PA: American College of Physicians; 2007.
58. Hall HI, Song R, Rhodes P, et al. HIV Incidence Surveillance Group estimation of HIV incidence in the United States. *JAMA*. 2008;300(5):520–529.
59. Subpopulation estimates from the HIV Incidence Surveillance System: United States, 2006. *MMWR*. 2008;57(36):985–989.
60. Millett GA, Flores SA, Peterson JL, Bakeman R. Explaining disparities in HIV infection among black and white men who have sex with men: a meta-analysis of HIV risk behaviors. *AIDS*. 2007;21(15):2083–2091.
61. Wilton L, Herbst JH, Coury-Doniger P, et al. Efficacy of an HIV/STI prevention intervention for black men who have sex with men: findings from the Many Men, Many Voices (3MV) Project. *AIDS Behav*. 2009;13(3):532–544.
62. Jones KT, Gray P, Whiteside YO, et al. Evaluation of an HIV prevention intervention adapted for black men who have sex with men. *Am J Public Health*. 2008; 98(6):1043–1050.
63. Kent CK, Chaw JK, Wong W, et al. Prevalence of rectal, urethral, and pharyngeal chlamydia and gonorrhea detected in 2 clinical settings among men who have sex with men: San Francisco, California, 2003. *Clin Infect Dis*. 2005;41(1):67–74. Epub May 26, 2005.
64. Gunn RA, O'Brien CJ, Lee MA, Gilchick RA. Gonorrhea screening among men who have sex with men: value of multiple anatomic site testing, San Diego, California, 1997–2003. *Sex Transm Dis*. 2008;35(10):845–848.
65. Centers for Disease Control and Prevention. Primary and secondary syphilis: United States, 2003–2004. *MMWR*. 2006;55:269–273.
66. Clinic-based testing for rectal and pharyngeal neisseria gonorrhoeae and chlamydia trachomatis infections by community-based organizations: five cities, United States, 2007. *MMWR*. 2009;58(26):716–719.
67. Coutinho RA, Schut BJT, Albrect-Van NN, Reerink-Brongers EE, Stoutjesdijk L. Hepatitis B among homosexual men in the Netherlands. *Sex Transm Dis*. 1981; 8:333–335.
68. Dietzman DE, Harnisch JP, Ray CG, Alexander ER, Holmes KK. Hepatitis B surface antigen HBsAg and antibody to HBAg: prevalence in homosexual and heterosexual men. *JAMA*. 1977;238:2625–2626.
69. Szmuness DE, Much MI, Prince AM, et al. On the role of sexual behavior in the spread of hepatitis B infection. *Ann Intern Med*. 1975;83:489–495.
70. Centers for Disease Control and Prevention. Hepatitis A among homosexual men: United States, Canada, and Australia. *MMWR*. 1992;41(9):155,161–164.
71. William DC, Shookhoff HB, Felman YM, DeRamos SW. High rates of enteric protozoal infections in selected homosexual men attending a venereal disease clinic. *Sex Transm Dis*. 1978;5:155–157.
72. Meyers JD, Kuharic HA, Holmes KK. Giardia lamblia infection in homosexual men. *Br J Vener Dis*. 1977;53:54–55.
73. Alter MJ, Hadler SC, Margolis HS, et al. The changing epidemiology of hepatitis B in the United States. *JAMA*. 1990;263:1218–1222.

74. Ghosn J, Larsen C, et al. Evidence for ongoing epidemic sexual transmission of HCV (2006 to 2007) among HIV-1-infected men who have sex with men: France: Abstract #800: 16th Conference on Retroviruses and Opportunistic Infections; February 8–11, 2009; Montreal, Canada.
75. Urbanus AT, van de Laar TJ, Stolte IG, et al. Hepatitis C virus infections among HIV-infected men who have sex with men: an expanding epidemic. *AIDS*. 2009;23(12):F1–7.
76. Danta M, Brown D, Bhagani S, et al. HIV and Acute HCV (HAAC) group: recent epidemic of acute hepatitis C virus in HIV-positive men who have sex with men linked to high-risk sexual behaviors. *AIDS*. 2007;21(8):983–991.
77. Sanchez T, Finlayson T, Drake A, et al. Human Immunodeficiency Virus risk, prevention, and testing behaviors: United States, National HIV Behavioral Surveillance System: men who have sex with men, November 2003–April 2005. *MMWR*. http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5506a1.htm.
78. Molitor F, Truax SR, Ruiz JD, Sun RK. Association of methamphetamine use during sex with risky sexual behaviors and HIV infection among non-injection drug users. *West J Med*. 1998;168:93–97.
79. Palefsky JM. Anal cancer prevention in HIV-positive men and women. *Curr Opin Oncol*. 2009;21(5):433–438.
80. Palefsky J. Human papillomavirus-related disease in people with HIV. *Curr Opin HIV AIDS*. 2009;4(1):52–56.
81. Hoots BE, Palefsky JM, Pimenta JM, Smith JS. Human papillomavirus type distribution in anal cancer and anal intraepithelial lesions. *Int J Cancer*. 2009; 124(10):2375–2383.

This page intentionally left blank

CHAPTER 15

Public Health Issues in Oral Health

María Teresa Canto, DDS, MS, MPH,
and Gustavo D. Cruz, DMD, MPH

> Ignoring oral health problems can lead to needless pain and suffering, complications that can devastate well-being, and financial and social costs that significantly diminish quality of life.
>
> —Donna Shalala, former Health and Human Services Secretary

INTRODUCTION

The significant advances made in the area of oral health care over the past decades have done much to improve the quality of life for millions of individuals. Yet, in many countries, including the United States, profound oral health disparities still exist. Although common dental diseases such as dental caries, also known as tooth cavities or tooth decay, are easily preventable and treatable, there remains a large underserved population whose basic oral health care is not being met. Even with fluoridation and advances in oral hygiene, tooth decay is still the most common chronic disease affecting all families and having a similar economic impact of heart disease, obesity, and diabetes. This chapter focuses on the epidemiology of oral health, the global burden of dental diseases, and social and economic costs associated with the "silent epidemic" of dental and oral diseases that demands an expanded effort to improve oral health for all.

BRIEF HISTORICAL BACKGROUND

Oral hygiene and dental care are not new concepts or practices specific to the modern age. As far back as 7000 BCE, for example, dentistry was being practiced in the Indus Valley (what is now modern-day India and Pakistan).[1] Ancient texts from Egypt, China, and Japan provide evidence that "tooth worm" was the cause of dental caries.[2] Greek and Roman writings, too, clearly show knowledge of oral diseases as well as dental treatments including dental extractions. Indeed, historically, dental extractions were used to treat a variety of illnesses. During the Middle Ages and throughout the 19th century, dental procedures predominantly were performed by

barbers who limited their activities to extracting teeth to alleviate the pain associated with chronic tooth infection. The science of modern dentistry evolved from tooth extraction to tooth preservation during the 17th and 18th centuries. French physician Pierre Fauchard, who is considered to be the father of modern dentistry, advanced the development of dental care significantly by using dental prostheses and dental fillings.[3]

It was only in the early part of the 20th century that the concept of professional dental hygiene care was developed. Whereas in the past tooth loss was quite common, now with advances in preventive oral health many more people can smile confidently and happily. The term "dental hygiene" is attributed to Dr. Alfred Civilion Fones, a major creative force in the dental hygiene movement, and the founder in 1949 of the Fones School of Dental Hygiene at the University of Bridgeport, the first dental hygiene school in the world. Until this time, the concepts of dental hygiene and disease prevention were almost unknown. Today, dentistry encompasses the evaluation, diagnosis, prevention, and treatment of diseases, disorders, and conditions of the soft and hard tissues of the jaw (mandible); the oral cavity (teeth and gums, gingiva); and the maxillofacial area and the adjacent and associated structures (mucosal tissue, muscles, bones, joints and salivary glands).

Post–World War II saw tremendous advances in the health and well-being of Americans, including advances in oral health. In 1948, the National Institute of Dental Research (now the National Institute of Dental and Craniofacial Research) was established as part of the National Institutes of Health. The institute's focus on dental caries and the effectiveness of fluoride in preventing dental caries ushered in a new era of health promotion and disease prevention. The empirical documentation that people who lived in communities with naturally fluoridated drinking water had far fewer dental caries than those without fluoride in their water supply showed the way to fluoridation of drinking water, 1 of the 10 great public health achievements in the 20th century.[4]

COMPONENTS OF ORAL HEALTH

Oral health means much more than just healthy teeth; it includes healthy gingiva (gums) and periodontum, mucosal tissue, craniofacial muscle, bones, joints, and salivary glands. Good oral health should be considered inseparable from general health. Good oral health is being able to eat, chew, talk, and smile without pain, discomfort, and embarrassment. It is affected by such factors as the biological makeup of individuals, behavioral practices, environment, and the organization of the health care system. Behavioral practices are a consequence of beliefs and attitudes that affect lifestyle choices—for example, food choices and use of products such as tobacco. The environment refers to the communities and neighborhoods where people live and it may reflect factors such as levels of income, education, availability of care, and access to fluoridated water. The organization of the health care system includes the role of providers, and use and delivery of services. A strong relationship exists between poor oral health and factors such as low education, low literacy,[5] income, and lack of health insurance coverage. In addition, poor oral health may be more prevalent in certain racial and ethnic groups, and may be related with acculturation and country of origin.[6]

Dental Caries

Dental caries is a common chronic disease that causes pain and disability across all ages. Caries is a bacterial disease that manifest as the localized destruction of the hard tooth structure (enamel, dentin, or cementum). These tissues progressively break down, producing dental cavities (holes in the teeth). Two groups of bacteria are responsible for initiating caries, Streptococcus mutans and Lactobacilli. If left untreated, individuals run the risk of pain, tooth loss, and infection. Caries can occur in the crown or root surfaces of a tooth.

Severe tooth decay in primary (baby) teeth early in life is known as Early Childhood Caries (ECC). ECC subsumes other descriptive terms of dental caries in young children, including "baby bottle tooth decay" and "nursing caries." ECC is the presence of one or more decayed, missing or filled tooth surfaces in any primary tooth in a child aged 71 months or younger.[1] The decayed surfaces appear as smooth, dull, white, or brown spots on the primary upper teeth. ECC may occur in children who are constantly drinking from a bottle with juice, milk, or formula during the day or overnight. These sugary liquids pool around the upper front teeth mixing with bacteria and causing a rapid destruction of the tooth structure. Other risk factors for ECC may include defects in the tooth enamel, chronic illness, altered salivary flow, and mouth breathing. At this time, the exact cause of ECC is unknown, but factors that have been implicated include large family size, extended use of the baby bottle with sugary liquids, nutritional status of the mother and infant, and the transfer of infectious organisms from caregiver to infant.[7, 8]

Data from the National Health and Nutrition Examination Survey (NHANES) III provided ECC estimates of 1 to 1.7 percent for children 12 to 23 months old. From a national perspective ECC appears to be rare among these children.[9] Children of low-income families suffer twice as many instances of dental caries as their more affluent peers, however, and their disease is more likely to be left untreated. Children whose parents or caregivers have less than a high school education or whose parents and caregivers are Hispanic, Native Americans, or Alaska Natives appear to be at markedly increased risk for developing ECC.[10] The prevalence of ECC in Pacific Islander children living outside of the mainland United States far exceeds those of U.S. mainland children.[11]

The Surgeon General's report on *Oral Health in America* [12] stated that tooth decay is the most prevalent childhood chronic disease and is five times more common than asthma. Data from the most recent NHANES show that the prevalence of dental caries experience increases with age. Dental caries experience is described by the number of decayed (untreated decay), missing, and filled teeth or surfaces. NHANES participants showed that among children ages 2 to 11 years, 41 percent had caries experience in their primary teeth; among adolescents ages 6 to 19 years, 42 percent had caries experience in their permanent teeth; and among dentate adults age 20 and over, approximately 91 percent had caries experience.[13] According to NHANES, approximately 21 percent of children 2 to 11 years old had untreated tooth decay in their primary teeth and

1. http://www.aapd.org/media/Policies_Guidelines/D_ECC.pdf, revised 2008, accessed August 5, 2009.

16 percent of children and adolescents 12 to 15 years old had untreated decay in their permanent teeth.[13] Furthermore, the incidence and prevalence of dental caries is more common among those living below the poverty level. Dental decay and untreated dental caries are both more common among those who are economically less well off compared with those who are above the poverty level. In addition, caries experience varies by race and ethnicity. Nonwhite children had lower prevalence of untreated tooth decay (18 percent) in their primary teeth when compared with non-Hispanic black (27 percent) and Mexican American children (32 percent). A similar pattern is observed for children and adolescents in their permanent dentition. Mexican American children and adolescents have the highest prevalence of untreated tooth decay.

Approximately 91 percent of adults older than 20 years of age had caries experience and more than 23 percent had untreated tooth decay. Prevalence of untreated tooth decay inversely correlates with higher level of education; prevalence was close to 41 percent among those with less than a high school education, 30 percent among high school graduates, and 14 percent among those with more than a high school education. Adults who were current smokers had a higher prevalence of untreated tooth decay (35 percent) than those who never smoked (19 percent) and former smokers (18 percent). Overall, the prevalence of adults with tooth decay decreased from approximately 28 percent in 1988–1994 to 23 percent in 1999–2002.[13] This reduction may be attributed to exposure to preventive measures for dental caries such as fluoride and dental sealants.

Caries may occur in the root of a tooth. Root caries are found where loss of periodontal attachment leaves the root surface exposed to the oral environment and then the accumulation of bacteria and plaque occur around the exposed roots. Approximately 18 percent of dentate adults, 20 years of age and older, have experienced root caries. Prevalence of root caries increase with age, close to 9 percent among persons between the ages of 20 and 39 years, 18 percent among those 40 and 59 years, and 32 percent among those 60 years of age and older. It is a particular problem among older people of lower socioeconomic status (SES), those who have lost some teeth, those who do not maintain good oral hygiene, and those who do not use dental services regularly.[13]

Gingivitis and Periodontal Diseases

Gingivitis and periodontal diseases are the second most common oral disease worldwide, after dental decay. Gingivitis or gingival inflammation is a reversible condition caused by bacteria accumulation in the dental plaque. Gingivitis is measured by gingival bleeding. Most children and adolescents worldwide have signs of gingivitis, which include gingival inflammation and bleeding. In the United States, studies indicate that gingivitis is more evident among Mexican Americans (64 percent) than among non-Hispanic blacks (56 percent) and non-Hispanic whites (49 percent).[14] The reasons for the difference are unknown, but it may relate to oral hygiene habits.

Gingival inflammation may occur as a side effect of medications such as phenytoin, nifedipine, and cyclosporine, which predispose one to gingival overgrowth in response to plaque. Also changes in hormone levels (for example, pregnancy and use of oral contraceptives) may increase severity of plaque-induced gingival inflammation.

Periodontitis involves progressive loss of the alveolar bone around the teeth, and if left untreated, can lead to the loosening and subsequent loss of teeth. It is a chronic bacterial infection initiated by the presence of Gram-negative bacteria, which accumulates in the gingival crevice region. Periodontal destruction is considered as a result of the response of a susceptible individual to the bacterial challenge. It is estimated that 30 to 50 percent of the adult population in the United States has some form of periodontal disease. Surveys of periodontal conditions usually show that men have poorer periodontal health than women.[2] Older seniors, black and Hispanic seniors, current smokers, and those with lower incomes and less education are more likely to have periodontal disease.

Several conditions may increase the prevalence, incidence, or severity of periodontitis. The effects of a significant number of systemic diseases upon periodontitis are unclear and often it is difficult to causally link such diseases to it. Immunosuppression of the normal inflammatory and immune mechanism (such as HIV infection) may predispose the individual to periodontal destruction. Convincing evidence suggests that smoking is the strongest risk factor for periodontal disease. Additionally, periodontitis severity and prevalence are increased in diabetics and are worse in poorly controlled diabetics.[13]

Emerging research data show the potential link between periodontal disease and systemic health. Findings from secondary data analysis and some epidemiological studies have suggested an association between chronic oral infections and heart and lung diseases, and stroke.[13] Although the role of periodontal disease and preterm birth has been investigated extensively, the results have been inconclusive; a number of studies using a prospective cohort study design without treatment found an association, whereas other studies have not. For example, the *Obstetrics and Periodontal Therapy Study* randomly selected women to receive periodontal treatment before 21 weeks of gestation or after delivery. In this study, periodontal disease progression was not associated with an increased risk for delivering a preterm or a low–birth weight infant.[15] However, pregnancy is an opportune time to educate women about preventing dental caries in young children and their own prevention of gingivitis and periodontal disease. Control of oral diseases will improve the woman's quality of life and that of her children. In 2006, the New York State Department of Health developed guidelines to help health care professionals to educate women about oral health. For example,

all health care professionals should advise women that the following actions will improve their health:

- Brush teeth twice daily with a fluoride toothpaste and floss daily.
- Limit foods containing sugar to mealtimes only.
- Choose water or low-fat milk as a beverage. Avoid carbonated beverages during pregnancy.
- Choose fruit rather than juice to meet the recommended daily fruit intake.
- Obtain necessary dental treatment before delivery.[3]

---

2. http://www.nidcr.nih.gov/DataStatistics/FindDataByTopic/GumDisease/PeriodontaldiseaseAdults20to64, accessed July 15, 2009.
3. http://www.health.state.ny.us/publications/0824.pdf, accessed July 29, 2009.

Tooth Retention and Edentulism

Tooth retention and edentulism (tooth loss), the major cause of which are dental caries and periodontal disease, have declined in the United States. Increased adoption of preventive measures such as improved oral hygiene, dental sealants, community water fluoridation, use of fluoride toothpaste, and mouth-rinses may explain the observed trend. The prevalence of complete tooth loss is similar in men (close to 7 percent) and women (close to 8 percent).[13] Among SES groups, those living below the poverty level tend to have more tooth loss compared with those in higher SES groups. Overall, a higher number of individuals living below the poverty level are edentulous than are those living above. Despite improvements in tooth loss and edentulism, disparities remain. Older adults and smokers have higher prevalence of tooth loss than younger adults and nonsmokers. Racial and ethnic differences in tooth loss exist, with non-Hispanic black adults retaining fewer teeth than nonwhite and Mexican American adults. Mexican Americans are the least likely to lose all of their teeth.[13]

Tooth loss and edentulism are due to differences in healthy behaviors, attitudes toward oral health and care, and access to and use of dental services and types of treatment received. In addition, findings reported in the last NHANES are influenced by the oral health status of the group 60 years of age and older, which may have experienced higher rates of dental caries and tooth extraction and might have had a different expectation about retaining their teeth earlier in life that those individual under 60 years of age.[13]

Oral and Pharyngeal Cancer

Oral and pharyngeal cancer are cancers that occur in the mouth, in the throat, and on the lips. Oral cancers are the sixth most common cancer in the world, with more than 500,000 new diagnoses each year. Unfortunately, most diagnoses are made at an advanced stage of the disease and prognosis for patients is poor. Sharp increases in the incidence of oral and pharyngeal cancers have been reported for several countries and regions such as Denmark (3.8 per 100,000 world population), France (7.6 per 100,000), Germany (4.6 per 100,000), Scotland (4.5 per 100,000), and Central and Eastern Europe ($\geq$6.9 per 100,000), and to a lesser extent Australia (3.6 per 100,000), Japan (2.6 per 100,000), New Zealand (2.1 per 100,000), and the United States (2.9 per 100,000).[16, 17]

In the United States, oral cancers are the eighth most common cancer among males. In 2008, an estimated 35,720 people will be diagnosed with oral cancers representing 3 percent of all cancers; some 7,600 individuals will die from these cancers.[18] For all stages combined, the five-year relative survival rate is 60 percent. The survival rate for these cancers has changed slightly in the past 30 to 40 years, and it is lower than the survival rate for other major cancers such as breast and colorectal. Yet, when oral cancer is diagnosed early, the five-year relative survival rate is 82 percent; if the cancer has spread to lymph nodes, the five-year survival drops to 53 percent and when there is metastases to other organs, the five-year survival rate is 28 percent.[19] The data stress the importance for early detection.

Data from the Surveillance, Epidemiology and End Results (SEER) program show that oral cancers are more prevalent among African Americans when compared

with other ethnic groups. The data from the SEER program cover areas of the United States where Mexican Americans make up the predominant Hispanic/Latino population. However, oral cancer may be higher is some Hispanic/Latino groups. A study conducted in the state of New York showed that oral cancers are higher among Hispanics/Latinos in the State of New York and New York City. In the case of New York State, Puerto Ricans represent a large number of the Hispanic/Latino population. Overall, data show that Hispanics/Latinos living in the United States have higher rates of oral cancers than other ethnic groups.[20] Recent studies have suggested that Hispanic Americans tend to experience a cancer burden similar to that seen in their countries of origin. Interestingly, Puerto Ricans in Puerto Rico experience high rates of oral cancers and those rates are similar to those experienced by other Hispanics in New York State.[21]

The most well-established risk factors for oral cancer include use of tobacco products (cigarettes, cigars, snuff, spit tobacco, and pipes) and use of alcohol.[22] Other risk factors include age, a previous oral cancer lesion, certain oncogenic viruses such as human papillomavirus,[23] unprotected exposure to sunlight (lip cancer), and diet low in fruits and vegetables. Oral squamous cell carcinoma, which arises from the mucosal lining of the oral cavity, accounts for more than 90 percent of oral cancers.[24] It is often preceded by the presence of clinically identifiable premalignant changes of the oral mucosal. These changes often present as white, red, or mixed patches known as erythroplakia, or erythroleukoplakia, respectively. Oral cancers are highly preventable and treatments are less invasive when detected at early stages; the early stages of oral cancer are often not painful; if you are in pain, it may be too late. The "gold standard" for the early detection of oral cancers is an oral examination, a visual and tactile examination that can be performed by a health provider.

DEVELOPMENTAL DISORDERS

Developmental disorders include congenitally missing teeth; congenital disorders involving tooth enamel, pulp, or dentin; and craniofacial birth defects and syndromes. Among such disorders, cleft lip and palate are the most common congenital anomalies and may occur as isolated defects or as part of other syndromes. Both genetic and environmental factors play a role in the etiology of cleft lip and palate. Oral clefts are among the most common birth defects and affect facial appearance throughout life. In the United States, the prevalence rates in the general population are 10.5 per 10,000 live births for cleft lip with or without cleft palate and 6.4 per 10,000 live births for cleft palate alone. The rate of oral clefts for whites is more than three times that for blacks and they are more common among Native Americans.[25] Furthermore, for reasons that are not entirely clear, cleft palate is more frequently seen in females, whereas cleft lip or cleft lip/palate is more common in males. Numerous developmental disorders affect the oral, dental, and craniofacial complex. These include congenitally missing teeth (all or specific tooth types); congenital problems involving tooth enamel, pulp, or dentin; and craniofacial birth defects or syndromes.

A number of environmental factors have been examined as potential risk factors for oral clefts: maternal smoking, antiepileptic drugs, antiemetic agents, vitamin use

during the periconceptual period, maternal metabolic factors, alcohol consumption, and exposure to agricultural chemicals.[26] Maternal smoking is the more studied environmental risk factor for oral clefts and is associated with a slight increase in risk for both cleft palate and cleft lip and palate.[27] A case-control study from Iowa showed a higher risk of isolated cleft lip and palate associated with increased maternal alcohol consumption. Vitamin A and folic acid have been implicated in the etiology of oral clefts.[28] High doses of folate supplementation have a protective effect, and vitamin supplementation also has a protective effect when taken early in pregnancy. Twin studies have demonstrated a major role for genetic factors.

INJURY

Injuries to the head, face, and teeth range in severity from the very mild to the most serious (death). That being said, the data are limited in that only injuries that involve a visit to the emergency room tend to be reported. Based on emergency room reporting, the leading causes of head and face injuries are a result of a fall, an assault, sports injury, and motor vehicle accidents. Males, in general, present with head and face injuries more frequently than females. Rates of injury are higher among the young and among older adults compared with those in middle age.[12] Injuries to the head, face, and teeth are common. Those head and face injuries related to sports may be prevented by the use of helmets, facemasks, and mouthguards.

TEMPOROMANDIBULAR DISORDERS

Temporomandibular muscle and joint disorders (TMJD) are a heterogeneous collection of disorders characterized by orofacial pain or masticatory dysfunction. TMJD can be organized into three broad categories: (1) functional disorders of the musculature of the face, head, neck, shoulders, and upper back, (2) disorders involving the soft tissues of the temporomandibular joint (TMJ), and (3) disorders involving the hard structures of the TMJ. Symptoms of temporomandibular joint and muscle disorders vary, but may include severe pain in the jaw musculature, severe pain or difficulty when opening the mouth and chewing, headaches, and ear pain. Based upon assessments of pain in or around the jaw joint, these disorders are estimated to affect 10 million Americans.[29] A range of risk factors have been implicated with TMJD: joint and muscle trauma, anatomical factors (skeletal and occlusal relationships), pathophysiological factors (bone and connective tissue disorders, hormonal differences, sensitization of peripheral and central nervous system pathways), and psychosocial factors (depression and anxiety), and emotional and perceptual responses to psychological stressors. It remains unclear which are the risk factors for developing acute or chronic forms of TMJD because most epidemiological studies have been limited to cross-sectional or case-control designs often relying on convenience samples.

The reported prevalence of TMJD ranges widely, depending on the definition used. In population studies examining for clinical signs such as TMJ sounds, the prevalence rates tend to be quite high, approximately 61 percent. When reported symptoms of TMJD are used as the criterion, the prevalence drops to around 32 percent. The prevalence goes down to 3 to 5 percent when pain and

dysfunction are severe and frequent. TMJD is higher among younger persons, is at least two times more prevalent in women than in men, and women using either supplemental estrogen or oral contraceptives are more likely to seek treatment for this condition.[30] Hormonal differences is one of the risk factors being studied in TMJD.

More than 39 million adults in the U.S. population (or 22 percent of adults over age 18) experienced at least one of six types of oral-facial pain during a six-month period—that is, tooth-related infections, mucosal sores, mucosal irritations, burning sensations, pain in the jaw-joint area, and aching pain across the face or cheek.[31] If untreated, oral-facial pain can affect quality of life. Jaw-joint pain and face-cheek pain appear to be more prevalent among women. Oral-facial pain can greatly reduce quality of life. Women are twice as likely as men to report two specific types of oral-facial pain: jaw/joint pain and face/cheek pain.

SJÖGREN'S SYNDROME

Sjögren's syndrome is an autoimmune disorder that causes: xerostomia (dry mouth), difficulty in swallowing, and xerophthalmia (dry eyes). This condition is estimated to affect approximately 1 to 2 million people in the United States with women being nine times more likely to have Sjögren's syndrome than men.[32] The diagnosis is most often made in women in middle age. Salivary gland hypofunction is part of this syndrome. Saliva is an important factor in the maintenance of oral health. It aids in digestion and helps to dislodge pathogens (viruses, bacteria and yeast) from teeth and mucosal surfaces. Also, saliva is rich in antimicrobial components and buffers that help maintain a normal pH when acidic foods and beverages are consumed. Lack of adequate saliva may cause difficulty in swallowing, chewing, and speaking as well as loss of enjoyment of food. Saliva controls the bacterial and fungal environment of the mouth and reduction in saliva increases the risk of developing dental caries, and other infections in the mouth.

SOCIAL AND ECONOMIC IMPACT OF ORAL DISEASES

Although improvements in oral health have occurred during the past century, certain groups (for example, low-income individuals, poorly educated, migrants, and uninsured individuals) still have great oral health unmet needs and limited resources to take care of these needs. In the United States, children lose 55 million school hours and adults miss 164 million work hours annually due to oral health problems. Low-income children have almost 12 times more days missed form school than do higher-income children.[12] Only about 35 percent of people over age 25 with less than 12 years of education have been to a dentist in the past year, compared with about 71 percent of those with more than 12 years of education. More than 20 percent of people with a family income of less than $10,000 have not been to a dentist for five years or more; close to 6 percent of Mexican Americans have never been to a dentist, compared with 4 percent of whites.[12] In 2004, almost 100 million Americans had no dental insurance, often wrongly perceived as a luxury that can be done without. More employers are either dropping dental coverage or not offering it at all.

Currently, routine oral health care services are not covered under the Medicare system and are not mandated for adults under the Medicaid program or for children under the State Children Health Insurance Program (SCHIP). Only 15.7 percent of adults—and only 2.5 percent of poor adults age 65 and over living in rural areas—have private dental insurance. More than 130 million children and adults in the United States lack dental insurance.[33] The effect of this lack of coverage is compounded by an oral health care delivery system that does not adequately meet the needs of the population. A recent study examining the disparities in dental insurance coverage and dental care among children found that disparities in access to dental care reflected family income, parental education, race and ethnicity, and urban versus rural residence. Even when insurance status was considered, these population groups still were less likely to receive preventive dental care. It is plausible that this situation exists because of the paucity of dentists who accept Medicaid. Also, many low-income families cannot afford the deductible even if they have insurance coverage. Previous research had shown that almost half of dental expenditures are paid out of pocket and only one-third by dental insurance.[34] Hence, addressing the issue of unmet dental needs, irrespective of dental insurance, must include provider availability as well as parental education.

ORAL HEALTH ON A GLOBAL SCALE

The Oral Health Program of the World Health Organization (WHO) has gathered results of epidemiological studies conducted in different countries over the past 20 years through its Global Data Bank.[17] The WHO measures the severity of dental caries by a Decayed, Missing, and Filled Teeth Index (DMFT). In 12-year-old children, dental caries experience is high in the Americas (DMFT=3) and in the European region (DMFT=2.6), whereas it is lower in most African countries (DMFT=1.7). In most developing countries, the level of dental caries was low until recent years, but prevalence rates of dental caries and dental caries experience are increasing. This is largely due to increase consumption of sugars and inadequate exposure to fluorides. Among adults 35 to 44 years old, the prevalence of dental caries is high in developed countries and some countries in Latin America (DMFT=14). Overall, the data confirm a decline in the prevalence of dental caries in children and adolescents in developed countries and an increase in dental caries in developing countries.

Regarding other oral diseases, severe periodontitis is found in 5 to 20 percent of most adult populations worldwide, and oral cancers are more common in developing than developed countries. For example, in south-central Asia, cancer of the oral cavity ranks among the three most common types of cancer mainly because of the high use of tobacco in combination with areca nut.

HIV infection is a major global health problem affecting developing and developed countries, and conditions of the oral cavity are prevalent among those with HIV/AIDS. Among those with HIV, for example, oral candidiasis (commonly called thrush) is quite common. Other oral manifestations include Kaposi's sarcoma and salivary gland disease.[35] If untreated, oral lesions associated with these diseases can and do affect the quality of life of HIV patients.

The current global and regional patterns of oral disease reflect living conditions, lifestyles, implementation, use of preventive measures, and systems of health care

delivery. Around the world, the greatest burden of oral disease is on the disadvantaged and poor population groups. It is critical to strengthen public health programs by including effective measure for prevention of oral diseases (for example, exposure to fluorides in different forms) and by addressing common risk factors with other diseases, making it essential to integrate oral health with national health programs.

EVIDENCE-BASED PREVENTION APPROACHES: DENTAL CARIES

Preventive measures for dental caries include exposure to fluoride, dental sealants, good oral hygiene, and reduction in sugar intake. In the United States, fluoride is added not only to the water supply but also to toothpaste, some processed foods, and some beverages. Adding fluoride to the water supply has been shown to help prevent cavities and it also helps to slow or reverse the progression of oral lesions. Fluoride is incorporated into the enamel before tooth eruption; it inhibits demineralization, enhances remineralization, and inhibits bacterial activity in dental plaque. Convincing evidence suggests that the widespread use of fluoride has been a major factor in the decrease in the prevalence and severity of dental caries in the United States.[13]

The U.S. Preventive Services Task Force and the Canadian Task Force on Periodic Health Examination [36] affirm that there is strong evidence to support the major methods for providing fluoride to prevent dental caries. Evidence from clinical trials conducted from the 1940s through the 1970s has shown fluoride's benefits in reducing dental caries.[37, 38] Regarding the safety of fluoride, evidence shows that when used appropriately fluoride is both safe and effective.[39] In fact, based on the available evidence, the Institute of Medicine in 1997 classified fluoride as a micronutrient and recommended its use in maintaining oral health.[40] The success of community water fluoridation trials in reducing dental caries led to the development of other important fluoride-containing products, such as fluoride-containing gels and pastes for topical use, dietary supplements, and of course in toothpaste.

Fluoride also has been shown to minimize hypomineralization of enamel (called enamel fluorosis), a condition characterized by visually detectable changes in the opacity of tooth enamel.[41] Most enamel fluorosis seen in the 21st century is of the mildest form, which affects neither aesthetics nor dental function. In cases in which a higher than optimal amount of fluoride is ingested, enamel fluorosis can be more severe. For example, cosmetically objectionable enamel fluorosis can occur in young children (younger than age six) when enamel is forming. The condition's severity depends on the amount, duration, and timing of fluoride intake. To control fluoride intake, the Centers for Disease Control and Prevention's recommendations for the use of fluorides in the United States include using just a "pea-size" amount of toothpaste for children age 6 and under, supervising children's tooth brushing to avoid excessive swallowing of toothpaste, and using other fluoride modalities based on the caries risk factor of the individual population.[42]

Salt fluoridation is another option of fluoride exposure for remote locations where no municipal water supplies exist. The history of salt fluoridation spans more than half a century, encompassing an effort in Europe and the Americas. In 1994, the Pan American Health Organization started a program

to evaluate the effects of salt fluoridation in the Americas. To date, Mexico, all of Central America, and a great part of South America (except Argentina, Brazil, Chile, and French Guyana) have or are about to launch salt fluoridation programs. These efforts have demonstrated that salt fluoridation is an effective and practical vehicle to provide appropriate levels of fluoride to the population and, therefore, to prevent dental caries.[43]

Dental sealants are highly effective in preventing dental caries that occur on the surfaces of teeth that have fissures. Dental sealants that remain intact after placement are 100 percent effective.[44] In examining the effectiveness of school-based programs, the Guide to Community Preventive Services documented a 60 percent decrease in tooth decay on the chewing surfaces of posterior teeth up to five years after sealant applications. In 2002, the Task Force on Community Preventive Services recommended school-based or school-linked sealant programs for the prevention and control of dental caries.[45]

The prevalence of sealants increased across all sociodemographic groups. Despite the gains, profound disparities still exist. Non-Hispanic white children and adolescents in families with higher incomes that have lower levels of tooth decay were at least 60 percent more likely to have received a sealant than were other racial or ethnic populations and those from lower SES groups.[13] School-based and school-linked programs in the United States generally target vulnerable populations less likely to receive private dental care. An expansion in the number of these programs may decrease disparities in the prevalence of sealants.

CHALLENGES

Even with many reports on the advances for the prevention and treatment of major oral diseases and conditions, not all individuals experience the same degree of improvement in oral health. The burden of oral diseases and conditions is greatest among the economically disadvantaged, which include a disproportionally large number of racial and ethnic minorities and underserved populations. The major risk factors for oral disease are known and they are common with those for other chronic diseases: diet, smoking, alcohol, and risky behaviors. It makes sense to incorporate oral health preventive activities in different health programs. For example, nutrition programs encouraging a reduction in sugar consumption, a risk factor for dental caries, would have the added benefit of a reduction in dental caries. Also, tobacco control programs should involve oral health professionals in tobacco cessation and prevention, as smoking and chewing tobacco is a well-known risk factor for diseases of the oral cavity.

Health promotion goes beyond education activities; it is "a combination of health education and/or organizational, political and economic interventions designed to facilitate behavioral and environmental adaptation that will improve and protect health."[46] Any oral health promotion programs should include the following: (1) enhancing the public's understanding of the meaning of oral health and the relationship of oral health with general health and other diseases/conditions; (2) raising awareness of the importance of oral health among government policy makers to create effective public policy that will improve oral health (for example, in the United States, Medicare should include dental services and State Medicaid

programs should include oral health for children and adults); (3) training non-oral-and-dental health professionals about oral health and diseases and about their role in ensuring that patients have good oral health; and (4) integrating oral health into primary care services.

The "silent epidemic" of dental and oral disease is a complex problem. It affects the health and quality of life of the individual and society. In addition, suffering from dental diseases or conditions has a psychosocial and economic impact. A toothless smile can be a liability in terms of personal confidence and may negatively affect an individual's ability to be hired. Evidence points to the fact that good oral health is conducive to productivity as well as general health. Because those with the greatest unmet needs are the underserved populations, which may include different racial and ethnic groups, it is important to address the issues of dental insurance and dental education. Concomitantly, there is a need to place emphasis on preventive care to avoid the more serious dental problems in the first place. Dentures, for example, are expensive whether one has insurance coverage or not. Good oral health, initiated in childhood and continued into adulthood, would do much to obviate the need for more expensive dental work as the individual ages. The unmet need for dental care is a serious problem that needs to be acknowledged, if not addressed, as Congress debates health care reform.

REFERENCES

1. Coppa A, et al. Early neolithic tradition of dentistry. *Nature*. 2006;440:755–756.
2. Suddick RP, Harris NO. Historical perspectives of oral biology: a series. *Crit Rev Oral Biol Med*. 1990;1:135–151.
3. Schulein TM. Significant events in the history of operative dentistry. *J Hist Dent*. 2005;53:64–69.
4. Centers for Disease Control and Prevention. Ten great public health achievements: United States, 1900–1999. *MMWR*. 1999;48:241–243.
5. Rudd RE, Horowitz AM. Health and literacy: supporting the oral health research agenda. *J Public Health Dent*. 2005;65:131–132.
6. Cruz GD, Chen Y, Salazar CR, Le Geros RZ. The association of immigration and acculturation attributes with oral health among immigrants in New York City. *Am J Public Health*. May 15, 2009; Epub ahead of print.
7. Horowitz AM. Response to Weinstein: public health issue in early childhood caries. *Community Dent Oral Epidemiol*. 1998;26:84–90.
8. Ismail AI. Prevention of early childhood caries. *Community Dent Oral Epidemiol*. 1998;26:49–61.
9. Kaste LM, Drury TF, Horowitz AM, Beltran E. An evaluation of NHANES III estimates of early childhood caries. *J Public Health Dent*. 1999;59:198–200.
10. Vargas CM, Crall JJ, Schenider DA. Sociodemographic distribution of pediatric dental caries: NHANES III, 1998–1994. *J Am Dent Assoc*. 1998;129:1229–1238.
11. Greer MH, Larson K, Sison S. Comparative analysis of oral health indicators among young children in Hawai'i, the Republic of Palau and Territory of Guam, 1999–2000. *Pac Health Dialog*. 2003;10:6–11.
12. Department of Health and Human Services. *Oral Health in America: A Report of the Surgeon General*. Rockville, MD: US Department of Health and Human Services, National Institute for Dental and Craniofacial Research/National Institutes of Health; 2000.

13. Centers for Disease Control and Prevention. Surveillance for dental caries, dental sealants, tooth retention, edentulism, and enamel fluorosis in the United States, 1998–1994 and 1999–2002. *MMWR CDC Surveill Summ*. 205;54:1–44.
14. Albander JM, Kingman A. Gingival recession, gingival bleeding and dental calculus in adults 30 years of age and older in the United States, 1988–1994. *J Periodontal*. 1999;70:30–43.
15. Michalowicz BS, Hodgers JS, Novak MJ, et. al. Change in periodontitis during pregnancy and the risk of pre-term birth and low birthweight. *J Clin Periodontol*. 2009;36:308–314.
16. Petersen PE, Bourgeois D, Ogawa O, Estupiñan-Day S, Hdiaja C. The global burden of oral diseases and risks to oral health. *Bull World Health Organ.* 2005;83(9):661–669.
17. Cancela M, Voti L, Guerra-Yi M, et al. Oral cavity cancer in developed and in developing countries: population-based incidence. *Head Neck*. 2009; Epub ahead of print.
18. American Cancer Society. *Cancer Facts and Figures 2009*. Atlanta, GA: American Cancer Society; 2009.
19. Ries LAG, Young JL, Keel GE, Eisna MP, Lin YD, Honer M-J, eds. Cancer of the head and neck. In: Piccirillo JF, Costas I, Reichman ME. *SEER Survival Monograph: Cancer Survival among Adults*. Bethesda, MD: National Cancer Institute, SEER Program; 2007: Chapter 2. NIH Pub No.07-6215.
20. Ries LAG, Melbert D, Krapcho M, et al., eds. *SEER Cancer Statistics Review, 1975–2004*. Bethesda, MD: National Cancer Institute; 2007. http://seer.cancer.gov/csr/1975_2004/. Published November 2006.
21. Cruz GD, Salazar C, Morse D. Oral and pharyngeal cancer incidence and mortality among Hispanics, 1996–2202: the need for ethnoregional studies in cancer research. *Am J Public Health*. 2006;96:2194–2200.
22. Hashibe M, Brennan P, Chuang SC, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis of the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiol Biomarkers Prev*. 2009;18:540–541.
23. Smith EM, Ritchie JM, Pawlita M, et al. Human papillomavirus seropositivity and risks of head and neck cancer. *Int J Cancer*. 2007;120:825–832.
24. Canto MT, Devesa SS. Oral cavity and pharyngeal cancer incidence rates in the United States, 1975–1988. *Oral Onc*. 2002;38:610–617.
25. Canfield MA, Honeim MA, Yuskin N, et al. National estimates and race/ethnic specific variation of selected birth defects in the United States, 1999–2001. *Birth Defects Res A Clin Mol Teratol*. 2006;76:747–756.
26. Wyszynski DF, Beaty TH. Review of the role of potential teratogens in the origin of human nonsyndromic oral clefts. *Teratology*. 1996;53:309–317.
27. Chung KC, Kowalski CP, Kim HM, Buchman SR. Maternal cigarette smoking during pregnancy and the risk of having a child with cleft lip/palate. *Plast Reconstr Surg*. 2000;105:485–491.
28. Prescott NJ, Malcolm S. Folate and the face: evaluating the evidence for the influence of folate genes on craniofacial development. *Cleft Palate Craniofac J*. 2002;39:327–331.
29. Lipton JA, Ship JA, Larach-Robinson D. Estimated prevalence and distribution of reported oralfacial pain in the United States. *J Am Dent Assoc*. 1993;124:115–121.
30. Isong U, Gansky SA, Plesh O. Temporomandibular joint and muscle disorder-type pain in US adults: the National Health Interview Survey. *J Orofacial Pain*. 2008;22:317–322.

31. McFarlane TV, Blinkhorn AS, Davies RM, et al. Oro-facial pain in the community, prevalence and associated impact. *Community Dent Oral Epidemiol*. 2002; 30:52–60.
32. Talal N. Sjogren's syndrome: historical overview and clinical spectrum of disease. *Rheum Dis Clin North Am*. 1992;18(3):507–515.
33. Liu, J. et al. Disparities in dental insurance coverage and dental care among US children: the National Survey of Children's Health. *Pediatrics*. 2009;119:512–521.
34. Vargas, CM, Manski RJ. Dental expenditures and sources of payment by race/ethnicity and other sociodemographic characteristics. *J Public Health Dent*. 1999;59:33–38.
35. Ranganathan K, Hemalatha R. Oral lesions in HIV infection in developing countries: an overview. *Adv Dent Res*. 2006;19:63–68.
36. Lewis DW, Ismail AI, The Canadian Task Force on the Periodic Health Examination. Periodic Health Examination 1995 update 2: prevention of dental caries. *CMAJ*. 1995;152:836–846.
37. Horowitz HS. Proper use of fluoride products in fluoridated communities. *Lancet*. 1999;353:1462.
38. Stookey GK. Caries prevention. *J Dent Educ*. 1998;62:803–811.
39. Newbrun E. The fluoridation war: a scientific dispute or a religious argument? *J Public Health Dent*. 1996;56:246–252.
40. Institute of Medicine. Dietary reference intake for calcium, phosphorus, magnesium, vitamin D and fluoride. http://fnic.nal.usda.gov/nal_display/index.php?info_center=4&tax_level=4&tax_subject=256&topic_id=1342&level3_id=5141&level4_id=10587. Published January 1, 1997.
41. Fejerskov O, Manji F, Baelum V. The nature and mechanism of dental fluorosis in man. *J Den Res*. 1991;69:(Spec 692):700.
42. Centers for Disease Control and Prevention. Recommendations for using fluoride to prevent and control dental caries in the United States. *MMWR Recomm Rep*. 2001;50(RR14):1–42.
43. Estupiñan-Day S, ed. *Promoting Oral Health: The Use of Salt Fluoridation to Prevent Dental Caries*. Washington DC: PAHO; 2005. Scientific and Technical Publication No.615.
44. Gooch BF, Truman BI, Griffin SO, et al. A comparison of selected evidence reviews and recommendation on interventions to prevent dental caries, oral and pharyngeal cancer, and ports-related craniofacial injuries. *Am J Prev Med*. 2002; 23:55–80.
45. Task Force on Community Preventive Services. Recommendations on selected interventions to prevent dental caries, oral and pharyngeal cancers, and sports-related craniofacial injuries. *Am J Prev Med*. 2002;23:16–20.
46. Green LW, Kreuter MW. *Health Program Planning: An Educational and Ecological Approach*. 4th ed. New York: McGraw-Hill; 2004.

This page intentionally left blank

CHAPTER 16

Aging and Public Health

Ronald D. Adelman, MD, Emily Finkelstein, MD, Sonal S. Mehta, MD, and Michele G. Greene, DrPH

INTRODUCTION

Individuals over 65 years represent almost 13 percent of the United States population,[1] and it is estimated that this number will increase dramatically in the next decades. By some estimates, it is predicted that the over-65 population will increase to 20 percent of the population by 2030. Moreover, the fastest growing subgroup of individuals over age 65 is the population 85 and over [2] and individuals reaching age 65 have an average life expectancy of another 19 years.[1] In addition, there has been a large increase in the centenarian population in recent years. Currently in the United States, there are 100,000 people who are at least 100 years old. America is an aging society not only due to the increasing longevity of its citizens but also because our population is having fewer children. The younger American population is much too small to counterbalance the growing number of aging individuals, a phenomenon that is also seen in other developed countries.[2] Described by some as the graying of the population, or the aging tsunami, aging baby boomers will significantly affect the nature of life as we know it in America.

WHO ARE THE ELDERLY?

What do we know about the over 65 population now? A report prepared by the U.S. Department of Health and Human Services' Administration on Aging, *A Profile of Older Americans: 2008,* provides a snapshot of the elderly in America. Fifty-eight percent of the older population is female, and, not surprisingly, the female-to-male sex ratio increases with age. While the majority of older men are married (73 percent), only 42 percent of older women are married. Furthermore, almost half of women 75 years and older live alone, and older women are more likely to be poorer or less economically well off compared with older men. In 2007, almost 10 percent of the older population lived below the poverty level. Although only 16 percent of individuals over age 65 worked in 2007, more than a third of these individuals worked full time.[1] Many retired older people engage in volunteer work, including

participating in family responsibilities such as looking after grandchildren and their aging parents. Such voluntary, unpaid labor is not calculated in gross domestic product (GDP) data. If these individuals continue with full-time work, then their unpaid caregiving or other volunteer work would need to be paid for or, unfortunately, might not get done at all.[3]

Between 2000 and 2020, the proportion of the elderly who are of an ethnic or racial minority is expected to increase from 16 percent to almost 24 percent. Given the changing demographic profile and the predictions of longevity, it is important to examine aging in the United States and to consider the public health implications of this growing and heterogeneous population. We must recognize that this current cohort has different medical, psychological, and social needs now than it did in the past. At the turn of the 20th century, the leading causes of death were largely infectious diseases, including influenza and pneumonia; whereas in the 21st century the leading causes of death for individuals age 65 and over are heart disease, cancer, chronic lower respiratory disease, and Alzheimer's disease (chronic conditions), which represents a profound shift in disease burden. Those with multiple chronic diseases often require costly long-term management and care. For example, expenditures for home care and long-term care have escalated significantly and it is projected that these costs will increase 20 percent between 2000 and 2020.[4] However, if public health interventions are developed to promote independent function and to decrease the burden of disability, projected expenditures can be reduced.

The controversy about the age at which old age begins is noteworthy to mention; that is, nothing magical occurs at the chronological age of 65 that marks a person as "older." The age of 65 was not derived from a biologic process; Bismarck originally conceived it through his evaluation of German social demographic data. Even so, in the United States, the age of 65 generally is perceived as the beginning of old age. It is clear to gerontologists, however, that the process of aging starts decades earlier in life and an individual's chronologic age often is not an accurate predictor of function.

If one conceptualizes the aging population to include individuals from 65 years of age to death, old age encompasses a span of as many as 35 years or more. Thus, some gerontologists view the elderly population as being composed of several age cohorts: the *young-old* (individuals 65 to 74 years old), the *middle-old* (individuals 75 to 84 years old), and the *old-old* (individuals 85 years and over). Each of these age-groups has its own unique and different social support and psychological needs as well as different types of medical problems. For instance, the old-old are more likely to have cognitive impairments, poorer physical health, and fewer financial and social resources. The old-old, for example, are less likely to be consumer-oriented than the young-old cohort. Also, health literacy decreases with age in patients 65 years of age and older. The combined effect of greater consumerism and a higher level of health literacy among the young-old than the old-old are likely to result in different interactions with physicians and with the health care system in general.

A public health perspective on aging mandates attention to aging beyond the biomedical focus. In this chapter, we provide an overview of aging, including an examination of the medical, psychological, and social components of older age. We know that adverse circumstances have a cumulative effect over the life span, so that factors such as poverty and discrimination over the years result in negative health

consequences later in life. This "cumulative adversity" refers to the "layering, intensification, or enlargement of the spheres of hardship that develop over a span of time."[5] Thus, it is worthwhile to explore the needs of the older population within the life course perspective. This chapter embraces a health promotion and disease prevention approach to organizing the public health issues facing the elderly. We prioritize some of the major issues confronting older people, although additional relevant topics are omitted due to space limitations.

PRIMARY PREVENTION AND HEALTH PROMOTION

The majority of adults aged 65 and older live with many different chronic diseases. It is estimated that approximately 80 percent of all persons age 65 and older have at least one chronic condition, and 50 percent have at least two.[4] Nonetheless, health promotion and primary prevention of disease remain vital for preserving health and reducing risk of disability in later life. Despite a high prevalence of chronic diseases, in general, older adults are healthier, and have fewer functional limitations now than they have had in previous generations.[6] This can be largely attributable not only to advances in medical therapies and health care technology, but also to primary prevention and health promotion. To prevent or decrease the likelihood of accruing additional chronic diseases, and to promote enhanced quality of life and independence, primary prevention remains critical for the overall health and well-being of older adults. The most important and effective aspects of primary prevention are reviewed below.

Nutrition

Adequate nutrition is a key component of the health, independence, and quality of life of older adults. It is estimated that 85 percent of noninstitutionalized older people have one or more chronic conditions that could improve with proper nutrition, and that up to 50 percent of these individuals have clinically diagnosed problems that require nutrition intervention.[7] Moreover, in the most recent data from the National Health and Nutrition Examination Survey, only one-third of adults over age 55 (33.6 percent) maintained a healthy weight.[7] It is especially important to consider how economic, social, and cultural factors influence older individuals' choices and access to good nutrition.

Obesity remains a major risk factor in later life for debility, functional decline, and other chronic illnesses, just as it is in earlier stages in life. In one large survey of noninstitutionalized adults over the age of 70, 36 percent had a body mass index of 27 or higher, indicative of obesity.[8] The majority of this group of older people consumed levels of dietary lipids such as saturated fats above the levels recommended for the population, and inadequate levels of one or more nutrients.[7] It is important for older adults to be educated about healthy dietary choices, and more important for clinicians to be aware of the nutritional health needs of the older patient.

On the flip side, nutritional assessment should in some cases focus on weight maintenance, as risk of undernutrition increases and becomes significant with advanced age. A thorough nutritional assessment is recommended for all older adults. As part of normal aging, a healthy person loses 0.1 to 0.2 kilograms per year after

age 70. This is due to loss of lean muscle mass. Weight loss is considered significant if it constitutes more than 5 percent of body weight in the past 6 to 12 months.[8] It is important to understand that undernutrition and malnutrition in older adults have numerous potential causes including the side effects of medications contributing to anorexia; mental health problems such as depression or late-life paranoia; alcohol abuse; medical conditions such as swallowing disorders, hyperthyroidism, or malabsorption; difficulty food shopping (due to mobility issues, safety of neighborhood); lack of financial resources to purchase food; and lack of access to adequate grocery stores.[9] Other important elements include education and cultural traditions that guide nutritional choices.

Certain nutritional deficiencies are particularly important for older adults. Calcium and vitamin D play a crucial role in the prevention and treatment of osteoporosis. Osteoporosis is highly prevalent and often undiagnosed in the geriatric population.[10] More than 25 million adults in the United States have or are at risk of developing osteoporosis, a condition that significantly increases the risk of fractures.[11] Osteoporosis is most often associated with inadequate calcium intakes (generally less than 1,000–1,200 milligrams/day), but insufficient vitamin D contributes to osteoporosis by reducing calcium absorption.[11] Vitamin D is unusual in that it is converted into its active form in the skin after exposure to ultraviolet light. Older adults are at increased risk of vitamin D insufficiency for several reasons: (1) decreased time spent outdoors and increased use of sunscreen, (2) decreased overall intake of milk products, (3) reduced ability for the skin to synthesize vitamin D efficiently, and (4) decrease in the ability of the kidney to convert vitamin D to its active hormone form. Screening for vitamin D deficiency has been recommended for all older adults. Supplementation of both calcium and vitamin D is beneficial for the prevention of fractures in all older adults.[12]

Additional attention to screening for other nutritional deficiencies should be considered for at-risk groups, such as those older individual who are alcoholic, poor, frail, and lack access to healthy foods.

Exercise and Physical Activity

The interaction of physical activity and exercise with health and aging is complex and multifaceted. Research has supported many potentially positive effects of physical activity on the aging process in general. More specifically, regular physical activity has been shown to have beneficial effects on almost every aspect of health—both psychological and physical well-being. In fact, regular physical activity has been shown to (1) minimize the physiological changes associated with typical aging, (2) increase longevity and decrease risk of many chronic diseases, (3) treat some chronic diseases, and (4) perhaps most dramatic and important, assist in the prevention and treatment of disability.[13] Because of the relatively higher prevalence of both chronic disease and disability in the older adult population, regular physical activity can play an important role in maintaining independence and overall health.

Hypertension, hyperlipidemia, and obesity are all independent, major risk factors for stroke, diabetes, and coronary artery disease. These risk factors, separately and additively, increase the risk of coronary artery disease and stroke by significant amounts. Vast and compelling data support that heart disease and stroke are potentially

treatable or at least partially reversible by physical activity. Moreover, studies have shown that regular physical activity decreases risk of falls and frailty in older individuals.[14, 15]

Regular physical exercise also has been shown to protect significantly against cognitive decline and dementia. An active lifestyle is associated with lower risk of depression and overall improved quality of life.[15, 16] Beneficial effects have been seen even with leisure-time physical activity, which include some type of repetitive bodily movement (for example, walking) that is performed without a specific intent of improving physical fitness. Exercise regimens or prescriptions should be tailored to the capabilities, restrictions, and needs of each person, to maximize safety and benefit, and to minimize the potential for injury.[16]

Immunizations

Immunizations are a mainstay of preventive treatments that have been shown to be effective in the care of older adults. Medicare covers the costs of influenza, pneumococcal, and tetanus immunizations, which will be discussed below. Vaccines for other illnesses such as herpes zoster, which can lead to the very painful condition called shingles, are also available but not offered as standard of care, and hence will not be discussed. Getting immunized for *seasonal influenza* is hugely important for the elderly. The current influenza vaccine is a killed virus that has an efficacy rate of 70 percent for illness and 90 percent for mortality. It has been shown to reduce the rates of respiratory illness, hospitalization, and mortality in the elderly. Annual vaccination must be provided because there is short-lived protection provided by the vaccine (four to five months) and because there is antigenic drift of the virus itself from year to year. It is recommended that all people over the age of 50 be vaccinated every year during the influenza season.[17]

Protective efficacy of the *pneumococcal vaccine* is estimated to be between 60 and 70 percent, and may be lower for high-risk older adults. Moreover, studies have revealed that although the vaccine does not protect against mucosal diseases such as sinusitis, and does not always protect against pneumonia, it does reduce the risk of invasive disease such as bacteremia.[18] Adverse effects following vaccination are rare and mild. Therefore, it is recommended that all adults over the age of 65 receive one dose of the vaccine. If more than five years has elapsed since the first dose, and the person was vaccinated before the age of 65, then a repeat vaccine is indicated.[19]

More than 60 percent of *tetanus infections* occur in people over the age of 60. Older adults who have never been vaccinated should receive two doses one to two months apart, and then an additional dose 6 to 12 months thereafter. The optimal interval for booster doses has not been well established. The U.S. Preventive Services Task Force recommends booster vaccinations every 10 years.

POLYPHARMACY

Polypharmacy (that is, the use of multiple medications or the administration of more medications than are clinically indicated, representing unnecessary drug use) is increasingly common among the geriatric population. Research has shown that polypharmacy continues to increase and represents a risk factor for both morbidity and mortality. The average person over the age of 65 uses two to three prescription

drugs and one to three over-the-counter medications or supplements. The most common uses of these medications are for cardiovascular disease and arthritis. Older adults are particularly susceptible to drug-related side effects because (1) aging affects how their body metabolizes medications and (2) they take more medications than younger individuals.[20, 21]

The most consistent risk factor for adverse drug reactions is increased number of drugs. The most common cause of adverse drug reactions is drug-drug interactions. It is not surprising then that the risk rises exponentially as the number of prescription drugs taken increases. Common side effects of polypharmacy reflect common complaints of the geriatric population—dizziness, confusion, loss of balance, constipation, depression, and fatigue. A health care provider should conduct a thorough review of all medications, both prescribed and over the counter (including vitamins and health food supplements), on a routine basis.

Tobacco Use

Tobacco use is a major risk factor for chronic obstructive pulmonary disease, cardiovascular disease, many cancers, and decreased functional status. For example, incidence of myocardial infarction in female smokers is estimated to be approximately six times that of female nonsmokers, and three times that of male smokers compared with male nonsmokers. Risk of myocardial infarction may decline by as much as 50 percent within one year of smoking cessation and may normalize to that of nonsmokers within two years. The benefits of smoking cessation are seen regardless of how long or how much the person smoked previously. Smoking cessation also leads to an improvement in exercise tolerance. Education and promotion of smoking cessation is equally important if not more so for older adults (because of their relatively higher risk profile for cardiovascular disease) than it is in younger adults. It truly is never too late to quit smoking.[22]

Life expectancy has increased dramatically in most developed countries over the last century. Concomitant with longevity is the issue of the quality of life that accompanies longer life span. The adoption of a healthy lifestyle and other preventive measures can exert their greatest influence by postponing the onset of morbidity and debility in old age. This is known as the "compression of morbidity" hypothesis.[14] Data support the hypothesis that there is significant reduction and postponement of disability with an overall healthier lifestyle in an aging population.

SECONDARY PREVENTION

A major component of a complete geriatric assessment is the identification of risk factors for diseases associated with high levels of morbidity and mortality. These practices are aimed toward early detection and prevention of disease progression. Stroke prevention, falls, and environmental modifications are discussed below because of their high association with morbidity in the older adult population.

Stroke Prevention

Stroke is a leading cause of death and long-term disability in the United States.[23] The risk of stroke doubles each decade after the age of 55.[23] Risk

factors for stroke in the older adult population include obesity, excessive alcohol intake, smoking, high blood pressure, diabetes, and atrial fibrillation. Controlling these lifestyle choices and medical diseases reduce the risk of stroke and its long-term outcomes.

Fall Prevention

Not surprisingly, the incidence of falls increases with age. Falls are the leading cause of injury-related deaths and approximately one-third of older adults fall each year.[24] Having a fall is one of the greatest predictors of loss of independence for an older person. Falls are associated with significant morbidity and mortality in older adults, and complications resulting from falls are among the leading causes of death of people age 65 and over. Falls also are associated with an overall decline in functional status, enhanced use of medical services, and higher risk of nursing home placement. Older adults who fall also often times develop an intense fear of falling, which contributes to their loss of independence as they become less mobile.

Environmental Geriatrics

Most houses and apartments are not designed with older people in mind. Environmental geriatrics focuses on fostering independent function by retrofitting living arrangements to accommodate those with balance issues, walkers, or wheelchairs. For older individuals who can afford it, telemonitoring devices, such as Lifeline, can extend safety and a sense of security to those living alone. Environmental geriatrics also acknowledges that individuals with dementia are often acutely sensitive to their environments. Appropriate colors and helpful orientation with clocks and calendars may be of benefit to this population.[25, 26]

TERTIARY PREVENTION

Tertiary prevention refers to efforts to promote maximum functionality and health promotion once an individual already has a disease. Given the often multiple chronic illnesses many older people have, maximizing efforts to maintain wellness and limit the impact of disease on functional living is an essential part of promoting health and delaying disease progression. For example, if a patient has chronic obstructive pulmonary disease (COPD), enrolling in a pulmonary rehabilitation program can provide significant benefits.[27] Tertiary prevention for older people also focuses on improving the health literacy of patients. Older patients have the lowest health literacy of any population group—if older patients can better understand their symptoms and their medications, and learn the skills to better communicate with their health professionals, this knowledge can advance their abilities to stay well.[28]

Care Transitions

Older adults with chronic illnesses often require multiple care providers and receive care in a variety of settings. Care transitions refers to the movement patients make through different health care settings as their disease and care needs change during the course of an acute or chronic illness.[26] This transition also includes a change in health care providers. For example, an older adult with COPD, coronary

artery disease, and osteoarthritis may require increased attention for exacerbation of a chronic disease or an acute illness such as pneumonia. This care may begin with the primary care physician in the office and move to the hospital setting where a new set of nurses and physicians will take over. This patient may then require transitioning to a skilled nursing facility for additional medical and rehabilitation therapy where another set of nurses and physicians will provide care before returning home to the care of a visiting nurse and the primary care physician again.

Transitional care is defined by the American Geriatrics Society as a set of actions designed to ensure coordination and continuity of health care as patients transfer between different locations or different levels of care within the same location.[26] Transitional care promotes a comprehensive plan of care, including a summary of care provided by one location; an updated list of medical diagnoses, medications, and allergies; and a follow-up plan and instructions to follow if a condition worsens. Excellence in transitional care management is essential for persons with complex care needs and plays a critical role in tertiary prevention in the elderly. A growing number of national, state, and institution-level initiatives have been designed specifically to improve this pivotal aspect of medical care.

In addition, the transfer from one setting and care provider to another is complex and challenging for older adults and their caregivers. Older adults are at increased risk for adverse events in a fragmented health care system where often the only constant member is the patient. Care transitions often place increased responsibility on patients and their caregivers to understand and communicate information about their hospital stay, laboratory test results, medications, and so on to the care providers at the next facility. One study demonstrated that 19.6 percent of Medicare beneficiaries were rehospitalized within 30 days of an initial admission. Readmissions are quite costly as unplanned rehospitalization costs are approximately $17.4 billion.[29]

Long-Term Care

Long-term care encompasses a variety of services and supports used by people who require assistance to function in their daily lives. This care can be provided to people of all ages; however, primary users are older adults who experience increased functional disability and chronic illnesses as they age. Older women are particularly at risk for requiring such care because of their higher rates of disability and longer life expectancy.[30] An estimated 12 million people over the age of 65 will require long-term care services by 2020.[31] Most long-term care involves nonskilled assistance with activities of daily living, such as bathing, dressing, toileting, transferring, and feeding. Other services included under the umbrella of long-term care are rehabilitation, home health care, assisted transportation, and care coordination. These services can be provided regularly or intermittently over the course of weeks, months, or years.

This type of care is provided along a spectrum of locations from home to facilities. For older adults living at home, they can receive long-term care through services such as senior centers, geriatric care managers, emergency response systems (such as Lifeline), volunteer visiting services, home health care, meals on wheels, and transportation services. Most long-term care is provided informally by family and friends in the home.[32] Facility-based long-term care can range from assisted

living to nursing homes. A new trend in facility-based care is a continuing care retirement community (CCRC). These communities provide a variety of levels of care from independent housing to nursing home care, all within one system and often at one location.

Paying for long-term care, however, is a significant problem for most elderly. Public funding in the form of Medicaid accounts for the largest payer of long-term care, but Medicaid is available only to the very poor. Medicare pays for a temporary long-term care services, specifically for skilled needs in a skilled nursing facility or at home with home health care. Medicare does not pay, however, for personal care and does not provide prolonged coverage in many cases. Family and significant others who provide care at home deal with significant physical, emotional, and financial costs.

Palliative Care

Planning for the future cannot be complete without a discussion about personal preferences regarding medical care. Setting goals for treatment and care may be more challenging with the elderly patient as many of the problems encountered in geriatric medicine cannot be resolved. The inevitable decline of the elderly patient spells a need for balanced and realistic goals for the physician, the patient, and the patient's family or other designated health care proxy.

As the majority of deaths in the United States occur in older adults, palliative care has an important role in geriatric medicine where the focus is on the prevention and relief of suffering. Geriatric palliative care focuses on reducing the severity of disease symptoms for all stages of illness while attempting to maintain quality of life, which usually includes aggressive symptom management for those with chronic illness such as heart failure and COPD; reducing symptoms for those going through life-prolonging treatments such as chemotherapy; and providing comfort and symptom relief to patients during end-of-life care. Additionally, palliative care incorporates support to the patients' family and friends who are experiencing a variety of emotions during such a critical stage of an older adult's life.

Health care providers must be able to recognize the time at which the potential risks of treatment outweigh the benefits and must help patients and families navigate those decisions carefully. Acceptance of death and skills in helping a patient die with dignity are prerequisites for the physician caring for older patients. These abilities require special insight, self-awareness, and training. It is not ageist to recognize that we are all mortal. As the end of life draws near, the patient, his or her family, and the physician need to carefully evaluate the goals of care and then specifically evaluate the utility or futility of each potential intervention and the possible suffering that such measures may produce. Indeed, a mature perspective on public health and aging needs to incorporate both appropriate health promotion and disease prevention measures to enable older people to age successfully, and also to support older adults and their families to orchestrate their most appropriate end-of-life care choices.

MENTAL HEALTH

An important public health epidemic in aging is the growing incidence and impact of dementia. The incidence of dementia among older people in their 60s is low (for example, the rate of moderate to severe dementia is about 2 percent among

persons 65 to 69 years of age); however, the incidence of cognitive dysfunction progressively increases with age. In the United States, the most common etiology for dementia is Alzheimer's disease, which accounts for approximately 60 to 80 percent of dementia cases; vascular dementia is the next most prevalent dementia, affecting approximately 10 to 20 percent.[33] Alcohol-related dementia and dementias secondary to medication side effects, depression, and other central nervous system illnesses account for the rest. Alzheimer's disease is estimated to affect at least 4 million mainly older Americans. Cognitive deficits often develop insidiously in Alzheimer's disease with the loss of memory of recent events being the essential characteristic of the disease. In the later stages of the disease, Alzheimer's victims are completely dependent on others for the basic activities of daily living. Eventually, almost 75 percent of patients with dementia end up in a long-term care institution and often stay for long periods of time.[34]

The impact of dementia on society is immense with an estimate of at least $100 billion spent annually. This cost includes the costs of medical care, long-term care, home care, and also the lost productivity of caregivers.[34] Furthermore, the emotional impact of dementia is devastating for both patients and their families. Nearly half of caregivers for patients with dementia have psychological stress, especially depression. In most urban areas, respite care and support groups are available through local agencies on aging or through the Alzheimer's Association. Information on the status of clinical trials in Alzheimer's disease is available on the Web (for example, see www.clinicaltrials.gov, which lists government-funded research). Though there have been advances in understanding the pathophysiology of dementing illnesses, management of these illnesses is still primarily symptomatic. Hopefully, the future will unfold both disease-specific and disease-modifying therapy to slow down the progression, delay the onset, and ideally prevent Alzheimer's disease.

Older people with serious mental illness are composed of two groups—those with persistent mental illness who are aging, and those who develop mental illness when they are older. Sadly, it is estimated that only 20 to 25 percent of older people with mental health problems actually receive professional care, although depression is one of the most prevalent mental health issues affecting the elderly population.[35] In particular, older men and older African Americans and Hispanics are at increased risk of undiagnosed depression. Yet, depression is often undetected in primary care settings and even when depression is diagnosed, it often is not treated. More than 80 percent of depression is treated in primary care settings and is left primarily for the primary care provider to detect, assess, and treat. Depression later in life leads to unnecessary suffering, increased mortality, decreased functional status, and excess use of the health care system. Unfortunately, depression in late life is often undetected and undertreated, although the elderly have the highest rates of suicide of any age group.[36]

Depression is not part of normal aging. Healthy community-dwelling older Americans actually have a lower prevalence rate of depression as compared with the nonelderly adult population. Rates of depression, however, increase among those older people with comorbid illness, and the prevalence of depression in hospitalized elderly is more than 30 percent. Some risk factors for late-life depression include female gender; social isolation; widowed, divorced, or separated marital status; lower socioeconomic status; comorbid medical condition; uncontrolled pain; insomnia;

functional impairment; and cognitive impairment. The prevalence of depression is extraordinarily high for older people residing in long-term care facilities.

Often depression can manifest differently in the elderly—possibly manifesting with decreased memory or lack of energy—and not necessarily with sadness or other classic symptoms. As part of basic training for individuals caring for the elderly, the syndrome of geriatric depression needs to be taught to and recognized by primary care providers and also by subspecialists (for example, increased depression after stroke, with Parkinson's disease, myocardial infarction, and cancer). In addition, it is important that family and hired professional caregivers learn about geriatric depression and bring it to the attention of the primary health care provider.

SOCIETAL IMPACTS

An increasing public health concern is the rapid growth in caregivers for elderly individuals that are in need of support and assistance. A significant emotional and financial cost is associated with caregiving. One report documents that 44 million caregivers in the United States rendered unpaid care to friends, relatives, and neighbors, which represents 21 percent of all U.S. households. The economic value for these unpaid efforts over one year was about $306 billion. It is important to develop public health interventions which address caregiver stresses and attend to their health needs.[37]

In most developed countries, only a small minority (3 to 6 percent) of older people live in long-term care institutions. In the United States and elsewhere, most older individuals would prefer to stay in their own homes. But, in the United States, paying for long-term care is extremely expensive and most insurance policies do not cover such costs. A few countries such as Germany, Japan, Luxembourg, and the Netherlands have introduced mandatory long-term care insurance plans. More work needs to be directed at adapting homes, helping families to care for elderly relatives, and finding support for the hard work of caregiving. One way to work with these increasing needs is life-care communities where an individual starts in independent living and then progresses through the system as medical, psychological, and social needs increase. This concept, however, has been directed mostly toward older people with sufficient financial resources. Helping the "sandwich" generation coping with the caregiving of both children and aging parents is hugely important to providing optimal health care. Now that many women work full time, traditional caregiving roles are harder to achieve, which can lead to inordinate caregiver stress and insufficient support.

STAFFING SHORTAGES

The 2008 Institute of Medicine report entitled "Retooling for an Aging America: Building the Health Care Workplace"[38] described the health care crisis that is looming before us and argued that the United States is ill-prepared to respond to both the social and health care needs of this burgeoning population. Compounding matters, the number of health care providers with the knowledge of geriatric medicine is inadequate to deal with the increasing numbers of older people with age-related complex health care problems. Ironically, as the elderly population increases, the workforce

necessary to meet the needs of this cohort is inadequate. The report calls for efforts to (1) increase the recruitment and retention of geriatricians as well as health care aides through better salaries and wages; (2) improve training of informal caregivers to enable them to better attend to the needs of aging friends and relatives and also alleviate the stress from providing this care; and (3) focus on new models of care delivery and payment that advance geriatric care for the nation, for example, enable financial feasibility of utilizing interdisciplinary teams.

Few health professionals are properly trained in geriatrics, few physicians choose to become geriatricians, and even fewer choose geropsychiatry. Currently, only approximately 7,100 certified geriatricians are in practice.[38] An Institute of Medicine report estimated that 36,000 geriatricians are needed to address the needs of the aging U.S. population by 2030. However, assuming current rates of growth and attrition, one estimate predicts that only 7,750 geriatricians will be practicing in 2030.[38] Unless interventions are made to encourage the training of geriatricians and other health professionals, there will be a dearth of professionals trained to care for the baby boomers who are approaching age 65.

Geriatric medicine is different from internal medicine, which tends to treat the younger patient. Older patients often have multiple medical problems that mask one another or make the treatment process for any one disease more difficult. Common diseases often present atypically in older people, yet many physicians have not been taught specific diagnostic skills for evaluating the geriatric patient. For example, cognitive dysfunction in an elderly patient must be carefully scrutinized to rule out underlying thyroid disease, vitamin B12 deficiency, depression, and other medical conditions. Classic signs for coronary artery disease, such as chest pain or shortness of breath on exertion, also may be more difficult to track in an older person with severe osteoarthritis who is unable to ambulate.

Older patients usually have accumulated more extensive medical histories than younger patients. Considerable expertise is required to discriminate important and relevant clinical problems in an initial evaluation. Many geriatricians will acknowledge that it may take two or three visits to assess the geriatric patient adequately. Not only is it important to obtain a comprehensive medical history, it is critical to secure a social and psychological history as well. This task can be daunting given the time constraints of contemporary medicine and the decreased reimbursement for the cognitive tasks of medicine as compared with reimbursements for procedures. Often in geriatric care, a team of nurses, physicians, social workers, mental health professionals, and others are necessary to address patients' multidimensional concerns. In many ways, the geriatric team is the "geriscope" of geriatric care; unfortunately, however, most geriatric medical care is practiced without such a team approach because of system-mediated financial constraints. Providing comprehensive, interdisciplinary geriatric care is expensive and, at present, not incorporated into Medicare reimbursement.

FINANCIAL IMPLICATIONS OF AN AGING SOCIETY

All is not bleak, however. For the most part, older individuals are in comparatively good health thanks to modern medicine. The period of poor health that usually precedes death has become shorter; the major part of health care spending is concentrated on the last year or two of life and, in particular, in the last six months

of life. But even older people who are fairly fit require more health care than younger individuals.[39]

How much leisure time can be expected in an American's "golden" years? Should older individuals be encouraged to work longer? Some countries already have raised their official retirement age; other countries are debating whether it makes sense to have a specific retirement age at all.[40] It would be economically advantageous to have people working at least a few more years than the typical retirement age of 65, as this would address some of the concerns (medical and economic) associated with aging populations.[41] As the goals and meaning of retirement have changed with the profound changes in longevity, getting people to work for somewhat longer would decrease the loss of resources from pension benefits, and these individuals would continue to pay taxes and contribute to Social Security. Many Americans do not see retirement in their foreseeable future because of financial concerns.[42] With the recent economic downturn and the shrinkage of their retirement accounts, they fear that they will not be able to afford retirement and will need to keep working—but also partly because they like to work. At the same time, though baby boomers see work as part of their future, they would like work to be less onerous, which can perhaps be best solved by working part time.

ELDER MISTREATMENT

Elder abuse and mistreatment are sad and unfortunately prevalent aspects of aging. For many reasons, elderly men and women, with or without impairments or dependency on family care, are vulnerable to mistreatment. Each year, about 1 million older Americans are injured physically, debilitated psychologically, or neglected by a family member.[43] Elder abuse is conceptualized as an act of commission; neglect as an act of omission. Abuse is thought to occur in 4 percent of the elderly population; as the geriatric population expands, the incidence will also increase. One part of the public health concern is that many older people are not aware of this phenomenon and suffer in silence or shame. Abuse of older people is associated with depression, cognitive impairment, loss of functional capacity, and increased mortality.[44] In addition to suffering physical injuries, these victims often develop overwhelming feelings of fear, isolation, shame, and anger and need extensive counseling to regain independence. Given the present incidence of mistreatment and the projected growth of the elderly, it is imperative that health care providers who care for the elderly learn to recognize and intervene on behalf of victimized patients. However, it must be recognized that the information base about abuse and neglect of the elderly is limited and more research is needed to better understand the causes and prevention of this problem, as well as ways to intervene.

In defining abuse and neglect of the elderly, the following questions need to be considered:

1. Who is being abused and neglected? The victim may be competent or incompetent, healthy or frail, male or female. Elder mistreatment affects all social classes, religions, and race-ethnic groups.
2. Who is doing the abusing or neglecting? Abuse or neglect may be family mediated, may be inflicted by a hired caregiver, or may be at the hands of unrelated perpetrators.

3. Where is abuse or neglect occurring? Abuse or neglect can occur in the elderly person's home, in the perpetrator's home, in a shared living arrangement, or in an institution.
4. What is the pattern of abuse and neglect? Although a one-time act of violence can be damaging, abuse or neglect of the elderly is commonly characterized by a pattern of violence increasing in severity and incidence over time.
5. What types of abuse or neglect are occurring? Abuse or neglect can include physical, psychological, or financial mistreatment.
6. Is the mistreatment intentional or unintentional? In some cases, abusers deliberately mistreat to cause harm; in other cases, the resulting harm is not deliberate, and interventions should vary according to the intent to harm. For example, improper medication of a relative because of poor understanding of the physician's directions may result in drug toxicity. Clearer instructions can remedy this unintentional mistreatment. However, deliberate inappropriate medication of an older relative to control his or her ability to function independently requires more intensive interventions.

The three categories of abuse and neglect are psychological, physical, and financial. All can be intentional or unintentional. Psychological abuse and neglect encompasses a range of behavior that causes emotional stress or injury to an older person. This behavior includes verbal abuse—threatening remarks, insults, or harsh commands—as well as being silent or ignoring the person. One form of psychological abuse is infantilism (a form of ageism), whereby the elderly individual is treated as a child, which both patronizes and encourages the person to passively accept a dependent role. Physical abuse and neglect produces a wide range of bodily injuries. Examples of this type of mistreatment include striking, shoving, shaking, beating, restraining, or improper feeding. Sexual assault, included in this category, requires special emphasis, because many health care providers find this form of violence inconceivable when an older individual is involved. Sexual assault refers to any form of sexual intimacy without consent or by force or threat of force. Financial abuse is the misuse or exploitation of or inattention to an older person's possessions or funds. This form of mistreatment includes conning, pressuring the victim to distribute assets, or irresponsibly managing the victim's money.

After mistreatment is identified, a comprehensive assessment needs to be performed, preferably with the participation of an interdisciplinary team. After the assessment is done, an intervention plan can be developed. With competent victims, the health care provider presents options and the victims decide how to proceed. With incompetent victims, decisions are made by the courts. With a cognitively impaired, incompetent victim, an interdisciplinary team is formed to make a decision. Decisions need to be made with a full awareness of the severity of the violence, the lifestyle choice history of the individual, and the legal ramifications. Often, there is no single correct decision, and each case must be carefully followed up. Because various state protective service laws govern intervention and reporting for incompetent victims, health care providers should be familiar with the laws and protective service resources available.

AGEISM

Profound opportunities exist to promote a healthy, productive, and meaningful old age for Americans. Unfortunately, a negative perception of aging and older people persists in youth-oriented U.S. culture. While stereotyping based on sexism and

racism is considered taboo, ageist biases abound. Jokes about the physical, cognitive, and sexual problems of older people are present in the popular culture, as is evident on every media front, ranging from greeting cards to national television. We laugh at the thought of old people having sex; we never think about old age as being productive or exciting. Ageist biases ultimately get played out in the workplace, within our health care systems, and within society at large. The United States is not alone with regard to attitudes toward aging; a study in Great Britain demonstrated that ageist attitudes far eclipsed sexism, racism, and negative attitudes to younger people with disabilities.[45]

Ageism, the system of destructive false beliefs about older people, is pervasive in America and its origins are multidimensional. Fears of aging and death, a subject still taboo in U.S. society, play a large role in the development of ageist attitudes. Fears of obsolescence and physical and mental losses in a society preoccupied with productivity and youth advance an ageist perspective. Distinct from other -isms, such as sexism and racism, ageism has a dangerously personal focus; we all become old—that is, if we are fortunate enough to survive. If an individual becomes old and has incorporated a significant dislike for aging or older people, however, this older individual then becomes the object of his or her own discrimination. This self-directed prejudice has ominous implications for successful aging—that is, older people may be ignorant about normal aging or have preconceived ageist attitudes that affect medical care, accessing entitlements, or impact on socialization with other older individuals. In terms of accessing medical care, if an older individual believes, for example, that urinary incontinence or impaired memory is part of normal aging, he or she will not readily seek medical attention and possible treatment will not be pursued.

Ageist bias is relevant especially in the context of public health and medical care, as these attitudes may cause health providers and patients to trivialize or deny needs for care. It is likely that health care professionals are more susceptible than the layperson to the development of ageist attitudes. By definition of their work, they primarily are exposed to the most vulnerable elderly population: the ill, the frail, the confused, the demented, and the hospitalized. Robust older people generally are not part of the patient sample, especially for the professional working in a tertiary care institution. Ageism, therefore, may be an occupational hazard of the health care professional and can undermine medical care.[46] For example, the physician who believes that depression is a component of normal aging will miss the opportunity to treat this potentially reversible condition. Several studies have documented that older people are treated less aggressively in the treatment of breast cancer [47] and the use of thrombolytic therapy for coronary artery disease.[48] In addition, negative ageist attitudes may influence the receptivity and willingness of the health care system to invest in programs that promote health and prevent disease for the elderly. It is time to abandon ageist perceptions and their permeation throughout the U.S. social and health care systems. It is time to develop a new paradigm for the different stages of aging and use collective creativity to provide appropriate societal supports to promote the best possible old age.

CONCLUSION

The United States, as well as other developed nations, is on the cusp of an explosion in the number of older individuals in its population. Life expectancy is

increasing; indeed, with improved health many in the field of aging consider individuals in their 60s and 70s to be in late middle age. Yet, the number of trained health professionals is vastly lacking to meet the growing demand of this aging population. We need to think through the social, political, economic, and medical implications of this growing heterogeneous population. For example, a large subset of older people are living productive and independent lives. What are the factors that promote successful aging in the United States and what innovative programs can be developed to advance this optimal old age? On the other hand, there is a burgeoning population of elderly with cognitive impairment. What can be done to improve the quality of the life for these individuals and their caregivers? A research focus on aging and public health is vital to carefully determine what resources are necessary to promote health and improve quality of life at each stage of aging.

REFERENCES

1. US Department of Health and Human Services, Administration on Aging. A profile of older Americans: 2008. http://www.mowaa.org/Document.Doc?id=69. Accessed December 9, 2009.
2. Burton L, Kasper J. Demography. In: Cobb E, Duthie E, Murphy J. *Geriatric Review Syllabus: A Core Curriculum in Geriatric Medicine*. 6th ed. New York: American Geriatrics Society; 2006.
3. A slow-burning fuse: a special report on ageing populations. *The Economist*. June 27, 2009.
4. Centers for Disease Control and Prevention. Public health and aging: trends in aging United States and Worldwide. *MMWR*. 2003;52:101–106.
5. Zarit S, Pearlin L. Special issue on health inequalities across the life course. *J Gerontol Psychol Sci Soc Sci*. 2005;60:S6.
6. Federal Interagency Forum on Aging-Related Statistics. Older Americans 2008: key indicators of well-being. http://www.agingstats.gov/agingstatsdotnet/Main_Site/Data/Data_2008.aspx. Accessed December 9, 2009.
7. Posner BM, Jette AM, Smith KW, Miller DR. Nutrition and health risks in the elderly: the nutrition screening initiative. *Am J Public Health*. 1993;83:972–978.
8. Schoenborn C, Heyman K. Health characteristics of adults aged 55 years and over: United States, 2004–2007. *Natl Health Stat Reports*. 2008;16:1–32.
9. Chernoff R. Nutrition and health promotion in older adults. *J Gerontol A Biol Med Sci*. 2001;56:47.
10. Thomas D. But is it malnutrition? *J Am Med Dir Assoc*. 2009;10:295–297.
11. Kamel H, Hussain M, Tariq S, Perry H, Morley J. Failure to diagnose and treat osteoporosis in elderly patients hospitalized with hip fracture. *Am J Med*. 2000;109:326–328.
12. Is O. Osteoporosis prevention, diagnosis, and therapy. *JAMA*. 2001;285:785–795.
13. Raisz L. Screening for osteoporosis. *N Engl J Med*. 2005;353:164.
14. Singh M. Exercise comes of age: rationale and recommendations for a geriatric exercise prescription. *J Gerontol A Biol Med Sci*. 2002;57(5):262–282.
15. Hubert H, Bloch D, Oehlert J, Fries J. Lifestyle habits and compression of morbidity. *J Gerontol A Biol Med Sci*. 2002;57:347.
16. Nelson M, Rejeski W, Blair S, et al. Physical activity and public health in older adults: recommendations from the American College of Sports Medicine and the American Heart Association. *Circulation*. 2007;116(9):1094–1105.

17. Singh N, Clements K, Singh M. The efficacy of exercise as a long-term antidepressant in elderly subjects: a randomized, controlled trial. *J Gerontol A Biol Med Sci*. 2001;56:497.
18. Couch R. Prevention and treatment of influenza. *N Engl J Med*. 2000;343:1778.
19. Advisory Committee on Immunization Practices. Prevention of pneumococcal disease: recommendations of the advisory committee on immunization practices. *MMWR Recomm Rep*. 1997;46(RR-8):1–24.
20. Jackson L, Neuzil K, Yu O, et al. Effectiveness of pneumococcal polysaccharide vaccine in older adults. *N Engl J Med*. 2003;348:1747.
21. Hanlon J, Schmader K, Boult C, et al. Use of inappropriate prescription drugs by older people. *J Am Geriatr Soc*. 2002;50:26–34.
22. Rathore S, Mehta S, Boyko W Jr, Schulman K. Prescription medication use in older Americans: a national report card on prescribing. *Fam Med*. 1998;30(10):733–739.
23. Lloyd-Jones D, Adams R, Carnethon M, et al. Heart disease and stroke statistics 2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation*. 2009;119:480–486.
24. Centers for Disease Control and Prevention. Falls among older adults: an overview 2009. http://www.cdc.gov/HomeandRecreationalSafety/falls/adultfalls.html. Accessed November 18, 2009.
25. This Caring Home. http://thiscaringhome.org. Accessed December 10, 2009.
26. Coleman EA. Falling through the cracks: challenges and opportunities for improving transitional care for persons with continuous complex care needs. *J Am Geriatr Soc*. 2003;51:549–555.
27. Lacasse Y, Brosseau L, Milne S, et al. Pulmonary rehabilitation for chronic obstructive pulmonary disease. *Physiotherapy*. 2002;88:720–721.
28. Scott T, Gazmararian J, Williams M, Baker D. Health literacy and preventive health care use among Medicare enrollees in a managed care organization. *Med Care*. 2002;40:395–404.
29. Jencks S, Williams M, Coleman E. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med*. 2009;360:1418.
30. Houser, AN. Women and Long Term Care. April, 2007. http://www.aarp.org. Accessed December 10, 2009.
31. National Clearinghouse for Long-Term Care Information. Own your future. http://www.longtermcare.gov/. Accessed December 10, 2009.
32. Houser, AN. Long term care. Oct 2007. http://www.aarp.org/ppi. Accessed December 10, 2009.
33. Plassman B, Langa K, Fisher G, et al. Prevalence of dementia in the United States: the aging, demographics, and memory study. *Neuroepidemiology*. 2007;29:125–132.
34. Barton C, Small G, Yaffe K. Dementia. In: Cobb E, Duthie E, Murphy J. *Geriatrics Review Syllabus: A Core Curriculum in Geriatric Medicine*. 6th ed. New York: American Geriatrics Society; 2006.
35. Alexopoulos G, Kelly RJ. Research advances in geriatric depression. *World Psychiatry*. 2009;8(3):140–149.
36. Conwell Y, Thompson C. Suicidal Behavior in Elders. *Psychiatr Clin North Am*. 2008;31:333–356.
37. Centers for Disease Control and Prevention, Kimberly-Clark Corporation. *Assuring Healthy Caregivers, A Public Health Approach to Translating Research into Practice: The RE-AIM Framework*. Neenah, WI: Kimberly-Clark Corporation; 2008.

38. Institute of Medicine. *Retooling for an Aging America: Building the Health Care Workforce.* Washington, DC: The National Academies Press; April 2008.
39. Beisecker A. Older persons' medical encounters and their outcomes. *Res Aging*. 1996;18:9.
40. Cooke M. Policy changes and the labour force participation of older workers: evidence from six countries. *Can J Aging*. 2006;25:387–400.
41. Bamia C, Trichopoulou A, Trichopoulos D. Age at retirement and mortality in a general population sample: the Greek EPIC study. *Am J Epidemiol*. 2008;167(5): 561–569.
42. Cahill K, Giandrea M, Quinn J. Retirement patterns from career employment. *Gerontologist*. 2006;46:514.
43. Aravanis S, Adelman R, Breckman R, et al. Diagnostic and treatment guidelines on elder abuse and neglect. *Arch Fam Med*. 1993;2:371.
44. Adelman R, Breckman R, Lachs M. Elder abuse and neglect. In: Ammerman R, ed. *Assessment of Family Violence: A Clinical and Legal Source*. 2nd ed. Somerset, NJ: John Wiley and Sons, Inc.; 1998.
45. Ray S, Sharp E, Abrams D. *Ageism: A benchmark of public attitudes in Britain*. Canterbury, UK: Age Concern England/University of Kent; 2006.
46. Adelman RD, Greene MG, Ory MG. Communication between older patients and their physicians. *Clin Geriatr Med*. 2000;16:1.
47. Greenfield S, Blanco D, Elashoff R, Ganz P. Patterns of care related to age of breast cancer patients. *JAMA*. 1987;257:2766.
48. Cobbs E, Duthie E, Murphy J. *Geriatrics Review Syllabus: A Core Curriculum in Geriatric Medicine*. 4th ed. Iowa: Kendall/Hunt; 2006.

SECTION 4

PUBLIC HEALTH PREVENTION

This page intentionally left blank

CHAPTER 17

Adolescent Substance Abuse Prevention and Cessation

Gilbert J. Botvin, PhD, Kenneth W. Griffin, MPH, PhD, and Madhuvanti M. Murphy, DrPH

Adolescent substance abuse is a global health concern, which, in recent decades, has gained significant attention from public health researchers, practitioners, and the media. The use of alcohol, tobacco, and illicit drugs, drugs during adolescence can result in lifelong habits that can have pervasive negative effects on physical, psychological, social, and occupational functioning.[1] Although U.S. national survey data have shown that prevalence rates for many substances have gradually declined among adolescents in recent years, substance use and abuse continue to be major public health problems. The abuse of prescription medications among adolescents is a relatively recent phenomenon that is now a leading problem among adolescents, and the abuse of over-the-counter (OTC) medications is also a growing problem. Because the use and abuse of alcohol, tobacco, illicit drugs, and prescription drugs contributes significantly to morbidity and mortality in the United States, reducing substance use and abuse among youth is an important public health priority.

In this chapter, we discuss a number of important issues related to adolescent substance use, including current prevalence rates and recent trends, the etiology of substance use, and types of prevention based on both traditional and current classification methods. Next, we describe different prevention approaches and evidence of effectiveness, with particular emphasis on school-based prevention. Finally, we discuss smoking-cessation approaches designed for adolescents and summarize the available evidence of effectiveness.

PREVALENCE AND CURRENT TRENDS

Prevalence rates of adolescent alcohol, tobacco, and illicit drug use peaked in the late 1970s, began to decrease during the 1980s, and increased again during much of the 1990s. Figure 17.1 illustrates the trends in drug use from 1991 to 2009. Despite a slight uptick in marijuana use in the past few years, prevalence rates for current (past month) use of cigarettes, alcohol (drunkenness), and illicit drugs other than marijuana gradually decreased in recent years among high school seniors. In the

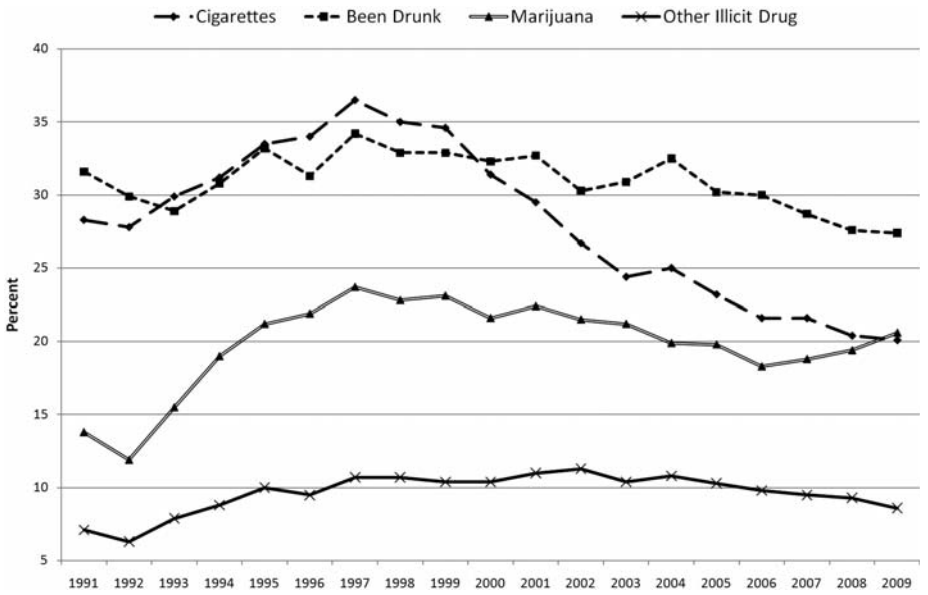

Figure 17.1 Past 30-Day Use among 12th Grade Students, 1991 to 2009. (Source: Monitoring the Future Study, 2009.)

most recent *Monitoring the Future Study,*[2] about half (45 percent) of high school seniors reported ever using cigarettes, and 20 percent reported being current (past month) smokers. This represents a significant improvement since the late 1970s; at that time, twice as many high school seniors (40 percent) were current smokers. Among teens, alcohol use (drinking alcoholic beverages) is the most prevalent substance use behavior, and marijuana is the most prevalent illicit drug used.

In 2008, more than half (55 percent) of 12th graders reported having been drunk at least once in their life, and approximately 25 percent reported binge drinking (five or more drinks in a row) in the past two weeks. Almost half (47 percent) of high school seniors reported ever using marijuana, and 32 percent reported use in the past year. After marijuana, the most commonly abused drugs among high school seniors were amphetamines (11 percent), inhalants (10 percent), sedatives (9 percent), tranquilizers (9 percent), hallucinogens (9 percent), and ecstasy (MDMA) at 6 percent lifetime prevalence. Recent data show that the misuse of OTC medications is a growing problem among young people. Specifically, when taken in large doses, the active ingredient in many cough suppressants (dextromethorphan, DXM) can produce hallucinations or dissociative experiences. Approximately 6 percent of 12th graders reported ever using an OTC cough and cold medication to get high in the past year.[3]

Substance Use Progression

Research has shown that experimentation with one substance frequently will lead to experimentation with others in a generally predictable progression. This developmental progression corresponds closely to the prevalence of substance use in the United States, with alcohol being the most widely used, followed by tobacco

and marijuana. Thus, the majority of individuals begin substance use involvement by using alcohol or tobacco. Some, but not all, will later progress to the use of marijuana. Because of wide availability, some youth use inhalants early in this sequence. For some young people, this progression eventually may lead to the use of depressants, stimulants, hallucinogens, and other drugs. On the other hand, many discontinue use or do not progress from one substance to other more serious substances after a short period of initial experimentation. Knowledge of the developmental progression of substance use during adolescence is important because this progression has implications for the focus and timing of preventive interventions. Programs that target substances that occur toward the beginning of the developmental progression have the potential to prevent the use of these substances and may decrease or eliminate the risk of more serious substance use involvement.

ETIOLOGIC RISK AND PROTECTIVE FACTORS

Knowledge of the etiologic risk and protective factors associated with the initiation and maintenance of tobacco, alcohol, and drug abuse provides a context for understanding the primary targets of preventive interventions. This is because effective preventive interventions target the risk and protective factors for substance use rather than the behavior itself. Thus, to determine the most appropriate intervention, it is important to identify when the onset and escalation of substance use occurs as well as the key etiologic factors that predict the behavior. In the following paragraphs, we review the multiple risks and protective factors for early-stage substance use that have been identified. A large and growing body of literature shows that substance abuse is the result of the complex interaction of a variety of different factors, which include cognitive, attitudinal, social, personality, pharmacologic, biologic, and developmental factors.[4, 5]

Social Factors

The most powerful influences promoting the initiation of tobacco, alcohol, and drug abuse are social factors, including the behavior and attitudes regarding substance use among significant others such as parents, older siblings, and friends.[6] Studies have found that parents' use of alcohol, tobacco, marijuana, and other illicit drugs, and parental attitudes that are not explicitly against use, are associated with higher levels of use among children and adolescents. High-conflict family interactions, distant or uncommunicative parent-child relationships, inconsistent discipline, and inadequate parenting practices (that is, lack of parental monitoring) have been identified as risk factors for youth substance use. Additional social influences include popular celebrities or media personalities portraying substance use as a way to become popular, sophisticated, and successful, and promote sex appeal and good times.[7] Positive messages about substance use in popular music, movies, and other media are another powerful source of negative social influences that promote and support substance use.[8]

Cognitive and Attitudinal Factors

Individuals who are unaware of the adverse consequences of substance use, and those who have positive attitudes or expectancies regarding use, are more likely to

become substance users compared with those who have more knowledge or more negative attitudes or expectancies.[9, 10] Individuals who believe that substance use is "normal" and that most people smoke, drink, or use drugs, are more likely to be substance users.

Personality Factors

Substance use is associated with a variety of psychological characteristics. Substance users have a lower self-esteem, assertiveness, personal control, and self-efficacy than nonusers. They are also more anxious, impulsive, and rebellious than nonusers.[4] As described in the clinical literature, individuals with specific psychiatric conditions or symptoms (for example, anxiety, depressive symptoms) may attempt to alleviate such feelings by using particular substances. For instance, through experimentation with different substances, highly anxious individuals may find that alcohol or other depressants help them to feel less anxious, and may use those substances as a way to regulate feelings of anxiety. In the scientific literature, this has been referred to as the self-medication hypothesis.[11]

Pharmacologic Factors

Animal research has found that several drugs of abuse (cocaine, amphetamine, morphine, nicotine, and alcohol) affect the brain in a similar way—that is, by increasing strength at excitatory synapses on midbrain dopamine neurons.[12] Additionally, virtually all of these substances produce effects that are highly reinforcing and dependency-producing. For tobacco, alcohol, and most illicit drugs, tolerance develops rapidly, leading to increased dosages and an elevated frequency of use. Once a pattern of dependency has been established, termination produces dysphoric feelings and physical withdrawal symptoms.

Behavioral Factors

Substance use is associated with an assortment of health-compromising or problem behaviors. Individuals using one substance are more likely to use others. Certainly, among young people, substance abuse is often considered to be part of a general syndrome reflecting a particular value orientation.[13] Youth who smoke, drink, or use drugs are likely to get lower grades in school, and generally are not involved in adult-sanctioned activities such as sports and clubs. They are more likely than nonusers to become involved in antisocial or delinquent behavior, aggressiveness, and premature sexual activity.[14, 15] Moreover, it is believed that these problem behaviors stem from the same or highly similar causes, which has significant implications for prevention. Notably, it may be possible to develop a single preventive intervention capable of having a significant effect on several associated behaviors at the same time, thereby eliminating the need for multiple (overlapping) prevention programs.

TYPES OF PREVENTION

The term "prevention" is used to describe a range of intervention activities that are intended to deter or disrupt the development of a disease or disorder. Prevention

can be classified according to the point at which intervention activities occur along the disease-disorder developmental progression or according to the groups to whom the interventions are directed. Consistent with traditional usage in the field of public health, prevention has been divided into primary, secondary, and tertiary prevention. *Primary* prevention is designed to reach individuals before they have developed a specific disorder or disease. As such, primary prevention targets a general population of individuals. For the most part, these individuals have not yet started to use tobacco, alcohol, or other drugs. The goal is to prevent substance use and abuse by intervening on individual or environmental factors viewed as promoting or supporting this type of health-compromising behavior. *Secondary* prevention involves screening and early intervention. *Tertiary* prevention involves preventing the progression of a well-established disorder to the point of disability. Nevertheless, one criticism of this classification system is that it is difficult to distinguish between tertiary prevention and treatment. This is because both involve care for persons with an established disorder.

In 1994, the Institute of Medicine [16] proposed a new framework for classifying intervention programs as part of a continuum of care. While initially proposed as a system to categorize interventions for mental disorders, the framework has been universally adopted and the terminology is now applied to other types of interventions. The continuum of care that was proposed includes prevention, treatment, and maintenance of treatment effects (see figure 17.2). In this framework, prevention is reserved only for interventions occurring before the initial onset of a disorder.

Prevention is additionally divided into three types: universal, selective, and indicated preventive interventions. These classifications define prevention according to the groups to whom the interventions are directed. *Universal* prevention programs focus on the general population. They aim to deter or delay the onset of a condition or risk behavior. Universal programs recognize that all members of a population share some level of risk and can benefit from prevention programs that provide information and skills to help individuals avoid the outcome or condition.

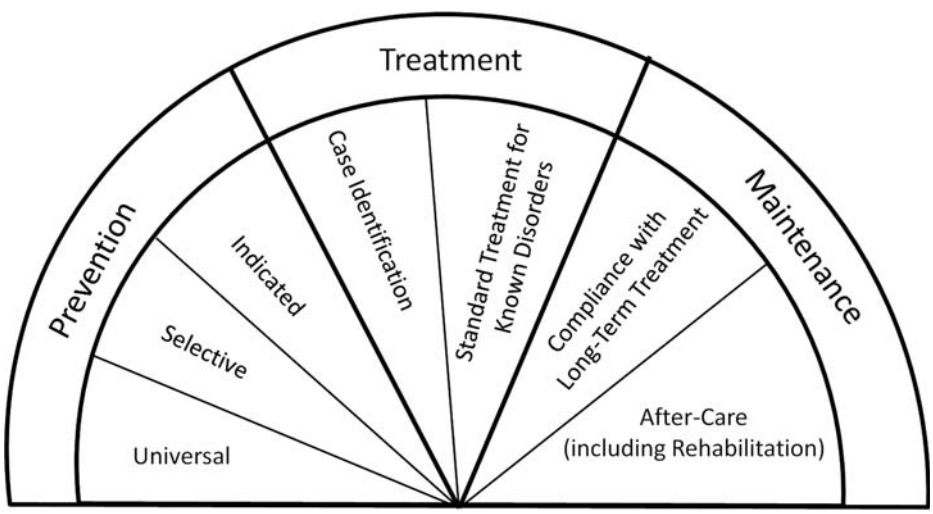

Figure 17.2 Institute of Medicine Continuum of Care Model. (Source: Institute of Medicine, 1994.)

Selected prevention programs target specific high-risk groups or subsets of the general population that are believed to be at high risk because of membership in a particular group. Risk groups for selected interventions may be based on biologic, social, psychological, or other risk factors. For example, selected interventions for drug abuse prevention might recruit groups such as children of drug users or residents of high-risk neighborhoods. An individual's level of risk is believed to be greater than average because of their membership in the selected group. *Indicated* prevention programs are designed for those already engaging in the behavior, showing early danger signs, or who are engaging in related high-risk behaviors. For example, indicated programs for drug-abuse prevention would be appropriate for individuals that recently initiated illicit substance use. The goal would be to reduce their chances of a further escalation in use that can develop into a drug abuse problem. Recruitment and participation in a selective intervention is based on subgroup membership, while recruitment and participation in an indicated intervention is based on early warning signs demonstrated by an individual.

SCHOOL-BASED SUBSTANCE ABUSE PREVENTION

Over the past 30 years, efforts to prevent adolescent use of tobacco, alcohol, and illicit drugs have taken different forms and utilized a range of approaches. These include prevention approaches designed for schools, families, and the larger community. Schools are the most common implementation site for adolescent drug abuse prevention programs. School settings are advantageous because they provide access to large numbers of youth in a structured setting. In addition to their traditional academic mission, schools have increasingly focused on a variety of social and health problems, particularly those that have the potential for undermining the achievement of educational objectives. For example, the U.S. Department of Education has included "drug-free schools" as one of its goals for improving the quality of education in the United States. Many states mandate tobacco, alcohol, or drug education during one or more years of middle or junior high school as well as information related to teenage pregnancy and HIV/AIDS. Although the most common and logical place in the school curriculum to address these issues is in health education, some schools have even integrated elements of prevention into academic subjects such as science or social studies.

Existing prevention programs use one or more of the following five approaches: (1) information dissemination, (2) alternative activities, (3) psychological inoculation, (4) resistance skills training, and (5) competence enhancement. Each of these approaches teach personal coping skills and general social skills. Table 17.1 summarizes these prevention strategies, which are discussed in the following sections below.

Information Dissemination

The most widely used substance abuse prevention approach involves educational efforts to increase knowledge and awareness of the dangers of smoking cigarettes, drinking alcoholic beverages, or using illicit drugs. The primary focus of this approach is to provide factual information about a range of drugs, including their pharmacology, methods of use, classification, prevalence, and the adverse consequences of use. Information dissemination approaches are based on a rational model of human behavior.

Table 17.1.
Overview of Major School-Based Prevention Approaches

Approach	Focus	Methods
Information dissemination	Increase knowledge of drugs, their effects and the consequences of use; promote antidrug use attitudes	Didactic instruction, discussion, audio/video presentations, displays of substances, posters, pamphlets, school assembly programs
Alternatives	Increase self-esteem, self-reliance; provide variable alternatives to drug use; reduce boredom and sense of alienation	Organization of youth centers, recreational activities; participation in community service projects; vocational training
Psychological inoculation	Increase resistance to social influences to smoke, drink, or use drugs through exposure to weaker doses of those influences	Class activities exposing students to appeals by advertisers or peers to engage in tobacco, alcohol, or other drug use; teach counterarguments to those appeals; in-class practice resisting those appeals
Resistance skills	Increase awareness of social influence to smoke, drink, or use drugs; develop skills for resisting substance use influences; increase knowledge of immediate negative consequences; establish nonsubstance use norms	Class discussion; resistance skills training; behavioral rehearsal; extended practice via behavioral "homework"; use of same-age or older peer leaders
Competence enhancement	Increase decision making, personal behavior change, anxiety reduction, communication, social, and assertive skills; application of generic skills to resist substance use influences	Class discussion; cognitive–behavioral skills training (instruction, demonstration, practice, feedback, reinforcement)

Source: Botvin, Griffin, and Murphy

The underlying assumption of this approach is that adolescents begin to use tobacco, alcohol, marijuana, and other drugs because they are unaware of the adverse consequences of using these substances. This approach assumes that providing individuals with factual information about the dangers of using drugs will lead to changes in attitudes, which, in turn, will deter substance use.

Prevention efforts using the information dissemination approach have taken the form of public information campaigns and school-based tobacco, alcohol, and drug education programs. Public information campaigns involve the use of pamphlets, leaflets, posters, and public service announcements to increase public awareness of the problem of tobacco, alcohol, or drug abuse and alter social norms concerning use. School programs involve classroom curricula, assembly programs featuring guest speakers (frequently policemen or health professionals), and educational films.

Studies testing tobacco, alcohol, and drug abuse prevention approaches have consistently found that information dissemination approaches are not effective. They can increase knowledge and, in some instances, change attitudes. However, increasing knowledge has little or no impact on substance use behavior.[17] As knowledge regarding the etiology of substance abuse has grown, it has become increasingly clear that the information dissemination approach is too narrow in its focus to be effective. Although knowledge about the adverse health consequences of substance use is important, it is only one of many factors promoting the initiation of adolescent substance use.

Alternatives

A prevention approach often incorporated into both community-based and school-based interventions involves restructuring part of the adolescents' environment to provide them with alternatives to substance use and activities associated with substance use. The original model for alternatives consists of establishing youth centers to provide teens with a place to go after school where they can participate in a range of activities, such as hobbies, games, sports, and community service. The underlying assumption of this approach is that creating a safe and healthy environment where adolescents could participate in structured activities would not only occupy their time in a constructive way but that participating in these activities could take the place of involvement with drugs.

Outward Bound and similar programs represent another type of alternatives approach. Proponents of these programs argue that they can alter the affective-cognitive state of an individual by promoting teamwork, self-sufficiency, self-confidence, and self-esteem. No evidence exists to support the effectiveness of the alternatives approach. Although some of the activities included in alternatives approaches (such as those related to academics, religion, or physical fitness training) are associated with nonsubstance use,[18] they generally are not appealing to high-risk youth. In addition, activities must be selected with care to avoid increasing risk for substance abuse, because some activities may expose adolescents to environments or settings in which substance use is considered normal and acceptable, such as some social, entertainment, and vocational activities.

Psychological Inoculation

The recognition that social factors play a major role in the onset and early use to tobacco, alcohol, and illicit drugs led to the development of prevention approaches designed to address the social influences to smoke, drink, or use drugs. Evans [19] developed a novel prevention approach containing several components targeting social and psychological factors associated with smoking onset. This and other similar approaches are often referred to as "social influence" approaches.

A central feature of this prevention strategy was based on persuasive communications theory and a concept called "psychological inoculation."[20] Psychological inoculation is analogous to the conventional notion of inoculation used in medicine with respect to the prevention of infectious diseases. Smoking is conceptualized as being the result of social influences (persuasive messages) to smoke either from peers or the media. These can be direct (offers to smoke from other adolescents or cigarette advertising) or indirect (exposure to high-status role models who smoke).

Following the inoculation analogy, social influences to smoke are conceptualized as "germs" that can infect an individual with prosmoking attitudes and norms that in turn can lead to cigarette smoking. To prevent "infection," it necessary to expose an individual to weaker doses of "germs" (social influences to smoke) to produce "antibodies" or resistance to exposure to more virulent social influences to smoke. If adolescents are faced with peer pressure to try cigarettes or if they are exposed to tobacco advertising, for example, they can be forewarned and prepared by exposing them to the kind of persuasive messages that they are likely to hear from peers or see in tobacco ads, thus providing them with counterarguments to those prosmoking appeals and teaching them the skills necessary for resisting these appeals. For example, adolescents can be taught what to say in specific situations to diffuse or negate attempts at peer pressure and be prepared with counterarguments for situations involving older youth posturing and acting "tough" by smoking.

The first study testing this approach [21] used a series of films designed to increase students' awareness of the various social pressures to smoke, which they would likely encounter as they progressed through the critical middle and junior high school years. These films included demonstrations of specific techniques that could be used to effectively resist various pressures to smoke. Other prevention components in the approach developed by Evans included (1) the periodic assessment of smoking rates in the participating schools with feedback to students about those rates, and (2) information about the immediate physiologic effects of smoking. Smoking was assessed by questionnaire on a biweekly basis and saliva samples were collected as an objective measure of smoking status. The rate of smoking in each classroom (which was considerably lower than most adolescents estimated) was publicly announced to correct the misperception that cigarette smoking is a highly normative behavior (that is, that everybody is doing it). This is now referred to as "norm setting" or "normative education." In an evaluation of this approach, students were randomly assigned to (1) a preventive intervention that included psychological inoculation plus periodic monitoring and feedback, (2) monitoring and feedback alone, or (3) a no-treatment control group. Results showed that smoking onset rates in the two prevention conditions were about 50 percent lower than in the control group. The success of this study by Evans provided the first real evidence that preventive interventions could produce effects on substance use behavior and ushered in a new era of prevention research.

Resistance Skills Training

The "resistance skills training" or "refusal skills training" approach is a variation on the above described model. A distinctive feature of this approach is an emphasis on teaching skills for resisting peer and media pressures to smoke, drink, or use drugs. The resistance skills training approach is based on a conceptual model

stressing the importance of social factors in promoting the initiation of substance use among adolescents. These influences come from the family (parents and older siblings), peers, and the mass media.

Social factors influence behavior by interacting with individual characteristics and shaping attitudes, norms, and expectations.[22] New behaviors are learned through a social learning process that involves observation and imitation. Through this social learning process, adolescents may expect various positive outcomes from substance use, such as increased alertness, relief from anxiety, or enhanced social status. While positive expectations (benefits) can be learned from personal experience with substance use, they are initially learned by observing the expectations, attitudes, and behaviors regarding substance use of significant others like parents, siblings, peers, and media personalities. Thus, an adolescent risk for becoming a substance user is increased if family members (parents or older siblings) or friends smoke, drink, or use illicit drugs, as well as through exposure to the mass media, which often glamorizes drug use by celebrities and other high-status role models.

Social resistance skills programs teach adolescents to recognize and avoid "high-risk" situations—that is, situations in which they are likely to be under peer pressure to smoke, drink, or use drugs. Adolescents are also taught how to handle situations involving peer pressure to use drugs, including both what to say (that is, specific refusal statements) and how to say it most effectively. These prevention programs often have included material to combat the perception that substance use is more widespread than it actually is. The tendency to overestimate the prevalence of smoking, drinking, and drug use can be challenged by providing adolescents with the actual prevalence rates of substance use among their peers, by presenting information regarding prevalence rates from national surveys, or by presenting students with information obtained from classroom or school-wide surveys conducted at their own school. Interventions using the resistance skills training approach often have included a component designed to increase students' awareness of the techniques used by advertisers to promote the sale of tobacco products or alcoholic beverages and to teach techniques for formulating counterarguments to the appeals used by advertisers.

The effectiveness of the resistance skills approach has been tested on tobacco, alcohol, and other drugs, typically targeting middle or junior high school students. Relatively few studies have involved elementary school students, partly because substance use is rare in this population. Considerable evidence indicates that well-designed and rigorously implemented interventions using the social resistance skills approach can deter adolescent substance use. In a comprehensive meta-analysis of school-based smoking prevention programs, 23 randomized controlled trials (RCTs) were identified that met rigorous inclusion criteria for research design and execution quality.[17] The main outcome variable was prevalence of nonsmoking at follow-up among those students not smoking at the baseline assessment. Of 13 RCTs testing the social resistance skills approach and using appropriate research designs, nine were found to produce positive intervention effects on cigarette smoking, while four RCTs showed no effects. Overall, this meta-analysis indicated that the social skills training approach produced relatively short-term effects. Findings from a separate review of the smoking prevention literature evaluated studies comparing social influence programs to no-intervention control-group conditions and found that the social influence programs were capable of producing prevention effects lasting for from one to four years.[23] The most effective of these

school-based smoking prevention programs include multiple social influence components (for example, resistance skills and norm setting), booster sessions conducted over two or more years, and reinforcement in the community, including the involvement of parents or the mass media.

Despite the positive short- and intermediate-term results of prevention approaches using one or more components addressing the social influences promoting the use of tobacco, alcohol, and other substances, the results of long-term follow-ups raise questions about the durability and sustainability of these prevention effects. Long-term follow-up studies of school-based social influence approaches indicate that these prevention effects typically are not maintained. Therefore, some have argued that these school-based prevention approaches may not be powerful enough to produce lasting prevention effects.

Others have argued, however, that the apparent failure of studies that test resistance skills training approaches to produce long-term prevention effects may relate either to the type of intervention tested or the way these interventions were implemented. For example, the lack of effects in several long-term follow-up studies may have occurred because (1) the length of the intervention may have been too short (that is, the prevention approach was effective, but the initial prevention "dosage" was too low to produce a long-term effect); (2) booster sessions were either inadequate or not included (that is, the prevention approach was effective, but it eroded over time because of the absence or inadequacy of ongoing intervention); (3) the intervention was not implemented with enough fidelity to the intervention model (that is, the correct prevention approach was used, but it was implemented incompletely, improperly, or both); or (4) the intervention was based on a faulty assumptions, was incomplete, or was otherwise deficient.[24] Although prevention programs using the resistance skills training approach either alone or in combination with other social influence components (such as norm setting or psychological inoculation) are able to produce initial prevention effects, evidence of long-term effectiveness is limited. The extant prevention literature now suggests that to be powerful enough to have a durable impact on adolescent substance use, prevention programs need to focus on a broad and comprehensive set of skills, have a strong initial dosage, include at least two additional years of (booster) intervention, and be implemented with fidelity (adherence) to the underlying prevention model.

Competence Enhancement

The competence enhancement approach is more comprehensive than either traditional information-based health education approaches or the more recent resistance skills approach. The competence enhancement approach emphasizes the acquisition of general personal and social skills, which can be taught either alone [25] or in combination with components of the social resistance skills model.[26] The competence-enhancement approach is based on a theoretical model that posits that youth with poor personal and social skills are more susceptible to influences that promote drug use and also are more likely to use drugs to achieve desired goals (such as popularity or self-esteem) or as a coping strategy for dealing with stress and anxiety.[26] The competence enhancement approach is based on social learning theory [22] and problem behavior theory.[13] Substance abuse is conceptualized as

a socially learned and functional behavior resulting from the interplay of social and personal factors. Substance use behavior is learned through modeling and reinforcement and is influenced by cognitions, attitudes, and beliefs. Prevention programs using the competence enhancement approach teach two or more of the following:

- General problem-solving and decision-making skills
- General cognitive skills for resisting interpersonal or media influences
- Skills for increasing self-control and self-esteem
- Adaptive coping strategies for relieving stress and anxiety through the use of cognitive coping skills or behavioral relaxation techniques
- General social skills
- General assertive skills

Prevention programs using this approach emphasize skills acquisition and use interactive teaching methods. For example, personal self-management and general social skills are taught using instruction, demonstration, feedback, reinforcement, behavioral rehearsal (practice during class), and extended practice through behavioral homework assignments. Unlike the resistance skills training approach that teaches skills with a specific focus on the problem of substance abuse, the competence enhancement approach teaches a more general set of skills intended to be applicable to a wide range of situations. The personal and social skills taught can be used to deal with situations related to substance use and abuse (for example, the use of assertive skills to situations involving peer pressure to smoke, drink, or use drugs). But, these same skills also can be used to deal with many of the other challenges confronting adolescents in their everyday lives. In other words, by teaching a general set of skills, this approach is designed to facilitate the development of general personal and social competence. Although the defining feature of this prevention approach is its emphasis on teaching general personal and social skills, it also incorporates elements of the social influence approaches (particularly resistance skills training and norm setting).

A number of school-based studies have tested the competence enhancement approach ranging from small-scale pilot studies to large-scale randomized trials. These studies have demonstrated significant behavioral effects on alcohol, tobacco, and other drug use. The most extensively researched prevention program using the competence enhancement approach is the *Life Skills Training* (LST) program. The LST program is a universal school-based prevention approach that teaches general personal and social skills training combined with drug-refusal skills and norm-setting activities. The program consists of 12 to 15 classes taught in the first year of middle or junior high school. Some studies have also included "booster sessions" in the second year (10 classes) and third year (5 classes) of middle and junior high school. Studies range from small-scale efficacy studies to large-scale effectiveness trials. These studies consistently have shown that LST produces significant reductions in tobacco, alcohol, and other drug use relative to controls not receiving the program.

Early research with LST tested its efficacy for preventing cigarette smoking and involved predominantly white, middle-class students. Additional studies examined its effectiveness with different delivery formats and different program providers, as well as tested its impact on different substances. These studies found that the prevention

approach was made more effective by the inclusion of booster sessions after the initial year of intervention. Furthermore, LST is effective whether taught by teachers, peer leaders, and health educators, and also can prevent the use of alcohol and marijuana.[27] Additional studies testing LST have focused on its long-term effectiveness in preventing tobacco, alcohol, and marijuana use, its effects on more serious levels of drug involvement, including illicit drug use, its impact on hypothesized mediating variables, and its effectiveness with urban, suburban, and rural youth.

Several large-scale randomized trials on different population groups have demonstrated the long-term effectiveness of the LST program. One study involved nearly 6,000 predominantly (95 percent) white students from 56 junior high schools in New York State. Schools were randomly assigned to prevention and control conditions. Students in the prevention condition received the LST program in the seventh grade and were given booster sessions in the eighth and ninth grades. Students who received the LST program had lower rates of cigarette smoking, alcohol use, and marijuana use than students in the control condition both at the end of the ninth grade [28] and at the end of high school.[29] The effect of the prevention program on illicit drug use was examined using data from a confidential and random subsample of these students who were found to have less illicit drug use (using an index of overall illicit drug use) as well as less use of hallucinogens, heroin, and other narcotics relative to illicit drug use among students in the control group.[30]

A second large-scale prevention trial tested the LST approach with a population of predominantly (98 percent) inner-city minority youth ($N = 3,621$) from 29 middle and junior high schools in New York City. Results at the post-test and one-year follow-up indicated that those who received the prevention program reported less smoking, drinking, drunkenness, inhalant use, and polydrug use (that is, the use of multiple drugs) relative to those in the control group who did not receive the prevention program.[31] A series of analyses using the same data set were conducted to further examine several specific prevention issues to determine the impact of the LST approach on cigarette smoking among adolescent girls, binge drinking (defined as five or more drinks per drinking occasion), and to test its effectiveness on a cohort of high-risk youth (students with poor grades in school, friends who engage in substance use). These analyses showed that the LST competence enhancement approach not only prevented the onset of cigarette smoking, but also reduced the escalation of cigarette smoking by 30 percent among adolescent girls.[32] Other analyses showed that this prevention approach cut binge drinking by 50 percent for up to three years among inner-city boys and girls [33] and also was effective in reducing binge drinking in high-risk youth.[34]

Finally, a randomized trial conducted on 1,677 rural youth from 36 Iowa schools demonstrated the long-term effectiveness of the LST approach.[35] Schools were randomly assigned to the following conditions: (1) LST, (2) LST plus a family-based program, and (3) controls. Long-term follow-up data were collected at the end of the 12th grade, five and a half years after baseline. Results replicated previous research with LST, demonstrating prevention effects for cigarette, alcohol, and marijuana use initiation for both LST conditions compared with controls. Taken together, these findings provide considerable evidence in support of the effectiveness of both the LST program and the competence enhancement approach to substance abuse prevention.

FAMILY-BASED SUBSTANCE ABUSE PREVENTION

Parents play a critical role in child and adolescent development and have a strong influence on the developmental pathways that ultimately contribute to substance use involvement. Parents have great potential to positively influence the decisions that young people make about many important issues, including substance use and abuse. A variety of family-based programs have been designed to prevent adolescent substance use. Programs focusing on preadolescents, for example, typically are designed to provide parents with the skills necessary to understand key precursors that can lead to substance abuse—that is, conduct problems, poor social skills, and school failure. These programs, often implemented with parents of young children without the children present, teach parents ways to help children develop prosocial skills and social resistance skills; ways to reduce aggressive, antisocial, or other unwanted behavior; and ways to nurture, bond, and communicate with children. Because parental monitoring plays a central role in mediating the negative impact of peer influences (which play an increasingly important role as children transition into adolescence) these programs frequently provide instruction to parents on best ways to supervise or monitor their children's day-to-day activities and whereabouts.

Another type of family-based prevention program, typically for older children and adolescents, includes the entire family in the intervention (with or without additional parent-only training). These programs focus on the quality of family relationships, and explore the impact these relationships have on the child or adolescent's behavior. Parents are taught how to correct inappropriate interactions. Interventions that focus on both parenting skills and family bonding appear to be the most effective in reducing or preventing substance use in children.[36]

An example of an evidence-based family prevention program that includes the entire family as part of the intervention is the Creating Lasting Family Connections (CLFC) program. CLFC is a selective intervention designed to prevent substance abuse and violence among adolescents and families in high-risk environments, such as in poor urban communities. The CLFC program promotes healthy beliefs and attitudes that are consistent with staying away from drugs and is designed to enhance family bonding and communication skills among parents and youth. Two or more CLFC facilitators run the weekly parent and youth training sessions over a 20-week period. The program facilitators administer six interactive modules, three to parents and three to youth. Within each module, there are five to six individual sessions each lasting up to 2.5 hours. The sessions focus on topics including substance use issues, personal and family responsibilities, and communication and refusal skills. In addition to the parent and youth training sessions, the CLFC program aims to encourage the use of community services in resolving family problems and addressing youth problem behavior.

Evidence supporting the effectiveness of the CLFC program has shown that when the key resiliency factors targeted by the program improve, the intervention produces behavioral effects. Findings indicated that as family pathology decreased, CLFC reduced the frequency of alcohol and other drug use among youth at the 12-month assessment. In addition, CLFC was found to produce effects on other outcomes, including use of community services and parent knowledge and beliefs about alcohol and other drug use.[37]

A recent systematic review of the family-based prevention literature examined programs to prevent alcohol, tobacco, or drug abuse in children under 18 years of age.[38] Findings from the 20 controlled studies indicated statistically significant reductions in tobacco use in 9 out of 13 studies, reduction in alcohol use in 6 of 14 studies, and reduction in drug use in 5 of 9 studies. However, three interventions produced increases in alcohol, tobacco, or drug use. Overall, the review showed that parenting programs could be effective in reducing or preventing substance use. The most effective drug prevention programs were those combining both active parental involvement and skills development in the areas of parenting, social competence, and self-regulation skills. Little is known about the change processes involved in such interventions; therefore, more research needs to be done to determine the long-term effectiveness of these programs.

To summarize, a variety of parenting skills programs and family-based drug prevention programs are available. Programs focusing on both parenting skills and family bonding appear to be the most effective in reducing or preventing substance use in children. A limitation of family-based prevention is the difficulty in getting parents to participate. Also, families that are at the greatest risk for drug use are the least likely to participate in a drug prevention program.[39]

COMMUNITY-BASED SUBSTANCE ABUSE PREVENTION

Community-based drug abuse prevention programs typically combine school-based programs, family or parenting interventions, mass media campaigns, public policy components (such as restricting youth access to alcohol and tobacco), and other types of community organization and activities. Generally, community-based programs are coordinated by a coalition of stakeholders including parents, educators, and key leaders in the community. An example of an evidence-based community prevention program is Community Trials Intervention to Reduce High-Risk Drinking (RHRD). RHRD is a universal intervention that aims to alter community-wide alcohol use patterns, such as drinking and driving, underage drinking, and binge drinking. The program relies on several environmental intervention strategies to increase community awareness, prevent access to alcohol for underage drinkers, and enforce laws regarding alcohol use and sales. The first of these strategies involves reducing underage alcohol access by training alcohol retailers to not to sell to minors, and pressing for increased enforcement of laws regarding alcohol sales to minors. Furthermore, the RHRD providers work with community leaders to ensure that existing zoning and municipal regulations are enforced regarding the density of bars and liquor stores.

Responsible beverage service is another goal of the RHRD program. The providers train alcohol beverage servers and retailers on policies and procedures to reduce drunkenness and driving after drinking among their customers. Another component provides communities with the tools to form the coalitions needed to implement and support the various intervention components. To properly implement RHRD, project staff must—in consultation with local community organizations, opinion leaders, law enforcement, zoning and planning commissions, policy makers, and the public—assess community priorities and decide which interventions to use and how to adapt them. RHRD training manuals and brochures offer strategies and

tactics for reducing alcohol use within the community. In an evaluation of the RHRD program, alcohol consumption was assessed via telephone surveys to randomly selected individuals in the intervention and comparison communities. When compared with individuals living in comparison communities, those in the intervention communities had significant reductions in the amount of drinking, the rate of driving when having had too much to drink, and the rate of driving when over the legal limit.[40, 41]

A recent *Cochrane Review* survey used a qualitative narrative synthesis to examine 17 community-based programs to prevent smoking initiation in children and adolescents.[42] The findings suggested that multicomponent community-based prevention programs can be effective in preventing adolescent substance use, especially when multiple components focus on a coordinated, comprehensive message. Of course, a limitation of such coordinated interventions with multiple components is the cost. Also, a high degree of coordination is necessary to implement and evaluate the type of comprehensive program most likely to be effective.

YOUTH SMOKING CESSATION

Data from the 2007 Youth Risk Behavior Survey (YRBS) indicated that almost 69 percent of high school students who ever smoked cigarettes daily had tried to quit smoking cigarettes; unfortunately, only 12 percent were successful in quitting, which was defined as not having smoked cigarettes for 30 days prior to completing the survey.[43] According to the Centers for Disease Control and Prevention, a reason why youth may be unsuccessful in quitting is because they do not use evidence-based approaches to assist them in their efforts, and also it may be easier for "sometime" smokers (who smoke on occasion) to quit compared with "regular" smokers (who may smoke daily). For example, youth are more likely to use self-help strategies that are not recommended by the Public Health Service, such as not buying cigarettes, exercising more, or quitting with a friend, as opposed to attending a support group, calling a help line, using nicotine gum, or talking to a counselor or health professional, which are recommended strategies to quit smoking.

Cessation programs geared to adolescent youth range from pharmacological to cognitive-behavioral strategies and motivational techniques to quit smoking. Programs also range from simple to complex interventions consisting of multiple components. Yet, compared to adolescent drug abuse prevention programs, of which there are more than 25 models, there are only two adolescent tobacco-cessation model programs. The limited number of meta-analyses of smoking-cessation programming found that many of the existing cessation programs have had limited effectiveness due to small sample sizes and the lack of control groups for comparison. This may be in part because the number of young people wanting to quit smoking is small, and the programs available are not as widely disseminated as with prevention interventions. Also, most studies did not follow up with participants to determine long-term smoking behavior.

A more recent meta-analysis included 48 studies on youth smoking-cessation programs.[44] As defined by Sussman et al., teen smoking-cessation programming included those programs targeting youth ages 12 to 19 who smoked cigarettes at baseline and who were encouraged to quit smoking. The interventions were broken

down into four categories: motivational-enhanced programs (15 studies), cognitive-behavioral programs (17 studies), and social influence programs (8 studies); only two studies used medical interventions such as nicotine replacement therapy (NRT). Six were categorized as "other" and were not included in the meta-analysis; instead a descriptive analysis was given. Nicotine replacement therapy, which addresses the physical addiction to smoking, was not as effective as the other methods of quitting smoking. Cognitive-behavioral, motivation-theory-related, and social influences programs were all found to be effective, especially when implemented in school settings. The authors concluded that the most effective interventions in a school setting were those that consisted of at least five sessions, and those that were interactive and interesting to the target audience.

A *Cochrane* review of adolescent smoking-cessation programs identified 15 studies, including RCTs, of youth under the age of 20 who were regular tobacco smokers.[45] The review did not include studies that had a follow-up of three months or less, as it would be difficult to determine effectiveness without longer follow-up. Cessation programs included in the review were geared toward individual youth or organizations associated with young people. The authors selected studies that were able to give biochemical verification of cessation with the primary outcome being smoking status six months after baseline. The authors categorized the interventions as psychosocial interventions, pharmacological interventions, and transtheoretical model approach. For this body of research, psychosocial interventions included motivational enhancement (for example, using motivational interviewing techniques in a group quit session) as well as behavioral management (for example, using cognitive behavior techniques, such as strategies for resisting social pressures related to smoking, or enhancing self-efficacy for refusing cigarettes). Pharmacological interventions such as NRT were the same as those used in adult populations. Transtheoretical model approaches include interventions targeting individual participants' stage of change (for example, precontemplation, contemplation, preparations, action, maintenance, or termination) as they related to quitting smoking.

A major limitation in these studies, however, is that no standard definition was used to measure smoking and quitting across different types of interventions. For example, distinctions were not always made between regular smokers and occasional smokers, and quitting was defined in some studies as not having smoked for more than one day, while others use a baseline of 30 days. At this time, the general consensus is that there is insufficient evidence to support implementing any one of the discussed models over another, although they all show some degree of promise.

CONCLUSION

Adolescent substance abuse remains an important public health problem throughout the world. Research in the United States has shown that the prevalence of alcohol, tobacco, and other drug use increases rapidly from early to late adolescence, peaking during the transition to young adulthood. Over the decades, numerous prevention initiatives for adolescents have been developed for use by schools, families, and communities. The most effective approaches target salient risk and protective factors, and are guided by relevant psychosocial theories regarding the etiology of substance use and abuse. The degree of substance use involvement of

any one adolescent is often a function of the negative pro-drug social influences in their environment combined with their individual vulnerabilities to these influences.

Many school-based prevention programs focus on skill-building in the areas of drug resistance and life skills, or address inaccurate beliefs about the high prevalence of substance use. Reviews of the previous school-based prevention research have found that, overall, theory-based programs can reduce smoking and other forms of substance use by adolescents. Furthermore, approaches in preventing cigarette smoking also have been shown to be effective in preventing alcohol use (including binge drinking) and the use of marijuana and other illicit drugs. The most effective school-based prevention programs are those that are interactive, that focus on building skills in drug resistance and general competence, and that are implemented over multiple years. School-based programs that include a substantive community component seem to be more effective than school-only programs.

Family-based prevention programs include training for parents or group interventions for the entire family and focus on improving family functioning, communication, and family policies on substance abuse. Family interventions combining parenting skills and family bonding appear to be the most effective. Community-based drug abuse prevention programs typically combine components of school, family, mass media, public policy, and community organization programs. Community programs are most effective when they present a coordinated, comprehensive message across multiple delivery components.

There are several factors that may reduce the public health impact of effective school, family, and community prevention programs despite the progress made in the field of substance abuse prevention for children and adolescents. Specifically, most schools in the United States do not use evidence-based prevention programs; effective family programs often do not reach the families in greatest need; and community programs require substantial financial and human resources to be implemented effectively. Future research is needed to find ways to effectively disseminate the most promising adolescent substance abuse prevention programs into our schools, families, and communities. Additional research is needed to develop, test, and disseminate more effective youth smoking-cessation programs.

REFERENCES

1. Newcomb MD, Locke T. Health, social, and psychological consequences of drug use and abuse. In: Sloboda Z, ed. *Epidemiology of Drug Abuse.* New York: Springer; 2005:45–59.
2. Johnston LD, O'Malley PM, Bachman JG, Schulenberg JE. *Monitoring the Future National Survey Results on Drug Use, 1975–2009*; vol. 1. *Secondary School Students*. Bethesda, MD: National Institute on Drug Abuse; 2010.
3. Substance Abuse and Mental Health Services Administration. *The NSDUH Report: Misuse of Over-the-Counter Cough and Cold Medications among Persons Aged 12 to 25*. Rockville, MD: Office of Applied Studies; 2008.
4. Swadi H. Individual risk factors for adolescent substance use. *Drug Alcohol Depend*. 1999;55:209–224.
5. Hartel CR, Glantz MD. *Drug Abuse: Origins and Interventions*. Washington, DC: American Psychological Association, 1997.

6. Andrews J, Hops H. The role of peers and family as predictors of drug etiology. In: Scheier LM, ed. *Handbook of Drug Use Etiology*. Washington DC: American Psychological Association, 2010.
7. McCool JP, Cameron LD, Petrie KJ. Adolescent perceptions of smoking imagery in film. *Soc Sci Med*. 2001;52:1577–1587.
8. Snyder LB. Substance use and the media. In: Scheier LM, ed. *Handbook of Drug Use Etiology*. Washington DC: American Psychological Association, 2010.
9. Piko B. Smoking in adolescence: do attitudes matter? *Addict Behav*. 2001;26:201–217.
10. Simons-Morton B, Haynie DL, Crump AD, et al. Expectancies and other psychosocial factors associated with alcohol use among early adolescent boys and girls. *Addict Behav*. 1999;24:229–238.
11. Khantzian EJ. The self-medication hypothesis of substance use disorders: a reconsideration and recent applications. *Harv Rev Psychiatry*. 1997;4:231–244.
12. Saal D, Dong Y, Bonci A, et al. Drugs of abuse and stress trigger a common synaptic adaptation in dopamine neurons. *Neuron*. 2003;37:577–582.
13. Jessor R, Jessor SL. *Problem Behavior and Psychosocial Development: A Longitudinal Study of Youth.* New York: Academic Press, 1977.
14. Donovan JE, Jessor R. Structure of problem behavior in adolescence and young adulthood. *J Consult Clin Psychol*. 1985;53:890–904.
15. Resnicow K, Ross-Gaddy D, Vaughan RD. Structure of problem and positive behaviors in African-American youth. *J Consult Clin Psychol*. 1995;63:594–603.
16. Institute of Medicine. *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research*. Washington, DC: National Academies Press, 1994.
17. Thomas R, Perera R. School-based programs for preventing smoking. *Cochrane Database of Syst Rev*. 2006;(3). doi:10.1002/14651858.CD001293.pub2.
18. Collingwood TR, Sunderlin J, Reynolds R, et al. Physical training as a substance abuse prevention intervention for youth. *J Drug Educ*. 2000;30(4):435–451.
19. Evans RI. Smoking in children: developing a social psychological strategy of deterrence. *Prev Med*. 1976;5:122–127.
20. McGuire WJ. The nature of attitudes and attitude change. In: Lindzey G, Aronson E, eds. *Handbook of Social Psychology*. Reading, MA: Addison-Wesley; 1968:136–314.
21. Evans RI, Rozelle RM, Mittlemark MB, et al. Deterring the onset of smoking in children: knowledge of immediate physiological effects and coping with peer pressure, media pressure, and parent modeling. *J Appl Soc Psychol*. 1978;8:126–135.
22. Bandura A. *Soc Learn Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1977.
23. La Torre G, Chiaradia G, Ricciardi G. School-based smoking prevention in children and adolescents: review of the scientific literature. *J Public Health*. 2005;13:285–290.
24. Resnicow K, Botvin GJ. School-based substance use prevention programs: why do effects decay? *Prev Med*. 1993;22:484–490.
25. Caplan M, Weissberg RP, Grober JS, et al. Social competence promotion with inner-city and suburban young adolescents: effects on social adjustment and alcohol use. *J Consult Clin Psychol*. 1992;60:56–63.
26. Botvin GJ. Preventing drug abuse in schools: social and competence enhancement approaches targeting individual-level etiological factors. *Addict Behav*. 2000;25:887–897.
27. Botvin GJ, Griffin KW. Life Skills Training: empirical findings and future directions. *J Prim Prev*. 2004;25:211–232.

28. Botvin GJ, Baker E, Dusenbury L, et al. Preventing adolescent drug abuse through a multimodal cognitive-behavioral approach: results of a three-year study. *J Consult Clin Psychol*. 1990;58:437–446.
29. Botvin GJ, Baker E, Dusenbury L, et al. Long-term follow-up results of a randomized drug abuse prevention trial in a white middle-class population. *JAMA*. 1995;273:1106–1112.
30. Botvin GJ, Griffin KW, Diaz T, et al. Preventing illicit drug use in adolescents: long-term follow-up data from a randomized control trial of a school population. *Addict Behav*. 2000;5:769–774.
31. Botvin GJ, Griffin KW, Diaz T, et al. Drug abuse prevention among minority adolescents: one-year follow-up of a school-based preventive intervention. *Prev Sci*. 2001;2:1–13.
32. Botvin GJ, Griffin KW, Diaz T, et al. Smoking initiation and escalation in early adolescent girls: one-year follow-up of a school-based prevention intervention for minority youth. *J Am Med Womens Assoc*. 1999;54:139–143.
33. Botvin GJ, Griffin KW, Diaz T, et al. Preventing binge drinking during early adolescence: one- and two-year follow-up of a school-based preventive intervention. *Psychol Addict Behav*. 2001;15:360–365.
34. Griffin KW, Botvin GJ, Nichols TR, et al. Effectiveness of a universal drug abuse prevention approach for youth at high risk for substance use initiation. *Prev Med*. 2003;36:1–7.
35. Spoth RL, Randall G, Trudeau L, Shin C, Redmond C. Substance use outcomes 5 1/2 years past baseline for partnership-based, family school preventive interventions. *Drug Alcohol Dep*. 2008;96:57–68.
36. Lochman JE, van den Steenhoven A. Family-based approaches to substance abuse prevention. *J Prim Prev*. 2002;23:49–114.
37. Johnson K, Bryant DD, Collins DA, Noe TD, Strader TN, Berbaum M. Preventing and reducing alcohol and other drug use among high-risk youths by increasing family resilience. *Soc Work*. 1998;43(4):297–308.
38. Petrie J, Bunn F, Byrne G. Parenting programs for preventing tobacco, alcohol or drugs misuse in children <18: a systematic review. *Health Educ Res*. 2007;22:177–191.
39. Díaz S, Secades-Villa R, Pérez JE, et al. Family predictors of parent participation in an adolescent drug abuse prevention program. *Drug Alcohol Rev*. 2006;25:327–331.
40. Holder HD, Gruenewald PJ, Ponicki WR, et al. Effect of community-based interventions on high-risk drinking and alcohol-related injuries. *JAMA*. 2000;284(18):2341–2347.
41. Treno AJ, Gruenewald PJ, Lee JP, Remer LG. The Sacramento Neighborhood Alcohol Prevention Project: outcomes from a community prevention trial. *J Stud Alcohol Drugs*. 2007;68(2):197–207.
42. Sowden A, Stead L. Community interventions for preventing smoking in young people. *Cochrane Database Syst Rev*. 2003;(1).
43. Centers for Disease Control and Prevention. High school students who tried to quit smoking cigarettes: United States, 2007. *MMWR*. 2009;58(16):428–431.
44. Sussman S, Sun P, Dent CW. A meta-analysis of teen cigarette smoking cessation. *Health Psychol*. 2006;25:549–557.
45. Grimshaw G, Stanton A. Tobacco cessation interventions for young people. *Cochrane Database Syst Rev*. 2006;(4).

CHAPTER 18

Violence as a Public Health Issue

Matthew Robert Anderson, MD, MSc, and
Victor W. Sidel, MD

There have been few years in the past century when the United States has not been involved in a "military action" (the preferred euphemism for war) in far-off countries. Popular news programs nightly bring the ravages of war, criminality, and suicide into our homes. The media, be it television, radio, or newspapers, reports on acts of violence almost on a daily basis. One suspects that few readers of this chapter will not have had a personal experience with violence touching their families. Violence, as defined in *Webster's* dictionary, is the "exertion of physical force so as to injure or abuse"[1] and seems a ubiquitous presence in our world. In this chapter we discuss the health impact of violence and the development of a public health approach to violence, including measures to prevent or reduce the harm of violence.

DEFINITION AND CLASSIFICATION OF VIOLENT ACTS

Over the past decades, the issue of violence has been the subject of numerous reports by private and public organizations. The World Health Organization (WHO) issued a definition and classifications of violence in its 1996 WHO Global Consultation on Violence and Health report defining violence as follows:

> The intentional use of physical force or power, threatened or actual, against oneself, another person, or against a group or community, that either results in or has a high likelihood of resulting in injury, death, psychological harm, maldevelopment or deprivation.[2]

In this chapter, we will follow the 2002 WHO classification of violence [3] that groups violence into three categories based on the perpetrator: (1) *Self-Directed Violence:* "violence a person inflicts on him or herself"; (2) *Interpersonal Violence:* "violence inflicted by another individual or by a small group of individuals"; and, (3) *Collective Violence:* "violence inflicted by larger groups such as states, organized political groups, militia groups and terrorist organizations." The WHO

classification represents one possible topology of violent acts, but the three categories do not necessarily function well as conceptual entities. Interpersonal violence, for example, includes violence within families (such as child abuse) as well as violence committed between strangers (for example, acts committed in the course of a crime). Some acts of violence may not be easily classified within this typology. Is a rape committed by a soldier an act of interpersonal violence, collective violence, or both? How does one classify the suicide of a soldier? Finally, the classification scheme does not imply common approaches—that is, an intervention that reduces gang violence may not affect the risk of war. That being said, the WHO classification is probably the best available classification of violence at this point in time.

HEALTH CONSEQUENCES OF VIOLENCE

The WHO estimates that 1.6 million people die each year from violence.[3] Suicide accounts for 54 percent of these deaths, homicide for 35 percent, and war and other armed conflict for 11 percent. The burden of these deaths falls primarily among young people. Among those ages 15 to 44, violence is responsible for 14 percent of male deaths and 7 percent of female deaths. In the United States, based on data from the National Violent Death Reporting System, the Centers for Disease Control and Prevention (CDC) estimates that there are about 50,000 violent deaths each year.[4] In agreement with international statistics on the majority of these deaths, in 2006, 33,000 were from suicide and 18,000 were due to homicide. Again, the burden of death falls primarily on the young.

Most violence occurs among poor populations, in the United States as well as in other countries. While it is true that 91 percent of deaths due to violence take place in low- and middle-income countries, this may be due to the large number of people living in poverty or near-poverty in these countries.[5] More data on the national, gender, age, and social class rates of violent death are needed to understand the issue more clearly.

Violence is associated not just with fatalities, but also with substantial morbidity. An estimated 2.2 million violence-related injuries occur yearly in the United States.[4] To consider just one example, a survivor of rape may have suffered physical injuries; may have been infected with various sexually transmitted pathogens, including HIV; may find herself with an unwanted pregnancy; and likely will suffer considerable psychological sequelae from the event. It is sobering, therefore, to consider that one in six U.S. women report being the victim of a rape or a rape attempt.[6]

Corso et al. estimated that in 2006 the health costs of violence (both fatal and nonfatal) in the United States were more than $70 billion. Of this expenditure, $5.6 billion was for medical care and $64.7 billion was due to lost productivity.[7] The authors mentioned that, because of the underreporting of injuries, the true economic costs were likely to be much higher.

Violence and war contribute significantly to economic disruption cost and social instability, thus compromising the foundations upon which the public's health is built. For example, the cost of war and preparation for war drain both human and economic resources from needed human and social services, and companies in industrial nations are less likely to invest in countries with high levels of violence, thus draining them of resources as well.

DEVELOPING A PUBLIC HEALTH APPROACH TO VIOLENCE PREVENTION

Two official public health bodies provide guidance on violence prevention. Within the United States, violence prevention is part of the CDC's National Center for Injury Prevention and Control. Internationally, the WHO's Program on Violence and Injury Prevention and Disability (VIP) works in coordination with the Global Campaign for Violence Prevention to promote research and evidence-based violence prevention programs.

Historically, during the 1970s, violence began to be conceptualized and addressed as a public health problem. This approach has been adopted and promoted by both the CDC and the WHO primarily based on evidence showing that public health programs have been successful in reducing certain forms of violence. In 1979, the U.S. surgeon general's *Healthy People* report identified the control of stress as 1 of 15 priority areas for action during the 1980s; both suicide and homicide were seen as preventable effects of failure to cope with stress.[8] The priority areas were later linked with measurable objectives: reductions in injuries and death due to child abuse, decreases in homicides among 15- to 24-year-old black males, decreases in suicides by 15- to 24-year-olds, decreases in the number of privately held handguns, and better data collection regarding child abuse and domestic violence.[9]

The surgeon general's report initiated a period of intense public health activity in the field of violence. By 1983, the CDC had established a Violence Epidemiology Branch. Two years later, researchers from this branch investigated two clusters of teen suicides in Texas, applying a case-control methodology in which each suicide victim was paired with three matched controls. The investigators found that suicide victims were no more likely to have been exposed to suicide in a friend or in the media, but they were more likely than controls to have threatened suicide or physically hurt themselves. This study is considered "the first demonstrated use of field epidemiological techniques to identify suicide clusters."[10] In 1985, the surgeon general organized a Workshop on Violence and Public Health, and the following year the CDC established a Division of Injury Epidemiology and Control, now the National Center for Injury Prevention and Control.[11]

The increasing public health attention to violence was recognized internationally by the WHO. The 49th World Health Assembly, held in 1996, adopted Resolution WHA49.25, "Prevention of Violence: A Public Health Priority." This resolution declared that "violence is a leading worldwide public health problem" and called upon the WHO director general "to initiate public health activities to address the problem of violence."[12] That same year, WHO organized a Global Consultation on Violence and Health and issued a report, *Violence: A Public Health Priority*.[2] The message inherent in these reports is that violence is pervasive and, in many societies, is so dominant that it thwarts efforts of economic and social development. The reports stressed the importance of primary prevention in an effort to root out violence.

The WHO continued to play a leading international role in public health efforts to prevent violence. It organized the 2001 Conference on Small Arms and Light Weapons that resulted in the United Nations Program of Action to Prevent, Combat,

and Eradicate the Illicit Trade in Small Arms and Light Weapons in All Aspects. For the conference, the WHO prepared a report entitled *Small Arms and Global Health,* which stated "Violence is . . . an important health problem—and one that is largely preventable. Public health approaches have much to contribute to solving it."[13] Recommendations for action at the local, national, and international levels provided guidelines to address the issue.

In 2002, WHO published its landmark *World Report on Violence and Health*.[3] This report justified a public health approach to the question of violence on a pragmatic basis:

> Violence can be prevented and its impact reduced, in the same way that public health efforts have prevented and reduced pregnancy-related complications, workplace injuries, infectious diseases, and illness resulting from contaminated food and water in many parts of the world. The factors that contribute to violent responses—whether they are factors of attitude and behaviour or related to larger social, economic, political and cultural conditions—can be changed.[3]

More recently, the WHO Commission on the Social Determinants of Health has reemphasized social approaches to violence prevention, which it summarized in these terms:

> Ensuring that all groups in society live in safety and are secure from crime and violence poses a major societal challenge. Reducing the prevalence of violent behaviour involves integrated strategies that target key domains for violence prevention such as nurturing and safe relationships between children and parents; reducing violence in the home; reducing access to alcohol, drugs, and lethal means; enhancing the life skills and opportunities of children and youth; and improving criminal justice and social welfare systems. Newer approaches to violence prevention include regulatory control—including alcohol sales designed so that harmful drinking is reduced—conflict transformation, [and] crime prevention through social capital. The Commission points to the need for national and local government to invest in street lighting, early closing of nightclubs and bars, gun control, establishment of neighbourhood watch initiatives, and educational and recreational activities (including job training opportunities).[14]

Also in 2002, the WHO launched a Global Campaign for Violence Prevention to support science-based efforts to define the causes of violence and to design and deliver prevention programs to reduce these causes. The WHO Global Campaign for Violence Prevention has cosponsored a Violence Prevention and Resources Web site run by the Centre for Public Health, Liverpool John Moores University. This site provides abstracts of systematically reviewed studies on violence prevention.

A PUBLIC HEALTH APPROACH TO VIOLENCE

Following the CDC's original study of teenage suicide clusters, the public health approach to violence has adopted a traditional public health model. This follows a methodology involving (1) problem identification, (2) identification of risk and protective factors, (3) the development and testing of prevention strategies, and (4) the assurance of widespread adoption (see figure 18.1).

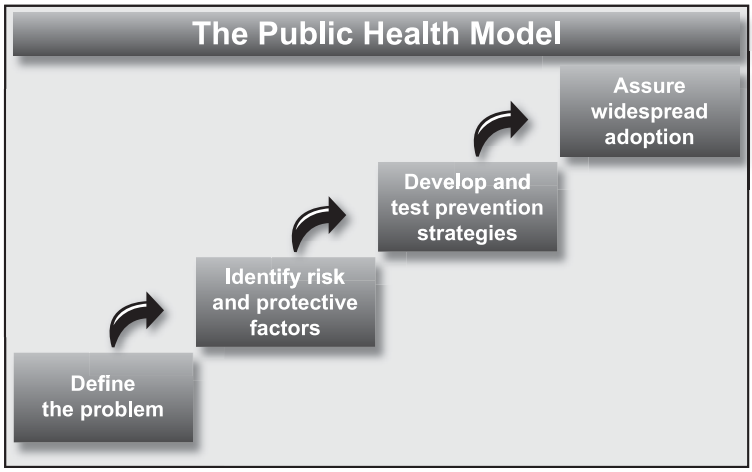

Figure 18.1 The Public Health Model.

The public health approach to violence strives to be scientifically based and rigorously apolitical. While recognizing its value, this approach has been subject to several critiques. First, it ignores the human rights dimension of violence, which recognizes that violence is wrong intrinsically and not just because it harms health.[15] Second, there may be larger, societal-level determinants of violent actions that are not captured by the "risk factor" approach. A review of the public health literature on violence noted that it primarily advocated individual-level and public health approaches to violence, thus implicitly deemphasizing the role of social causes and human rights.[16, 17]

Within the conceptual framework of a public health, there is a need for data collection and evaluation. Indeed, prevention programs must be assessed and tested for efficacy and impact. Sources of data include medical records (hospital, emergency room, outpatient clinics); vital statistics (death certificates); police, judiciary, and prison records; media reports; pertinent government or legislative documents; and local surveys. The hospital emergency room, in particular, often is an important source of data regarding violence. Yet, in the words of the WHO Global Campaign Report of 2007, currently there is "Plenty of activity, not enough evaluation."[18]

Data collection serves a number of important functions:

- It documents problems (such as violence against women and children), bringing such issues to public attention and (hopefully) on a political agenda.
- Good documentation allows for a better understanding of the scope of the problems and should be the first step in a research program to examine the causes of violence.
- Evaluation programs allow for more accurate assessment of the impact of violence prevention programs.

In summary, although national and a global responses to violence prevention have been acknowledged, and a myriad of reports have been issued and conferences held, we are far from achieving a goal of a violence-free society. Violence persists despite the tools and knowledge to make a difference. Perhaps part of the problem

is the multifaceted face of violence. Violence is not evenly distributed across population groups or settings; its form is different depending on the mechanism. The following discusses specific types of violence in an effort to better understand these complex issues.

PREVENTION OF GUN VIOLENCE

Firearms play an important role in all forms of violence. Both nationally and internationally the prevention of gun violence is seen as a priority. Globally, more than 200,000 people are killed each year by small arms and light weapons, about 30,000 annually in the United States.[19, 20, 21] Guns play a particular role in the United States, where more than 200 million firearms, including 65 million handguns, are registered. Americans, perhaps more so than in other societies, are enamored with guns of all types. Gun organizations, such as the National Rifle Association, are powerful lobby groups that are in business to uphold the Second Amendment, the right to bear arms. Perhaps because guns are ubiquitous in U.S. society, firearms of all types are used frequently during violent acts. In 1998, 57 percent of suicides were committed with firearms, 87 percent of which are committed by men.[22] A telephone survey of 800 gun-owning households conducted in 1994 found that 21 percent stored a loaded and unlocked gun. A total of 56 percent of gun owners had received formal training in the proper use and handling of guns. Those who had received training in the proper use and handling of guns were as likely to keep a gun unlocked and loaded as those who had not.[23]

How, then, can we address reducing gun violence? While the 2004 WHO report on violence emphasized the importance of decreasing firearm availability, this is not an easy task to accomplish. Restricting the import of guns is problematic as the gun trade is highly profitable and military weapons are readily available in many countries. Gun buyback programs are commonly used to remove street guns, but, as noted in the surgeon general's Youth Violence Report, buybacks are "a particularly expensive strategy, [that] have consistently been shown to have no effect on gun violence, including firearm-related homicide and injury."[24] The United States has relied heavily on the regulation of firearm usage. Attempts to prevent gun violence by laws prohibiting carrying concealed weapons have been enacted in almost every state, but the laws are uneven and generally ineffective.

Concern over gun violence has led to suggestions that individual clinicians screen for gun ownership and then provide counseling. For example, it might be useful for clinicians who are concerned that a patient is suicidal to use not only their standard clinical strategies to assess and manage suicidal risk, but also go one step further and talk about whether firearms and other lethal means are available in the home. The mnemonic GUNS has been proposed as a screening method: Is there a **G**un in your home? [One in every four homes in the United States contains a handgun]; Are you around **U**sers of alcohol or other drugs, or is there a history of mental illness or violence in the family? Do you feel the **N**eed to protect yourself? Do you have **S**chool-age children or adolescents in your home?[25] For individuals who screen positive, clinicians should note that unintentional shooting, criminal assault or homicide, or committed or attempted suicide are much more likely uses of the gun in the home than for protection. Patients with guns are advised to store them

unloaded, in an impenetrable container, and with a trigger lock. Ammunition should be stored and locked separately.[26]

A report by the Center for Substance Abuse Treatment in 2008 discusses the relationship of substance abuse to suicide. The call to Action to Prevent Suicide issued by the surgeon general in 1999 recognized suicide as an issue of public health concern and initiated national action. Since then progress has been made, but much more needs to be done to recognize the role of substance abuse in suicide and to develop programs for prevention.[27]

PREVENTION OF SELF-DIRECTED VIOLENCE

Self-directed violence includes both self-abuse and suicide. Self-abuse includes self-mutilation and self-injury. Modalities include cutting, scratching, burning, infecting oneself, bruising or breaking bones, inserting objects in body openings, and other forms of bodily harm. An estimated 2 million people in the United States suffer from the effects of self-inflicted harm, the majority of whom are women between the ages of 13 and 30 years.[28] Among those between the ages of 15 to 44 years old, self-inflicted injuries are the fourth leading cause of death and the sixth leading cause of ill-health and disability.[29]

Suicide is the 13th leading cause of death worldwide. In 2000, an estimated 815,000 people died from suicide. Perhaps not surprisingly, alcohol plays an important role in the epidemiology of suicides. In 2009, the CDC reported on a large sample of suicide victims whose blood alcohol levels were measured postmortem. Findings showed that one in four had been legally drunk, with a blood alcohol content at or above the federal standard of 0.08, or 8 milligrams of alcohol per 100 milliliters of blood. Alcohol, the author of the report commented, "can enhance feelings of hopelessness and depression. Alcohol impairs judgment and can lead to much more impulsive behavior. . . . Any suicide prevention efforts must take that into account and address alcohol and substance abuse as well."[30]

In 1998, firearms were the leading method of committing suicide for both men and women in the United States. Rates of firearm suicide have changed little over the past two decades and consistently have exceeded rates of firearm homicide. The firearm suicide rate among men is approximately six times that of women. While firearm suicide rates are highest among the elderly, the majority (66 percent) of firearm suicides are among persons under 55 years of age. Firearm suicide rates among women of all ages have dropped modestly, whereas rates among elderly men have risen considerably. Whites have roughly twice the rate of firearm suicide as do blacks and other race or ethnicity groups. Individual-level empirical studies consistently have indicated that keeping firearms in the home is associated with an increased risk of suicide.

The U.S. Army reported in January 2009 that the suicide rate among U.S. soldiers in 2008 rose for the fourth year in a row, reaching the highest level in nearly three decades. The U.S. Army's suicide rate surpassed that of civilians for the first time since the Vietnam War. Although the suicide rate in all four services were higher than the national civilian rates, they were particularly higher in the Army and the Marines.[30, 31] This is one of the many psychological manifestations of the war on U.S. soldiers.[32]

It should go without saying that every individual who succeeds in killing himself or herself leaves behind family and friends whose lives are often profoundly affected emotionally, socially, and economically.

INTERPERSONAL VIOLENCE AND ITS PREVENTION

Interpersonal violence spans a broad sweep of actions that include various forms of family or intimate partner violence, child and elder abuse, and spousal battering.[33, 34] The term also includes violence between acquaintances and strangers, such as violence during arguments and criminal acts such as robberies. Intimate partner violence (IPV) is especially noteworthy to discuss. IPV in the form of sexual abuse includes unwanted pregnancy, sexually transmitted infections and infertility, and posttraumatic stress disorder (such as anxiety, suicidal tendencies, depression, and eating and sleeping problems). IPV also has been associated with the development of high blood pressure, heart disease, and stroke. Beatings and other forms of physical violence against a partner leave more than the physical bruises and scars; the mental scars fade less quickly.

Each year, women experience about 4.8 million intimate partner–related physical assaults and rapes.[35] Men are the victims of about 2.9 million intimate partner–related physical assaults. IPV resulted in 1,544 deaths in 2004. Of these deaths, 25 percent were males and 75 percent were females. Intimate partner violence accounts for approximately 40 percent of the women who are killed in each year. While the human cost is incalculable, the economic cost of IPV is huge: an estimated $8.3 billion (in 2003 dollars).[36] This cost includes medical care, mental health services, and lost productivity (for example, time away from work). Physical injuries often are serious and result in disability or even death. In New York City, approximately 4,000 women and 900 men are treated in the city's emergency rooms for partner violence every year.

Abuse of children or of older people (elder abuse) encompasses many forms, including neglect or exploitation. The abuse may be intentional or unintentional; it may be physical, psychological (emotional or verbal aggression or abuse), or economic (involving financial misappropriation). Regardless of the type of abuse, the result usually is unnecessary pain, suffering, injury, loss of violation of human rights, and decreased quality of life. Whether the act is characterized as being abusive, neglectful, or exploitative depends on the cultural context, the frequency of the mistreatment, and its duration, severity, and consequences.[3]

The 2002 WHO World Report on Violence and Health [3] discusses a wide range of interventions across the life cycle that have been attempted to reduce interpersonal violence. The WHO report offers a variety of interventions targeted at children from infancy to age 11. These interventions encompass a variety of methodologies, ranging from home visitation after birth to "deconcentrating poverty."

PREVENTION OF COLLECTIVE VIOLENCE

Collective violence includes gang and mob violence, war, and terrorism and is often intended to further the aims of a group or cause. These acts involve the use of

torture and rape and indiscriminate and purposeful attacks on civilian populations. Extreme versions may include "ethnic cleansing" and genocide. The consequences extend to the families and the communities of the victims. Gang and youth violence is especially prevalent in the United States and in other countries. Over the last 20 years in the city of Los Angeles, at least 10,000 African American and Latino young men have died as a result of gang activity and gang warfare.[37] As large as that figure is, it is only a portion of the deaths attributable to gang violence in the United States. Although several studies have been conducted on the nature of gangs and gang warfare, few have examined the various reasons that teenagers and children join gangs. Studies on gangs and gang violence have examined the question: What causes a teenager or child to join a gang? Researchers have found that entrance into a gang starts between the ages of 9 and 11 years old.[38, 39, 40] These same studies also pointed to a need for children and teenagers to feel that they "belong." Membership in a gang fulfills this need, especially when gang members belong to families that neglect, abuse, or ignore them. In particular, it seemed that those least at risk for joining a gang had two parents living at home who consistently were involved in their family, church, and community. Additionally, being a victim of bullying, which often leads to being a bully, was an important factor for a child or teen to decide to join a gang.

The prevention of youth violence is a focus area of the CDC, which has funded a variety of programs to evaluate the extent of youth violence, examine its characteristics, and run pilot prevention trials.[41] A CDC-sponsored review of the literature published in *Morbidity and Mortality Weekly Report* in 2007 concluded that there was "strong evidence that universal school-based [antiviolence] programs decrease rates of violence and aggressive behavior among school-aged children. Program effects were demonstrated at all grade levels. An independent meta-analysis of school-based programs confirmed and supplemented these findings."[42]

The Blueprints Program at the University of Colorado conducts rigorous methodological evaluation of youth violence prevention programs. It has identified 11 model programs, descriptions of which are available on its Web site.[43] Among these programs is the Big Brothers Big Sisters of America, a mentoring program targeted at children 6 to 18 years old from single-parent homes.[44] Another successful program is the Nurse-Family Partnership, which is based on intensive home visits by a nurse to young low-income women in their first pregnancy. The visits start during the pregnancy and continue until the baby is 2 years old.[45] An article in the *New Yorker* describes an innovative approach to prevention of gang violence in Cincinnati.[46] The Einstein Youth Violence Project was a primary-prevention antiviolence program involving fourth- and fifth-graders. An afterschool program for fourth graders focuses on important violence-prevention concepts, such as conflict resolution, communication skills, and awareness of risk factors. Those who complete the afterschool program are enrolled in a mentorship program in which a medical student serves as mentor.

Terrorism adversely affects health in many ways, including causing injury, illness, and death; creating fear, anxiety, and other psychological reactions; destroying the physical infrastructure and social fabric of communities; and causing profound, adverse economic and political impacts on individuals, communities, nations, and the global society. Some responses to terrorism, however, can be harmful as well.

Vengeful responses to terrorist acts or threats may hurt innocent people. Attempts to locate, interrogate, and punish suspected terrorists may threaten civil liberties domestically and international justice abroad. And the U.S. government's diversion of resources from essential public health programs for homeland defense or the War on Terrorism may lead to a worsening of health problems in the United States and to a reduction in U.S. health-program assistance to developing countries.

Recommendations for preventing acts of terrorism and their adverse public health consequences include the following:

1. Focusing on poverty, social injustice, and health disparities that may contribute to the development of terrorism;
2. Providing humanitarian assistance to, as well as protecting the human rights of, civilian populations directly or indirectly affected by terrorism;
3. Advocating for the speedy end of the armed conflict in Afghanistan and promoting nonviolent means of conflict resolution;
4. Strengthening the public health infrastructure (workforce, laboratory, and information systems) and other components of the public health system (including education, research, and the faith community) to increase the ability to identify, respond to, and prevent problems of public health importance, including to heal the effects of terrorist attacks;
5. Ensuring the availability of, and accessibility to, health care—including medications and vaccines—for individuals exposed, infected, made ill, or injured in terrorist attacks;
6. Educating and informing health professionals and the public to better identify, respond to, and prevent the health consequences of terrorism, and promote the visibility and availability of health professionals in the communities that they serve;
7. Addressing the mental health needs of populations directly or indirectly affected by terrorism;
8. Assuring the protection of the environment, the food and water supply, and the health and safety of rescue and recovery professionals;
9. Assuring the clarification of the roles, relationships, and responsibilities among public health agencies, law enforcement, and first responders;
10. Working towards preventing hate crimes and ethnic, racial, and religious discrimination, including profiling; promoting cultural competence, diversity training, and dialogue among people; and protecting human rights and civil liberties;
11. Advocating the immediate control and ultimate elimination of biological, chemical, and nuclear weapons; and,
12. Building and sustaining the public health capacity to develop systems, which includes the collecting data on the health and mental health consequences of terrorism and other disasters on victims, responders, and communities; developing uniform definitions and standardized data-classification systems of death and injury resulting from terrorism and other disasters.[47]

According to Levy and Sidel, war accounts for more deaths and disability than many major diseases combined. It destroys families, communities, and sometimes whole cultures. It directs scarce resources away from protection and promotion of health, medical care, and other human services. It destroys the infrastructure that supports health. It limits human rights and contributes to social injustice. It leads many people to think that violence is the only way to resolve

conflicts—a mindset that contributes to domestic violence, street crime, and other kinds of violence. And it contributes to the destruction of the environment and overuse of nonrenewable resources. In sum, war threatens much of the fabric of our civilization.[48]

It is estimated that 200,000,000 people died as a result of war during the 20th century. Sadly, civilians represent an increasing percentage of war-related deaths.[49] The field of public health can provide needed guidance in efforts to stem or to prevent conflicts that lead to war. Public health workers should be encouraged to investigate and research causes of conflict in their communities in an effort to better understand how conflict can be resolved before it becomes violent. They should establish partnerships with medical and community-based organizations to build trust and understanding and to provide medical and public health services to community members affected by conflict.

Many public health workers have little experience with peace-building in their own institutions or in their own communities. Peace-building requires mediation and, when necessary, arbitration to address conflicts before they become irreconcilable. This problem was documented in a study of medical students' knowledge of the Geneva Conventions.[50] Only 37 percent knew that the Geneva Conventions apply whether or not war had formally been declared; one-third were unaware that the Geneva Conventions state that physicians should treat the sickest first, regardless of nationality; 37 percent did not know that the Geneva Conventions prohibit threatening or demeaning prisoners or depriving them of food or water for any length of time. Before public health or medical professionals can offer constructive ways to alleviate conflict, much more needs to be done to increase awareness and to provide the skills necessary for each to assume leadership roles in this area.

Events in the 21st century illustrate the urgency to control weapons used in collective violence. "Weapons of mass destruction" include chemical weapons, biological weapons, and nuclear weapons. Chemical agents have been used for thousands of years. They are compounds designed to kill or disable people through their toxic or poisonous effects. They are relatively simple to make and use. Their effects are often dramatic and immediate. Both combatant and noncombatant populations can be the targets of these weapons. Environmentally persistent chemical agents also can be used to deny terrain or to contaminate food and water. Like nuclear and biological weapons, chemical weapons have psychological, political, operational, and strategic impacts.[51]

Biological weapons have been used in warfare, although infrequently, since ancient times. These agents are feared because they are generally invisible and easy to disseminate, some may spread easily from person to person, and some can cause horrific diseases. The public and health professionals have focused much attention on these agents since the dissemination of anthrax spores through the U.S. mail in September and October 2001. In addition, tens of thousands of postal, news media, and other workers received prophylactic and antibiotic treatment. And millions more feared that they too could be at risk of developing anthrax. The biological agents in biological weapons are living organisms (usually microorganisms) or their toxic products. Although the main targets of these weapons are people, they also can be used against animals or plants to limit human food supplies or agricultural resources and thereby adversely affect human health and well-being indirectly.[52]

SO WHAT CAN PUBLIC HEALTH TRULY DO TO STEM VIOLENCE?

Violence is a complex problem shaped by a multitude of forces within families and communities, and it transcends national boundaries. Looking back, the 20th century will be remembered as a century marked by violence, a legacy of mass destruction of peoples and places (the Holocaust, for example), violence inflicted on a scale never seen before (Hiroshima, for example), and a hideous amount of killings and maiming against those different from others either because their skin color was different, their religion was different, or because of their gender. The 21st century seems to be perpetuating the legacy of the previous century. No country is immune, but as Nelson Mandela has said, neither are we powerless against it.

The public health approach has brought attention to the important impact of violence on the health and well-being of society. Violence prevention always occurs within a political context that cannot be ignored. In its strivings to be "apolitical," the public health model may not address the larger social context within which violent acts occur. The idea that social policies will be enacted simply because they are based on scientific evidence is naïve and perhaps even a bit dangerous. The run up to the Iraq invasion in 2003 displayed how policy makers in the Bush administration either ignored or fabricated evidence to justify a violent and destructive war.

The goals and disciplines of public health and the training and work of public health workers may be in a unique position to deal with violence.[53] The goals of public health—to prevent disease, injury, and premature death and to promote healthy living conditions for all—are identical to the goals of violence prevention. The disciplines and methods of public health—including analyzing the causes of diseases, injuries, and premature deaths, examining the causes of poverty and despair, and determining methods to counter them—can strengthen efforts to prevent violence. And the ability of public health workers to gain trust—both nationally and internationally—can bring new skills and vigor to violence prevention.

REFERENCES

1. Merriam-Webster's Online Dictionary. http://www.merriam-webster.com/dictionary/violence. Accessed May 30, 2009.
2. World Health Organization, Global Consultation on Violence and Health. *Violence: A Public Health Priority*. Geneva, Switzerland: World Health Organization; 1996.
3. World Health Organization. World report on violence and health. http://www.who.int/violence_injury_prevention/violence/world_report/en/index.html. Accessed September 26, 2009.
4. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based injury statistics query and reporting system (WISQARS). http://www.cdc.gov/ncipc/wisqars. Accessed September 26, 2009.
5. Mercy JA, Forjuoh SN. Editorial: violence prevention in an interconnected world. *Int J Inj Contr Saf Promot*. 2008;15(4):175–176.
6. Tjaden P, Thoennes N. *Extent, Nature, and Consequences of Intimate Partner Violence: Findings from the National Violence against Women Survey*. Washington, DC: US Department of Justice; 2000. http://www.ojp.usdoj.gov/nij/pubs-sum/181867.htm. Publication No.NCJ 181867. Accessed September 26, 2009.
7. Corso PS, Mercy, JA, Simon TR, et al., Medical cost and productivity loss interpersonal violence and self-directed violence. *Am J Prev Med*. 2007;32(6):474–482.

8. Public Health Service. *Healthy People: The Surgeon General's Report on Health Promotion and Disease Prevention*. Washington, DC: US Department of Health, Education, and Welfare, Public Health Service; 1979. http://profiles.nlm.nih.gov/NN/B/B/G/K/_/nnbbgk.pdf. DHEW publication no.(PHS)79-5507. Accessed September 26, 2009.
9. Public Health Service. *Promoting Health/Preventing Disease: Objectives for the Nation*. Washington, DC: US Department of Health and Human Services, Public Health Service; 1980. http://tobaccodocuments.org/ti/TIMN0364878-5080.html. Accessed September 26, 2009.
10. Davidson LE, Rosenberg ML, Mercy JA, Franklin J, Simmons JT. An epidemiologic study of risk factors in two teenage suicide clusters. *JAMA*. 1989;262(19):2687–2692.
11. Centers for Disease Control and Prevention. Violence prevention: a timeline of violence as a public health issue. http://www.cdc.gov/violenceprevention/overview/timeline.html. Accessed September 26, 2009.
12. Forty-ninth World Health Assembly. WHA49.25 Prevention of violence: a public health priority; May 20–25, 1996; Geneva, Switzerland. http://www.who.int/violence_injury_prevention/resources/publications/en/WHA4925_eng.pdf. Accessed September 26, 2009.
13. World Health Organization. *Small Arms and Global Health*. Geneva, Switzerland: World Health Organization; 2001. http://whqlibdoc.who.int/hq/2001/WHO_NMH_VIP_01.1.pdf. Accessed September 26, 2009.
14. Commission on the Social Determinants of Health. *Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health*. Geneva, Switzerland: World Health Organization; 2008. http://whqlibdoc.who.int/publications/2008/9789241563703_eng.pdf. Accessed September 26, 2009.
15. Heath I. Treating violence as a public health problem: the approach has advantages but diminishes the human rights perspective. *BMJ*. 2002;325:726–727.
16. Winett LB. Constructing violence as a public health problem. *Public Health Rep*. 1998;113(6):498–507.
17. Mann JM, Gostin L, Gruskin S, Brennan T, Lazzarini Z, Fineberg HV. Health and human rights. *Health Hum Rights*. 1994;1:6–23.
18. World Health Organization. *Third Milestones of a Global Campaign for Violence Prevention Report 2007*. Geneva, Switzerland: World Health Organization; 2007. http://whqlibdoc.who.int/publications/2007/9789241595476_eng.pdf. Accessed September 26, 2009.
19. Cukier W, Sidel VW. *The Global Gun Epidemic: From Saturday Night Specials to AK-47s*. Westport, CT: Praeger Security International; 2006:310.
20. Brady Campaign to Prevent Gun Violence. Firearms facts. http://www.bradycampaign.org/facts/factsheet/pdf/firearm_facts.pdfs. Accessed September 26, 2009.
21. Hemenway, D. *Private Guns, Public Health*. Ann Arbor: University of Michigan Press; 2004.
22. Romero MP. The epidemiology of firearms suicide in the United States. *J Urban Health*. 2002;79(1):39–48.
23. Hemenway D, Solnick SJ, Azrael DR. Firearm training and storage. *JAMA*. 1995; 273(1):46–50.
24. Surgeon General of the United States. *Youth Violence: A Report of the Surgeon General*. Washington, DC: Department of Health and Human Services; 2001. http://www.surgeongeneral.gov/library/youthviolence/chapter5/sec5.html. Accessed October 12, 2009.
25. GUNS acronym. *HELP Network of Concerned Professionals*. Chicago, IL: Children's Memorial Medical Center. www.childrensmemorial.org. Accessed October 12, 2009.

26. Romero MP. The epidemiology of firearms suicide in the United States. *J Urban Health*. 2002;79(1):39–48.
27. Center for Substance Abuse Treatment. *Substance Abuse and Suicide Prevention: Evidence and Implications*. Rockville, MD: Substance Abuse and Mental Health Service Administration; 2008. DHHS Pub. No.SMA-08-43352.
28. Absolve Self Abuse. Home page. http://www.selfabuse.org/self.abuse.html. Accessed October 12, 2009.
29. Roberts H, Meddings DR. *What Can be Done about the Social Determinants of Violence and Unintentional Injury: Background Paper of the Priority Public Health Conditions Knowledge Network of the Commission on Social Determinants of Health*. Geneva, Switzerland: World Health Organization; 2007.
30. Avarez L. Suicides of soldiers reach high of nearly 3 decades. *The New York Times*. January 30, 2009. http://www.nytimes.com/2009/01/30/us/30suicide.html?_r=1&pagewanted=print. Accessed September 30, 2009.
31. Michael WH. Suicide rates remain high: top officers urge supportive climate for troubled troops. *AirForceTimes*. March 23, 2009. http://www.airforcetimes.com/news/2009/03/airforce_suicide_032309w/. Accessed September 30, 2009.
32. Waitzkin H, Noble M. Caring for active duty military personnel in the civilian sector. *Soc Med*. 2009;4(3):56–69.
33. Tjaden P, Thoennes N. Extent, nature, and consequences of intimate partner violence: findings from the National Violence against Women Survey. Washington, DC: US Department of Justice; 2000. http://www.ojp.usdoj.gov/nij/pubs-sum/181867.htm. Publication No.NC 181867.
34. World Health Organization, International Society for Prevention of Child Abuse and Neglect. *Preventing Child Maltreatment: A Guide to Taking Action and Generating Evidence*. Geneva, Switzerland: World Health Organization. 2006.
35. Centers for Disease Control and Prevention. *Costs of Intimate Partner Violence against Women in the United States*. Atlanta, GA: CDC, National Center for Injury Prevention and Control; 2003. http://www.cdc.gov/ncipc/pub–res/ipv.Jost/ipv.htm. Accessed May 22, 2006.
36. Max W, Rice DP, Finkelstein E, Bardwell RA, Leadbetter S. The economic toll of intimate partner violence against women in the United States. *Violence Vict*. 2004;19(3):259–272.
37. Knox GW. *Introduction to Gangs*. 5th ed. Peotone, IL: New Chicago School Press; 2000.
38. Hill KG, Howell JC, Hawkins JD, Battin-Pearson SR. Childhood risk factors for adolescent gang membership: results from the Seattle social development project. *J Res Crime Delinq*. 1999;36(3):300–322.
39. Vigil, J. *Barrio Gangs: Street Life and Identity in Southern California*. Austin: University of Texas Press; 1998.
40. Sirpal SK. Causes of gang participation and strategies for prevention in gang members' own words. *J Gang Res*. 1997;4:13–22.
41. Centers for Disease Control and Prevention. Youth violence. http://www.cdc.gov/ViolencePrevention/youthviolence/index.html. Cited by: Blue Prints Program at the University of Colorado. Model programs center for the study of prevention of violence. http://www.colorado.edu/cspv/blueprints/modelprograms.html. Accessed September 18, 2009.
42. Centers for Disease Control and Prevention, Task Force on Community Preventive Services. The effectiveness of universal school–based programs for the prevention of violent and aggressive behavior. *MMWR Recomm Rep*. 2007;56(RR–7):1–12.

43. Center for the Study and Prevention of Violence. Blue Prints for violence prevention: Blue Prints Program at the University of Colorado. http://www.colorado.edu/cspv/blueprints/. Accessed September 20, 2009.
44. Big Brothers Big Sisters of America. Community Based Programs. http://www.bbbs.org. Accessed October 12, 2009.
45. Cabinet Office Social Exclusion Taskforce. Family nurse partnership. http://www.cabinetoffice.gov.uk/social_exclusion_task_nurse_partnership.aspx. Accessed September 20, 2009.
46. Seabrook J. "Don't shoot": a radical approach to the problems of gang violence. *The New Yorker*, June 22, 2009.
47. Guiding Principles for a Public Health Response to Terrorism. CDC Wonder and the American Public Health Association. http://wonder.cdc.gov/wonder/help/CMF/APHA-Terrorism.html. Accessed October 12, 2009.
48. Levy BS, Sidel VW. War and public health: an overview. In: Levy BS, Sidel VW, eds. *War and Public Health*. 2nd ed. New York: Oxford University Press; 2008: 3–22.
49. Garfield R. The epidemiology of war. In: Levy BS, Sidel VW, eds. *War and Public Health*. 2nd ed. New York: Oxford University Press; 2008:23–36.
50. Boyd WJ, Himmelstein D, Lasser K, et al. US medical student's knowledge about the military draft, the Geneva Convention, and military medical ethics. *Int J Health Serv*. 2007;37(4):643–650. http://www.baywood.com/hs/ijhs374.pdf. Accessed October 12, 2009.
51. Ernest CL, Stefanos NK. Chemical weapons. In: Levy BS, Sidel VW, eds. *War and Public Health*. 2nd ed. New York: Oxford University Press; 2008:117–134.
52. Barry SL, Sidel VW. Biological weapons. In: Levy BS, Sidel VW, eds. *War and Public Health*. 2nd ed. New York: Oxford University Press; 2008:135–151.
53. Gellert GA. *Confronting Violence*. 2nd ed. Washington, DC: American Public Health Association; 2002.

This page intentionally left blank

CHAPTER 19

Gender-Based Violence: A Call for Action

Nazanin Meshkat, MD, MHSc, and
Megan Landes, MD, MSc

> Violence against women both violates and impairs or nullifies the enjoyment by women of their human rights and fundamental freedoms. . . . In all societies, to a greater or lesser degree, women and girls are subjected to physical, sexual and psychological abuse that cuts across lines of income, class and culture.
>
> —*Beijing Declaration and Platform for Action, paragraph 112*[1]

INTRODUCTION

In the developed and developing world, gender-based violence (GBV) is endemic. Not only is it a major public health and human rights problem, but for the victims it can, and most often does, have devastating personal, health, societal, and economic consequences. Addressing the issue requires a multsectoral response among health policy makers, health care providers, governmental health, justice, and social service sectors. In this chapter, we examine the multifactorial issues of GBV, including the public health and health sector concerns and responses to GBV.

WHAT IS GENDER-BASED VIOLENCE?

The UN Declaration on the Elimination of Violence against Women defines violence against women as "Any act of gender-based violence that results in, or is likely to result in, physical, sexual or psychological harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or private life." In a further article of the UN declaration, it is specified that the acts may include but are not limited to "Acts of physical, sexual and psychological violence whether they be in the family or the community." The acts of violence specified in this chapter include the following: spousal battering, sexual abuse of female children, dowry-related violence, rape including marital rape, traditional practices harmful to women such as female genital mutilation, non-spousal violence, sexual harassment and intimidation, trafficking in women, forced

prostitution, and violence perpetrated or condoned by the state such as rape in war.[2]

Violence can take on many forms and can be perpetrated by either unknown or known individuals to the victim. Intimate partner violence (IPV) is any form of violence perpetrated by an intimate partner or spouse. Domestic violence is a term that often is used interchangeably with IPV; however, it is sometimes used to describe violence within the family, for example, by a father against his daughter.

Table 19.1 outlines different types of GBV throughout the life cycle.

Violence can be divided into physical, sexual, psychological, and emotional, and the threat of physical and sexual violence.[3] All forms of violence can have health implications for the victim. **Physical violence** is the intentional use of physical force with the potential to cause death, disability, injury, or harm. It can include punching, hitting, choking, biting, throwing objects at a person, kicking and pushing, and using a weapon such as a gun or a knife. **Sexual violence** is the use of physical force or coercion to compel a person to engage in a sexual act against his or her will whether or not the act is completed. An attempted or completed sex act involving a person who is unable to understand the nature or condition of the act, to decline participation, or to communicate unwillingness to engage in the sexual act (by illness, disability, influence of alcohol, or due to intimidation or pressure) constitutes sexual violence. Abusive sexual contact refers to intentionally touching, directly or through clothing, of the genitalia, anus, groin, breast, inner thigh, or buttocks of any person against his or her will or of any person who is unable to understand the nature or condition of the act, to decline participation, or to communicate unwillingness to engage in the sexual act. Insisting on unsafe sex is a form of sexual abuse.

The terminology of rape and sexual assault can present some confusion. Rape is the use of physical force, or threat of force or emotional coercion, to penetrate a woman's vaginal, oral, or anal orifices without her consent. Sexual assault is nonconsensual sexual contact that does not include penetration. Of note, sexual assault

Table 19.1
Gender-Based Violence through the Life Cycle

Phase	Type
Prenatal	Prenatal sex selection, battering during pregnancy, coerced pregnancy (rape during war).
Infancy	Female infanticide, emotional and physical abuse, differential access to food and medical care.
Childhood	Genital cutting, incest and sexual abuse, differential access to food, medical care, and education, child prostitution.
Adolescence	Dating and courtship violence, economically coerced sex, sexual abuse in the workplace, rape, sexual harassment, forced prostitution.
Reproductive Years	Abuse of women by intimate partners, marital rape, rape, dowry abuse and murders, partner homicide, psychological abuse, sexual abuse and harassment in the workplace.
Old Age	Abuse of widows, elder abuse.

Source: Heise L, Pitanguy J, Germaine A. *Violence against Women: The Hidden Health Burden*. World Bank Discussion Paper. Washington, D.C. World Bank; 1994.

and rape by an intimate partner is not considered a crime in most countries, and women in many societies do not consider forced sex as rape if they are married to, or cohabiting with, the perpetrator. Surveys in many countries reveal that approximately 10 to 15 percent of women report being forced to have sex by their intimate partner, and as high as 26 percent in Zimbabwe.[4, 5] However, the degree of sexual IPV varies greatly between studies and within contexts and has proven difficult to measure partially because of the cultural normalization of such acts.[6]

Psychological and emotional abuse can be defined as the trauma to the victim caused by acts, threats of acts, and coercive tactics when prior physical or sexual violence, or prior threat of physical or sexual violence, has occurred. Acts of humiliation; controlling what the person can and cannot do; withholding information or money; deliberately making the person feel diminished; creating, encouraging, and enforcing isolation from family and friends; threatening loss of custody of children; and smashing objects all constitute forms of psychological or emotional abuse.

INCIDENCE OF GBV

Many factors cloud efforts to determine the incidence of gender-based violence, such as the lack of awareness or normalization of the acts of violence that can lead to underreporting, methodological variability between studies and errors in data collection, and lack of definition and controversies of what constitutes GBV.[7] A landmark multinational study by the World Health Organization (WHO) has shed some light on the issue. Unless otherwise specified, the summaries in the following sections are from the WHO study.

IPV is the most common and endemic form of violence against women in the world. The percentage of women reporting physical assault by a male partner spans a wide spectrum, ranging from 13 percent in Japan to 67 percent in Papua New Guinea.[8] Sexual violence by intimate partners ranges from 6 percent in Japan and Serbia and Montenegro to 59 percent in provincial Ethiopia. In most settings, domestic sexual violence is less common than physical violence. This is not the norm, however, in provincial Bangladesh, provincial Ethiopia, and urban Thailand, domestic sexual violence is more common than domestic physical violence.

Emotional abuse by male partners is staggeringly high in most settings, ranging from 20 to 75 percent. Controlling acts tend to be higher among men who physically abuse their partners, ranging from 21 percent in Japan to 90 percent in urban United Republic of Tanzania. Controlling behaviors include controlling access to health care, with clear public health implications. The incidence of physical violence by nonpartners ranges from 5 percent in Ethiopia and Japan to 62 percent in Samoa. The incidence of sexual violence in nonpartners ranges from 1 percent in provincial Bangladesh and Ethiopia to 10 to 12 percent in Peru, Samoa, and Tanzania. Interestingly, despite the high rates of partner violence in Ethiopia, the incidence of nonpartner physical and sexual violence is relatively low, in contrast to Samoa where the opposite relationship is observed.

The trafficking of women and children for prostitution is one of the fastest growing areas of international criminal activity. Somewhere between and 1 and 2 million women and children of both sexes are trafficked each year worldwide, usually for forced labor, domestic servitude, or sexual exploitation.[9] Violence against

women (including rape, sexual abuse, slavery, and forced pregnancy) during armed conflict and its aftermath is increasingly apparent, and especially this type of violence has gained increased recognition in the international community. During the 1994 Rwanda genocide, for example, an estimated 500,000 women were raped; 60,000 women of all ages were raped during the Croatian and Bosnia-Herzegovina war; and 64,000 women were raped in the armed conflict in Sierra Leone. The practice continues in other areas of armed conflict, such as Darfur and the Democratic Republic of Congo.[10]

PRECIPITATING FACTORS OF GBV

The causes of gender-based violence are multifactorial, with complex interactions between each.[3] For simplification, the following section divides precipitating factors of GBV into two main categories: societal and individual factors.

Societal Factors

In many parts of the world, perceived male superiority and control over women is culturally engrained and reinforced by societal rules and institutions. Perceived male superiority can translate into acts of violence toward a woman and can be used as a tool to ensure women adhere to societal norms, such as performing household duties. Gender roles are supported by social institutions, such as the family, and by the law, as in the case of marriage and inheritance law.[3] Cross-cultural studies of wife abuse have found that nearly one-fifth of peasant and small-scale societies are essentially free of family violence. The existence of such cultures proves that male violence against women is not the inevitable result of male biology or sexuality, but more a matter of how society views masculinity. The prevalence of domestic violence in a given society, therefore, is the result of tacit acceptance by that society. The way men view themselves as men, and the way they view women, will determine whether they use violence or coercion against women.[11]

Taken to the extreme, cultural and societal norms can lead to harmful traditional practices against women. Female genital mutilation, dowry murder (a woman being killed by her husband or in-laws because her family is unable to meet their demands for her dowry), "honor" killings (rape victims, women suspected of engaging in premarital sex, and women accused of adultery who are murdered by relatives because the violation of a woman's chastity is viewed as an affront to the family's honor), early marriage, and persecution of "witches" (including torture and death) are examples of such practices.

Cultural attitudes and societal norms not only have significant impact on the incidence of GBV, but also on the designed interventions geared toward the victims to address the violence. Within a given society, gender-based violence can manifest itself differently in different cultural and religious subsets. For example, in Bangladesh, where the incidence of IPV is high, a study found that intimate partner violence was significantly higher among Muslim women compared with Hindu women (77 percent versus 62 percent).[12] The WHO multicountry study highlighted what many with experience in GBV have witnessed in the field. In many settings, women believe physical violence by their partners may be justified. The most accepted reasons are infidelity and disobeying a husband. There is a wide variation in women's

acceptance of violence, with the most marked variation being between the urban and rural settings. As an example, in provincial Ethiopia, about 80 percent of women believe that a husband is justified to beat his wife in the case of infidelity and approximately 80 percent believe this type of violence is justified if the wife disobeys her husband. In contrast, in urban Brazil and Serbia and Montenegro more than 90 percent of women do not accept any reason for violence. Similarly, attitudes to the right to refuse sex from a husband are varied. In provincial Bangladesh, Peru, Tanzania, Ethiopia, and Samoa, between 10 and 20 percent of women believe they do not have the right to refuse sex under any circumstance (including sickness). In contrast, in urban Serbia and Montenegro more than 95 percent of women believe that it is a woman's right to withhold sex.[8]

Individual Factors

While unequal gender relations lies at the core of GBV, within a given cultural context, a number of individual factors are associated with an elevated risk of experiencing such violence. These include age and marital status of the victim, education level, history of violence in the family, alcohol and drug abuse, personality and psychiatric disorders, and poverty.[13, 14] In the WHO multicountry study, age of women was identified as a risk factor for physical or sexual violence in most settings with the exception of Japan and Ethiopia. Similarly, separated or divorced women reported more partner violence, except in societies where the number of divorced and separated women is low (provincial Bangladesh and Ethiopia). Higher education was associated with less violence in most settings, even when controlling for income and age.

The interactions of poverty and GBV are complex. It is estimated that 70 percent of world's poor are women.[15] Studies have shown that women who occupy the lowest wealth quintile are more likely to experience victimization of all forms of violence, including IPV.[3, 12] The association is not fully understood: does poverty cause GBV, or does GBV cause poverty? On the one hand, poverty can place a strain on a relationship, and spark arguments and violence. Furthermore, poverty can be linked with lower education levels, a factor that is associated with increased rates of GBV. On the other hand, controlling behaviors of perpetrators, societal rules that restrict women's access to education and jobs, and the lower productivity of women who are burdened with the health and mental health consequences of GBV can create a cycle of poverty. The World Bank has found strong evidence that gender inequality undermines economic growth and constitutes an obstacle to poverty reduction.[16]

ADVERSE HEALTH EFFECTS OF GBV

Many victims of GBV often report being in poor physical health. In the WHO multicountry study, for example, the majority of women who had ever experienced physical or sexual violence reported poor or very poor health. Self-reported health is considered to be predictive of illness in a population-based survey.[20] Abused women often seek medical attention for conditions that at first glance may not alert the health provider to the existence of abuse. Unfortunately, the mechanisms leading to these chronic conditions have not been well studied.[21]

Women who are abused often do not seek care for their acute injuries. The prevalence of acute injury among abused women ranges from 19 to 55 percent.[8] These injuries are often in the form of soft tissue injuries, such as bruises, cuts, punctures, bites, broken bones, eye and ear injuries, and head injuries. Of note, most studies indicate that less than 30 percent of premenopausal women and less than 50 percent of postmenopausal women will have genital injuries visible to the naked eye after nonconsensual penetration.[17]

A study of U.S. hospital emergency departments found that the percentage of women who seek medical care for acute trauma from abuse is in the range of 2 to 3 percent.[14] The low rates could reflect poor access to health care resources because of controlling behaviors of perpetrators or because of the victim's shame or fear of consequences, the reluctance to report the true cause of injuries by the victim, or the health care provider's failure to screen appropriately.

Despite the apparent reluctance of abused women to seek care, the utilization of health care by victims of GBV is high. In a Canadian population-based study, battered women sought care from hospital emergency departments and saw a medical professional about three times more often than nonbattered women.[18] In a well-designed cost-analysis study, abused women generated around 92 percent more costs per year than nonbattered women, with mental health services accounting for most of the increased costs.[19] A 1994 World Bank report found that the medical costs of victims of GBV were estimated to be 2.5 times that of women who had not experienced such violence.[4] Another World Bank report estimated that GBV accounts for 5 percent of healthy years of life lost to reproductive-age women in developing countries.[16] Similarly, the Disability-Adjusted Life Years (DALY) lost by reproductive-age women by health consequences of GBV are high, estimated at 9.5 million DALYs, comparable to HIV (DALY 10.6 million) and sepsis during childbirth (DALY 10 million).[16] Clearly, both the economic and the psychological costs of GBV is high.

Table 19.2 summarizes some key health consequences of GBV.

REPRODUCTIVE HEALTH AND GBV

Research indicates that GBV has important implications in a woman's reproductive health. According to the 2005 UN World Population report, women's lack of reproductive control is a key limiting factor in major goals of economic development.[22] Adverse effects of GBV on a woman's reproductive health are numerous and include the following: gynecologic injuries, unwanted pregnancies, unsafe abortions, contraction of sexually transmitted diseases and HIV, and complications of pregnancy.[11]

Gynecologic injuries can occur during rape ranging in severity depending on the force of the assault. While many raped women present with no acute genital injuries,[17] most women will experience different levels of gynecologic dysfunction for years following sexual and physical violence. In a U.S. population-based study of self-reported data, the odds of having a gynecological problem were three times greater than average for victims of intimate partner abuse.[23] Gynecological problems can manifest as vaginal bleeding or infection, decreased sexual desire,

Table 19.2
Health Consequences of Gender Violence

Acute Physical
Fractures
Soft Tissue Injuries (including vaginal and rectal bruising, redness and swelling)
Lacerations (including vaginal and rectal tears)
Intracranial bleeding
Intra-abdominal bleeding
Pregnancy related (including abruptio placenta, premature labor, fetal distress and loss, uterine bleeding)
Death
Delayed Physical
Fistulas
Sexually transmitted disorders (including syphilis, chlamydia, gonorrhea, herpes, HIV, HPV)
Pelvic inflammatory diseases
Urinary tract infections
HIV and AIDS
Unwanted pregnancies
Unsafe abortions and sequel
Permanent disability
Chronic
Chronic Pain Syndromes (including headaches, chest pain, back pain)
Dizziness
Gastrointestinal symptoms
Sexual difficulties
Mental
Depression
Anxiety
Phobia/panic disorders
Post-traumatic stress disorder
Eating disorders
Suicide
Negative behavioral change (including poor self-esteem, sexual risk-taking, alcohol and drug abuse, smoking, physical inactivity, family instability, neglect of children, poor antenatal care)

dysmenorrhea, genital irritation, pain on intercourse, chronic pelvic pain, and urinary-tract infections.[21, 24]

A unique form of gynecologic injury occurs as a result of a "traditional" practice called female genital mutilation (FGM), which involves the removal of all or part of the female genitalia. It is estimated that more than 100 million women have been subjected to FGM, with millions more at risk annually.[25] FGM is predominantly practiced in Africa, West Asia, East Asia, and the Indian subcontinent. FGM exists in the developed world as well, primarily practiced by various immigrant populations. FGM is done for multiple reasons, including as an initiation ritual into

womanhood, as a means of controlling female sexuality, for cleanliness purposes, in the belief of enhancing fertility, as part of religious requirements, or as a prerequisite for marriage.

Apart from the painful experience (often done without anesthesia), FGM can have a number of immediate detrimental consequences, including hemorrhage, wound infection, sepsis, tetanus, urinary retention, and death. Long-term consequences are numerous and include formation of abscesses, fistulas and chronic pelvic infections, urethral damage and urinary incontinence, urinary tract infections, keloid scar formation, dyspareunia and sexual dysfunction, and menstrual flow obstruction. A study in six African countries found that adverse obstetric outcomes, such as Cesarean sections, prolonged or obstructed labor (increasing the risk of peripartum hemorrhage and infection), prenatal resuscitation, and death were significantly more common in women with FGM.[25]

There has been an international effort to reduce the incidence of FGM by means of education and public health campaigns, which highlight the associated morbidity and mortality associated with FGM. International and regional policies and laws have been passed to make FGM illegal; however, the prevalence of FGM still remains high and has been attributed in part to the emergence of a few alarming trends in communities where FGM continues to be practiced.[26] For example, there is now evidence of the "medicalization" of FGM, which refers to the practice of young children being taken to medical professionals to have the procedure performed. Another disturbing factor is that females are subjected to FGM at a younger age (as young as four years of age) to avoid detection by authorities. One rationale for this practice is the belief that younger girls will heal faster.[26]

On a positive front, a number of initiatives have proven effective in reducing the practice of FGM. In certain communities the practice is not completely abandoned, but "lesser cuts" are being practiced, suggesting a shift in cultural practices.[26] In other communities, such as among the Massai population in Kenya, the practice is substituted by alternative rites-of-passage ceremonies, through work done by community organizations.[27] The community organization also provides safe houses for girls who escape from FGM and forced child marriage, and provides them with counseling, educations, and reintegration services into their communities. Furthermore, those who practiced FGM were found alternative work. The grassroots work done in Kenya highlights the importance of an all-encompassing approach to FGM. Clearly, education campaigns and legal reforms, though important, cannot address the issue alone. The exploration of values underlying the practice, alternatives to the practice, and the impact on the community and individuals if they forgo the practice are important areas that require exploration and attention. Although some progress has been made to reduce the practice of FGM, in some societies, it still is viewed as an important, traditional practice with inherent social and cultural value.

IPV may have important public health implications for women and pregnancies. A number of studies have established that women subjected to IPV have more unplanned pregnancies than nonabused women.[28–35] In a multicountry study, findings showed that likelihood of an abused woman having an unwanted birth was approximately 40 percent higher than that among nonabused women.[36] These unplanned and unwanted pregnancies can at least partly be attributed to the lack of

control over reproductive decisions. Several studies found that many battered women experience sexually abusive and controlling acts by their partners, including the male's refusal to use condoms or other contraceptive methods.[37, 38] This situation is especially common in countries and cultures in which societal norms do not accord women equal say in reproductive decisions. Women in certain contexts are reluctant to raise the issue of contraception with their partners for fear of violence.[39]

In a recent study in Bangladesh, where the rates of IPV were as high as 75 percent, unwanted pregnancies were 50 percent more likely among those physically, as opposed to sexually, abused by their husbands, suggesting a reduced control over reproductive choices and family planning as a reason behind unwanted pregnancies.[12] A multicultural study found that even though abused women reported higher rates of having ever used contraception they were less likely to continue using contraception. Importantly, women also were less likely to use modern contraceptive methods; yet, in India, sterilization accounted for a majority of all contraceptive use.[36]

Unplanned pregnancies can have detrimental effects on the health of the woman and the fetus. According to recent estimates, up to 100,000 cases of maternal deaths per year and 4.6 million DALYs could be prevented globally if women had greater control over preventing unintended pregnancies.[40] There are numerous negative outcomes of unwanted pregnancies, including complications due to unsafe abortions, poor antenatal care with resulting poor pregnancy outcomes, and increased morbidity and mortality for mother and child associated with frequent pregnancies—that is, decreased "birth spacing." Women often resort to legal or illegal abortion to terminate an unwanted pregnancy, and studies have shown that rates of induced abortion are higher in those who are abused.[12] In the WHO multicountry study, women who were subjected to physical or sexual violence were more likely to report induced abortions.[8] In countries where women do not have access to legal and safe abortion, abortive techniques often are unsafe and unsupervised. In fact, unsafe abortion complications are one of the leading causes of maternal mortality in developing countries, causing 70,000 deaths annually, about 5 million women with temporary or permanent disability per year, and 5 million hospitalizations per year.[41] Unsafe abortions account for approximately 13 percent of maternal deaths, according to WHO estimates.[41] The complications, such as infection, bleeding, and long-term infertility, are more likely to occur in countries where abortion is illegal or severely restricted.[42, 43, 44] In countries with restricted access to safe abortion, the rate of abortion-related hospitalizations are significantly higher, ranging from 3 to 15 per 1,000 women of reproductive age.[45]

Women with unwanted pregnancies often fail to seek antenatal care early in pregnancy in both developed [46, 47] and developing countries,[48, 49] which could have an adverse impact on perinatal mortality and fetal birth weight. Research has shown mixed results between unintended pregnancies and perinatal mortality. Some researchers have found a positive relationship between unintended pregnancies and perinatal mortality in U.S. populations, even after controlling for other risk factors such as smoking and alcohol consumption.[50, 51] Yet similar findings from other studies looking at the association between low fetal birth weight and unintended pregnancies conducted in other countries are mixed.[52, 53]

Unintended pregnancies also can have a negative impact on a woman's health. One large study in Latin America and the Caribbean found adverse maternal outcomes associated with short pregnancy intervals (less than 5 months compared with 18 to 23 months), including third trimester bleeding, premature rupture of membranes, puerperal endometriosis, and anemia. These findings were evident even after adjusting for a variety of sociodemographic characteristics, fertility history, preexisting conditions, and behaviors.[54] Infant and child health outcomes also have been shown to be adversely affected when a pregnancy is unwanted. Significantly more women with unintended pregnancies refrain from breastfeeding [55, 56, 57] and unwanted pregnancies have been linked with subsequent child abuse and neglect.[58, 59]

The WHO multicountry study on violence against women reported prevalence rates for physical violence during pregnancy of between 1 percent and 28 percent.[8] Among all women who reported abuse during their lifetime by a partner, between 11 percent and 44 percent were also abused during their pregnancy. A majority of women reported that the abuse was the same or less severe or frequent than before the pregnancy, supporting findings from both developing and developed countries that pregnancy may have some protective effects. Abuse, however, also can start during pregnancy; the study reported that 13 to 50 percent of women were beaten for the first time during pregnancy.

Research has shown a link between domestic violence and adverse birth outcomes such as low birth weight, intrauterine growth retardation, preterm labor, fetal death, and miscarriages. A meta-analysis of studies in industrial countries showed a weak but significant association between abuse during pregnancy and low birth weight (odds ratio 1.4; 95 percent Confidence Interval (CI) 1:1–1:8).[60] The results of studies in developing countries showed a significant relation with low birth weight as well.[61–63] A number of studies have shown a correlation between domestic violence and preterm labor.[64, 65] Low birth weight and preterm labor can have negative consequences for infant outcomes, and be a significant cost to the health care system.[66]

Women who have ever experienced violence are more likely to have a nonlive birth (abortion, miscarriage, or stillbirth), even after controlling for confounding factors.[36] A number of studies support the association between domestic violence and pregnancy loss in the form of miscarriage or stillbirth.[12, 67–70] Sub-Saharan Africa and Southeast Asia have the highest rates of fetal death in the world (30 per 1,000 deliveries compared with 5 per 1,000 deliveries in developed countries).[12, 71] Perhaps not surprisingly, these regions also have the highest rates of GBV; it is estimated that more than 50 percent of women are subjected to domestic violence.

Maternal deaths have been linked to GBV. In the United States, homicide related to IPV accounts for as much as 20 percent of maternal deaths.[72] A West Indian study found that 16 percent of maternal deaths were associated with GBV.[73] A study from Bangladesh found 9 percent of maternal deaths were caused by injuries sustained during IPV.[74]

CHILD HEALTH AND GBV

Research indicates that GBV may have negative health effects among children of abused women, including increased mortality and malnutrition.[75–78] In a

Nicaraguan study, controlling for other known contributing factors to child mortality, about one-third of all child deaths were attributed to the experience of spousal violence. A multicultural study corroborated the finding that higher rates of infant and child mortality are observed among women who have ever been abused.[36] Although the pathways are not well understood, it has been shown that maternal violence can have detrimental effects on the mental health of children, including emotional and behavioral problems,[79, 80] and that children of abused women tend to have worse nutritional status (undernourished or anemic), especially in many of the developing countries.[36] Furthermore, vaccination of children against preventable diseases, a cornerstone of child health, has been shown to be lower among children of abused women.[36] Maternal violence was associated with decreased child immunization rates, decreased completion of immunization rates, and decreased rates of having received any immunizations.[36]

THE RELATIONSHIP BETWEEN GBV AND SEXUALLY TRANSMITTED INFECTION OR HIV

Globally, gender discrepancy in the spread of HIV/AIDS have been extensively documented: women are at greater risk. In Sub-Saharan Africa, women account for 60 percent of all HIV infections [81] and young women (15 to 24 years) account for 75 percent of all new infections.[82] The increased HIV risk for women has been linked to gender norms, which reinforce female subordination.[83] Commonly held ideas of masculinity, which encourage sexual risk-taking and the control of women, have been associated with infrequent or no condom use, wider sexual partner networks, and engagement in transactional sex.[84, 85] Such accepted behavior for men increases the risk of sexually transmitted infection (STI) and HIV for their female partners. A study of Indian truck drivers, for example, showed married men held extremely negative attitudes about condom use with their wives, most had had other sexual partners, and 74 percent reported intercourse with a commercial sex worker,[86] all of which place the female at high risk for acquiring a sexually transmitted infection and HIV.[87]

It should be surprising to see that many countries struggling with sizable HIV epidemics are experiencing concurrent GBV epidemics. Globally, these concurrent HIV and GBV epidemics share multiple risk factors and mounting evidence suggests a causal relationship between GBV and increased transmission of both STIs and HIV. Multiple cohort studies have demonstrated that women who experience GBV have higher rates of STI and HIV diagnosis than women who have not experienced GBV.[88–93] Several other studies have shown an increased risk in STI or HIV acquisition in rape survivors.[94–97] Additionally, violence in intimate partner relationships has been shown to increase the risk of STI acquisition [88, 98] and HIV transmission among previously HIV-discordant couples.[98–100] One large study of more than 24,000 Indian husband-wife dyads showed a seven-fold increased risk of HIV acquisition for women with abusive husbands when compared with women with nonabusive husbands.[101]

While GBV is a risk factor for HIV, further evidence suggests that the reciprocal relationship exists with HIV-positive women demonstrating higher rates of GBV than HIV-negative women.[102, 103] Cohen et al. reported a 12 times

increased risk of violence among HIV-positive women in the United States.[104] In Rwanda, HIV-positive pregnant women were more than twice as likely to experience physical IPV and had higher rates of all forms of violence than HIV-negative women.[105] The mechanisms of this relationship are complex and multidimensional. Direct and indirect mechanisms can increase the risk of HIV and STI transmission. Regarding direct mechanisms, forced sex can increase direct exposure to STI and HIV,[106] the degree of which depends on exposure to blood or genital trauma. Forced sex can lead to tears, lacerations, and disruption of the mucosal barrier, which increases susceptibility for transmission of both STIs and HIV. Heightened risk of transmission also may be associated with multiple-site genital trauma, multiple perpetrators in sexual violence, and anal penetration. Additionally, forced sex often occurs without the use of condoms.

The increased risk of STI and HIV transmission for women experiencing GBV is accentuated by both the risk profile and subsequent increased prevalence of STI and HIV among those men reporting GBV perpetration. Several studies indicate increased sexual risk-taking behaviors among perpetrators such as anal sex, unprotected sex, multiple partners, and buying sex.[107–110] Additionally, a greater prevalence of STI and HIV diagnosis among perpetrators is reported,[101, 108, 111] thus increasing the exposure of STI and HIV to women who are targets of sexual violence.

Perhaps a more disturbing trend placing a woman at risk for STI and HIV is the high prevalence of forced sex and rape during armed conflict.[112–114] Widespread sexual violence as an intentional tool of war has been documented in multiple conflicts in the past decade, including in the former Yugoslavia, the Democratic Republic of the Congo, Rwanda, and most recently in the Darfur region of Sudan.[115] Armed conflicts and forced migration greatly increase a women's vulnerability to coercive sex or transactional sex to gain basic needs such as food, safety, and shelter. A WHO study showed that 27 percent of single women in eastern and central Sudan had become sex workers to earn a living.[106] Other studies have shown persistently high or increasing levels of sexual violence in the transitional environment of postconflict states.[116]

Given the potential increased risk of STI/HIV transmission for women in conflict, humanitarian, public health, and medical resources clearly should be directed toward mitigating this risk.[117–119] Although widespread rape may not be yet shown epidemiologically to increase the risk for STI and HIV at a population level, it is important to remember that forced sex still poses a considerable threat for STI and HIV transmission to the individual.[120, 121]

Regarding indirect mechanisms, Andersson et al. [87] argue that the interaction between GBV and HIV goes well beyond increasing the direct physical risk for HIV transmission to indirectly influencing women's risk-taking behavior and mitigating their ability to engage in HIV prevention. Multiple studies have shown that women who have experienced GBV, either in childhood or as an adult, are more likely to engage in high-risk sexual behaviors, such as commercial sex work, having multiple partners or concurrent nonprimary partners.[88, 122–126] A Ugandan cohort showed those women reporting coercion during their first intercourse were less likely to report contraceptive use, condom use at last sex and consistent use of condoms in the last six months.[127] Several other studies support inconsistent condom use in women who have experienced, or are experiencing, violence.[128, 129]

GBV affects the health care–seeking behaviors of women and their ability to engage in prevention initiatives, such as voluntary counseling and testing. Women in violent relationships are less likely to disclose their status due to fear of harm and report negative experiences from disclosure.[130–132] This fear of violence inevitably clashes with prevention messaging to increase safer sex practices, such as condom use. In Chennai, India, women were shown to forgo safe practices for fear of violence,[133] and other studies have shown that women in violent relationships discuss and use condoms less often than women in nonviolent relationships.[128]

MENTAL HEALTH AND GBV

Abused women often seek help for mental health issues. The mental health effects of GBV can range from insomnia, emotional distress, and social dysfunction to full-blown clinical depression, post-traumatic stress disorder (PTSD), and anxiety. In the WHO multicountry study, emotional distress (identified through symptoms such as crying easily, inability to enjoy life, fatigue, and thoughts of suicide, in the four weeks before the interview) was more common in women experiencing physical or sexual violence. A Canadian population-based study demonstrated that in addition to depression, abused women had significantly more anxiety, insomnia, and social dysfunction than those not abused, with physical violence having a stronger effect than psychological abuse.[134] The elevated prevalence of emotional stress has been further corroborated in studies conducted in both developed and developing countries.[135–137] Additionally, alcohol and drug abuse often is observed in victims of IPV. The substance abuse may be used as a coping mechanism, or as self-treatment for symptoms of depression and PTSD.

Depression and PTSD are common mental health sequelae of intimate partner violence. A meta-analysis has shown the risk of depression and PTSD is higher among those who are victims of intimate partner violence than that resulting from childhood sexual assault.[138] The emotional distress, low self-esteem and isolation that ensue after exposure to IPV can result in a constellation of symptoms seen in *battered woman's syndrome,* a term coined in the 1990s after research on victims of IPV.[139, 140] PTSD has been shown to be higher in abused women compared with nonabused women, with the severity of abuse and partner dominance as important precursors.[141]

Suicide and GBV and IPV have been linked.[138, 142, 143] The WHO multicultural study found that women who have been abused by their partners were more than 50 percent more likely to contemplate suicide compared with their nonabused counterparts.[8]

U.S.-based studies established that victims of IPV are five times more likely to attempt suicide.[3] In Sri Lanka, the number of suicides by girls and women ages 15 to 24 years old is 55 times greater than the number of deaths resulting from pregnancy and childbirth.[144]

WHAT CAN BE DONE TO END GBV?

Public health initiatives often are tightly bound to existing legal frameworks, and this holds especially true in the case of GBV. Over the last two to three decades, the tireless work of women's groups and nongovernmental organizations has

succeeded in raising awareness about GBV. The Convention on the Elimination of All Forms of Discrimination against Women, adopted by the United Nations in 1979, legally binds all UN members to strive for accomplishing human rights for women.[145] Recent precedents regarding mass rape set at the International Criminal Tribunals for both Rwanda and the former Yugoslavia suggest that other violent acts such as sexual assault, sexual slavery, forced prostitution, forced sterilization, forced abortion, and forced pregnancy may qualify as crimes of torture, crimes against humanity, and crimes of genocide.[11]

It is excellent that GBV is now recognized as a major public health and human rights issue, but discussion has focused on how best to reduce and eliminate this type of violence around the world. Unfortunately, as of this writing, assessment and evaluation of programs and interventions strategies are few and far between, and those programs that have been assessed are more often than not plagued by methodological flaws, including lack of adequate follow-up.[146] Studies that look at the prevention of GBV show that despite the positive effects of some interventions, the rates of abuse even among the intervention groups remains stubbornly and substantially high.

Although a number of screening tools are available, studies have not comparatively assessed the effectiveness of such tools on reducing IPV. Counseling and empowerment of women who are victims of IPV have been suggested. A common counseling intervention is the empowerment approach directed toward increasing a woman's independence and control. But, the few studies that have been conducted fail to find long-term benefits of such intervention strategies.[147] Furthermore, the effectiveness of shelter stay in reducing violence, a mainstay intervention strategy for abused women, has not been demonstrated in any study.

There are a few studies that have evaluated the effectiveness of interventions to curtail the effects of GBV during pregnancy. A number of studies looking at empowerment intervention among pregnant woman have shown that women report decreased levels of violence post intervention.[148] It is unclear what level of intervention is most effective and how effective the intervention is in the long term. Perhaps the failure to show a reduction in abusive behavior can be attributed to the fact that the abuser is rarely if ever included in the intervention.

Because a high percentage of abused women present to a hospital emergency department, it is logical to think that screening intervention programs might be more successful in this setting. Given the nature of the hospital emergency department, which is more focused on treating and discharging patients rather than providing long-term follow-up care, intervention strategies have not been successful in the long term.

Because of the complex interaction between GBV and HIV, many researchers and advocates insist policy changes, programming, and interventions must concurrently address GBV and HIV.[83] To date, several primary prevention initiatives have attempted to decrease both the transmission of HIV and the prevalence of GBV. Two large programs that took place in South Africa, Sisters for Life and Stepping Stones, used training programs to address gender roles and communication as well as GBV and IPV and HIV prevention for men and women.[149, 150] These programs failed to demonstrate a significant reduction in HIV incidence, but both demonstrated improved risk behaviors in men and a significant reduction in IPV in

the short term. Postexposure prophylaxis (PEP) for victims of sexual violence has been explored as a means of primary prevention of HIV in GBV victims. Both the WHO and the Centers for Disease Control and Prevention recommend STI and HIV-PEP for nonoccupational exposures (that is, sexual assault) in cases in which the likelihood is a high or unknown that the assailant may have an STI or be HIV positive.[151, 152][1]

A number of studies have looked at primary prevention of abuse through education campaigns directed at young people, with some promising results.[153–155] These studies, however, are limited by short follow-up timelines, raising questions about sustainability. A focus on the male has led to some promising change in attitudes. For example, Program H, developed by *Instituto Promundo*, targeted young men between the ages of 15 and 24. After a two-year impact evaluation, results showed a significant increase in condom use and decrease in domestic violence.[156] Long-term success and sustainability is not known, however. So much more needs to be done to better understand how to effectively intervene to make a difference.

CONCLUSION

GBV is a major public health concern with potentially devastating consequences to the victim and her offspring. That such violence continues in spite of efforts by the United Nations and other world bodies to dictate against its practice illustrates how difficult it is to eliminate. This chapter highlighted the extent of the problem and illustrated the serious negative physical and mental health sequelae of both GBV and IPV. The prevention and management of GBV requires a culturally sensitive, multiprong approach at all levels of society. Unfortunately, the evidence indicates that intervention programs have done little to stem the practice. Fortunately, however, the media have brought this shameful practice into the public's consciousness, which is an important step forward. Nevertheless, much work remains both in developed and developing countries to ensure the safety and well-being of women of all ages against all forms of GBV.

REFERENCES

1. United Nations. Beijing declaration and platform for action. http://www.un.org/womenwatch/daw/beijing/platform/. Accessed October 27, 2009.
2. United Nations. *United Nations General Assembly Declaration on the Elimination of Violence against Women*. New York: United Nations; 1993. http://www.unhchr.ch/huridocda/huridoca.nsf/(Symbol)/A.RES.48.104.En?Opendocument. Accessed October 27, 2009.
3. World Health Organization. *Integrating Poverty and Gender into Health Programs: A Sourcebook for Health Professionals*. Geneva, Switzerland: World Health Organization; 2005.

1. For further details on the management of the rape patient refer to the following two documents: "Guidelines for the medico-legal care for victims of sexual violence," World Health Organization, 2003; and "Clinical Management of Rape Survivors: Developing protocols for use with refugees and internally displaced persons," World Health Organization/United Nations High Commissioner for Refugees, 2004.

4. Heise LL, Pitanguy J, Germaine A. *Violence against Women: The Hidden Health Burden*. Washington, DC: The World Bank; 1994.
5. Watts C, Keogh E, Ndlovu M, Kwaramba R. Withholding of sex and forced sex: dimensions of violence against Zimbabwean women. *Reprod Health Matters*. 1998;6:57–65.
6. Watts C, Zimmerman C. Violence against women: global scope and magnitude. *Lancet*. 2002;359:1232–1237.
7. De Waal A, Klot JF, Mahajan M, Huber D, Frerks G, M'Boup S. *HIV/AIDS Security and Conflict: New Realities, New Responses*. Brooklyn, NY: Social Science Research Centre; 2009. http://asci.researchhub.ssrc.org/rdb/asci-hub. Accessed September 22, 2009.
8. World Health Organization. *WHO Multi-Cultural Study on Women's Health and Domestic Violence against Women: Initial Results on Prevalence, Health Outcomes and Women's Responses*. Geneva, Switzerland: World Health Organization; 2005.
9. US Department of State. *Trafficking in Women and Children: The US and International Response.* Washington, DC: US Department of State; 2000.
10. United Nations Development Fund for Women. UNIFEM facts and figures on violence against women. http://www.unifem.org/gender_issues/violence_against_women/facts_figures.php?page=7. Accessed October 27, 2009.
11. United Nations Population Fund. Gender equality. http://www.unfpa.org/gender/violence.htm. Accessed October 27, 2009.
12. Silverman JG, Gupta J, Decker MR, Kapur N, Raj A. Intimate partner violence and unwanted pregnancy, miscarriage, induced abortion, and stillbirth among a national sample of Bangladeshi women. *BJOG*. 2007;114:1246–1252.
13. Gostin LO, Lazzarini Z, Alexander D, Brandt AM, Mayer KH, Silverman DC. HIV testing, counseling, and prophylaxis after sexual assault. *JAMA*. 1994;271:1436–1444.
14. Dearwater SR, Coben JH, Nah G, et al. Prevalence of intimate partner violence in women treated at community hospital emergency departments. *JAMA*. 1998;280:433–438.
15. United Nations Development Programme. *Human Development Report: Gender and Human Development*. New York: Oxford University Press; 1995.
16. The World Bank. *Engendering Development through Gender Equality in Rights, Resources, and Voice: A World Bank Policy Research Report*. Washington, DC: The World Bank; 2001.
17. Biggs M, Stermac LE, Divinsky M. Genital injuries following sexual assault of women with and without prior sexual experience. *Can Med Assoc J*. 1998;159:33–37.
18. Ratner PA. The incidence of wife abuse and mental health status in abused wives in Edmonton, Alberta. *Can J Public Health*. 1993;84:246–249.
19. Wisner CL, Gilmer TP, Saltzman LE, Zink TM. Intimate partner violence against women: do victims cost health plans more? *J Fam Pract*. 1999;48:439–443.
20. McDonough P, Duncan GJ, Williams D, House J. Income dynamics and adult mortality in the United States, 1972–1989. *Am J Public Health*. 1997;87:1476–1483.
21. Campbell JC. Health consequences of intimate partner violence. *Lancet*. 2002;359:1331–1336.
22. United Nations Department of Economic and Social Affairs. *Population Challenges and Development Goals*. New York: United Nations Department of Economic and Social Affairs; 2005.
23. Coker AL, Smith PH, Bethea L, King MR, McKeown RE. Physical health consequences of physical and psychological intimate partner violence. *Arch Fam Med*. 2000;9:451–457.

24. Reiter RC, Shakerin LR, Gambone DO, Milburn AK. Correlation between sexual abuse and somatization in women with somatic and nonsomatic chronic pelvic pain. *Am J Obstet Gynecol*. 1991;165:104–109.
25. World Health Organization. *Female Genital Mutilation and Obstetric Outcome: WHO Collaborative Prospective Study in Six African Countries*. Geneva, Switzerland: World Health Organization; 2006.
26. United Nations Children's Fund. *Female Genital Mutilation/Cutting: A Statistical Exploration*. New York: UNICEF; 2005.
27. United Nations Population Fund. *A Holistic Approach to the Abandonment of Female Genital Mutilation/Cutting*. Geneva, Switzerland: UNFPA; 2007.
28. Eby K, Campbell J, Sullivan C, Davidson W. Health effects of experiences of sexual violence for women with abusive partners. *Health Care of Women Int*. 1995;16:563–567.
29. Amaro H, Fried LE, Cabral H, Zuckerman B. Violence during pregnancy and substance use. *Am J Public Health*. 1990;80:575–579.
30. Hillard PJ. Physical abuse in pregnancy. *Obstet Gynecol*. 1985;66:185–190.
31. Pallitto CC, O'Campo P. The relationship between intimate partner violence and unintended pregnancy: analysis of a national sample from Colombia. *Int Fam Plan Perspect*. 2004;30:165–173.
32. Kaye DK, Mirembe FM, Bantebya G, Johansson A, Ekstrom AM. Domestic violence as risk factor for unwanted pregnancy and induced abortion in Mulago Hospital, Kampala, Uganda. *Trop Med Int Health*. 2006;11:90–101.
33. Martin SL, Kilgallen B, Tsui AO, Maitra K, Singh KK, Kupper LL. Sexual behaviors and reproductive health outcomes: associations with wife abuse in India. *JAMA*. 1999;282:1967–1972.
34. Muthal-Rathore A, Tripathi R, Arora R. Domestic violence against pregnant women interviewed at a hospital in New Delhi. *Int J Gynaecol Obstet*. 2002;76:83–85.
35. Glander SS, Moore ML, Michielutte R, Parsons LH. The prevalence of domestic violence among women seeking abortion. *Obstet Gynecol*. 1998;91:1002–1006.
36. Kishor S, Johnson K. *Profiling Domestic Violence: A Multi-Country Study*. Calverton, MD: ORC Macro International, Inc.; 2004.
37. Coker AL, Richter DL. Violence against women in Sierra Leone: frequency and correlates of intimate partner violence and forced sexual intercourse. *Afr J Reprod Health*. 1998;2:61–72.
38. Wood K, Jewkes R. Violence, rape and sexual coercion: everyday love in a South African township. *Gender Dev*. 1997;5:41–46.
39. Bawah, AA, Akweongo P, Simmons R, Phillips JF. Women's fears and men's anxieties: the impact of family planning on gender relations in northern Ghana. *Stud Fam Plann*. 1999;30:54–66.
40. Butler P. Effective contraception: more benefits than meet the eye. *Prog Reprod Health Res*. 2003;62:1–3.
41. World Health Organization. *Unsafe Abortion: Global and Regional Estimates of the Incidence of Unsafe Abortion and Associated Mortality in 2003*. 5th ed. Geneva, Switzerland: World Health Organization; 2007. http://www.who.int/reproductive health/publications/unsafe_abortion/9789241596121/en/index.html.
42. Baker J, Khasiani S. Induced abortion in Kenya: case histories. *Stud Fam Plann*. 1992;23:34–44.
43. Magadi MA. Unplanned childbearing in Kenya: the socio-demographic correlates and the extent of repeatability among women. *Soc Sci Med*. 2003;56:167–178.
44. Okonofua FE. Induced abortion: a risk factor for infertility in Nigerian women. *J Obstet Gynecol*. 1994;14:272–276.

45. Henshaw SK, Singh S, Haas T. The incidence of abortion worldwide. *Int Fam Plan Perspect*. 1999;25:S30–S38.
46. Marsiglio W, Mott FL. Does wanting to become pregnant with a first child affect subsequent maternal behaviors and infant birth weight? *J Marriage Fam*. 1988;50:1023–1036.
47. Korenman S, Kaestner R, Joyce T. Consequences for infants of parental disagreement in pregnancy intention. *Perspect Sex Reprod Health*. 2002;34:198–205.
48. Eggleston E. Unintended pregnancy and women's use of prenatal care in Ecuador. *Soc Sci Med*. 2000;51:1011–1018.
49. Marston C, Cleland J. Do unintended pregnancies carried to term lead to adverse outcomes for mother and child? an assessment in five developing countries. *Popul Stud*. 2003;57:77–93.
50. Bustan MN, Coker AL. Maternal attitude toward pregnancy and the risk of neonatal death. *Am J of Public Health*. 1994;84:411–414.
51. Laukaran VH, van den Berg BJ. The relationship of maternal attitude to pregnancy outcomes and obstetric complications: a cohort study of unwanted pregnancy. *Am J Obstet Gynecol*. 1980;136:374–379.
52. Gadow EC, Paz JE, Lopez-Camelo JS, et al. Unintended pregnancies in women delivering at 18 South American hospitals. *Hum Reprod*. 1998;13:1991–1995.
53. Sable MR, Spencer JC, Stockbauer JW, Schramm WF, Howell V, Herman AA. Pregnancy wanted-ness and adverse pregnancy outcomes: differences by race and Medicaid status. *Fam Plann Perspect*. 1997;29:76–81.
54. Conde-Agudelo A, Belizán JM. Maternal morbidity and mortality associated with inter-pregnancy interval: cross sectional study. *BMJ*. 2000;321:1255–1259.
55. Chinebuah B, Perez-Escamilla R. Unplanned pregnancies are associated with less likelihood of prolonged breast-feeding among primiparous women in Ghana. *J Nutr*. 2001;131:1247–1249.
56. Korenman S, Kaestner R, Joyce T. Consequences for infants of parental disagreement in pregnancy intention. *Perspect Sex Reprod Health*. 2002;34:198–205.
57. Perez-Escamilla R, Cobas JA, Balcazar H, Benin M. Specifying the antecedents of breast-feeding duration in Peru through a structural equation model. *Public Health Nutr*. 1999;2:461–467.
58. Barber JS, Axinn WG, Thornton A. Unwanted childbearing, health, and mother-child relationships. *J Health Soc Behav*. 1999;40:231–257.
59. Altemeier WA, O'Connor S, Vietze P, Sandler H, Sherrod K. Prediction of child abuse: a prospective study of feasibility. *Child Abuse Negl*. 1984;8:393–400.
60. Murphy CC, Schei B, Myhr TL, Du Mont J. Abuse: a risk factor for low birth weight? a systematic review and meta-analysis. *CMAJ*. 2001;164:1567–1572.
61. Heise LL, Ellsberg MC, Gottemoeller M. Ending violence against women. *Popul Rep*. 1999;27:1–43.
62. Nuñez-Rivas H, Monge-Rojas R, Grios-Davila C, Elizondo-Ureña AM, Rojas-Chavarria A. La violencia física, psicológica, emocional y sexual durante el embarazo: riesgo reproductivo predictor de bajo peso al nacer en Costa Rica [Physical, psychological, emotional, and sexual violence during pregnancy: reproductive risk predictor of low birthweight in Costa Rica]. *Rev Panam Salud Publica*. 2003;14:75–83.
63. Valladares E, Ellsberg M, Pena R, Hogberg U, Persson LA. Physical partner abuse during pregnancy: a risk factor for low birth weight in Nicaragua. *Obstet Gynecol*. 2002;100:700–705.

64. Berenson AB, Wiemann CM, Wilkinson GS, Jones WA, Anderson GD. Perinatal morbidity associated with violence experienced by pregnant women. *Am J Obstet Gynecol*. 1994;170:1760–1769.
65. Shumway J, O'Campo P, Gielen A, Witter FR, Khouzami AN, Blakemore KJ. Preterm labor, placental abruption, and premature rupture of membranes in relation to maternal violence or verbal abuse. *J Matern Fet Med*. 1999;8:76–80.
66. World Health Organization. *Health Benefits of Family Planning*. Geneva, Switzerland: World Health Organization, Department of Reproductive Health and Research; 1995. http://www.who.int/reproductivehealth/publications/family_planning/HRP_FHE_FPP_95_15/en/index.html. Accessed October 27, 2009.
67. Boy A, Salihu HM. Intimate partner violence and birth outcomes: a systematic review. *Int J Fertil Womens Med*. 2004;49:159–163.
68. Alio AP, Nana PN, Salihu HM. Spousal violence and potentially preventable single and recurrent spontaneous fetal loss in an African setting: cross-sectional study. *Lancet*. 2009;373:318–324.
69. Janssen PA, Holt VL, Sugg NK, Emanuel I, Critchlow CM, Henderson AD. Intimate partner violence and adverse pregnancy outcomes: a population-based study. *Am J Obstet Gynecol*. 2003;188:1341–1347.
70. Jejeebhoy SJ. Associations between wife-beating and fetal and infant death: impressions from a survey in rural India. *Stud Fam Plann*. 1998;29:300–308.
71. Stanton C, Lawn JE, Rahman H, Wilczynska-Ketende K, Hill K. Stillbirth rates: delivering estimates in 190 countries. *Lancet*. 2006;367:1487–1494.
72. Campbell J, Garcia-Moreno C, Charps P. Abuse during pregnancy in industrialized and developing countries. *Violence Against Women*. 2004;10:770–789.
73. Ganatra BR, Coyaji KJ, Rao VN. Too far, too little, too late: a community-based case-control study of maternal mortality in rural west Maharashtra, India. *Bull World Health Organ*. 1998;76(6):591–598.
74. Faveau V, Koenig M, Chakroborty, J, Chowdury A. Causes of maternal mortality in rural Bangladesh, 1976–1985. *Bull World Health Organ*. 1988;66:643–651.
75. Ramachandran P. Need of organization of antenatal and intrapartum care in India. *Demogr India*. 1992;21:179–193.
76. Ganatra BR, Coyaji KJ, Rao VN. Too far, too little, too late: A community-based case-control study of maternal mortality in rural west Maharashtra, India. *Bull World Health Organ*. 1998;76:591–598.
77. Jejeebhoy SJ. Associations between wife-beating and fetal and infant death: impressions from a survey in rural India. *Stud Fam Plann*. 1998;29:300–308.
78. Asling-Monemi K, Pena R, Ellsberg MC, Persson LA. Violence against women increases the risk of infant and child mortality: a case-referent study in Nicaragua. *Bull World Health Organ*. 2003;81:10–18.
79. Edelson JL. Children witnessing of adult domestic violence. *J Interpers Violence*. 1999;14:839–870.
80. Jouriles EN, Murphy CM, O'Leary KD. Interspousal aggression, marital discord, and child problems. *J Consult Clin Psychol*. 1989;57:453–455.
81. United Nations. *Report on the Global HIV/AIDS Epidemic 2008*. Geneva, Switzerland: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2008.
82. United Nations. *Report on the Global AIDS Epidemic 2006*. Geneva, Switzerland: Joint United Nations Programme on HIV/AIDS (UNAIDS); 2006.
83. Greig A, Peacock D, Jewkes R, Msimang S. Gender and AIDS: time to act. *AIDS*. 2008;22(Suppl 2):S35–S43.

84. Santana MC, Raj A, Decker MR, March AL, Silverman JG. Masculine gender roles associated with increased sexual risk and intimate partner violence perpetration among young adult men. *J Urban Health*. 2006;83:575–585.
85. Dunkle KL, Jewkes RK, Nduna M, et al. Perpetration of partner violence and HIV risk behavior among young men in rural Eastern Cape, South Africa. *AIDS*. 2006;20:2107–2114.
86. Bryan AD, Fisher JD, Benziger TJ. HIV prevention information, motivation, behavioral skills and behavior among truck drivers in Chennai, India. *AIDS*. 2000;14:756–758.
87. Andersson N, Cockcroft A, Shea B. Gender-based violence and HIV: relevance for HIV prevention in hyperendemic countries of southern Africa. *AIDS*. 2008;22(S4):S73–S86.
88. Wu E, El-Bassel N, Witte SS, Gilbert L, Chang M. Intimate partner violence and HIV risk among urban minority women in primary health care settings. *AIDS Behav*. 2003;7:291–301.
89. van der Straten A, King R, Grinstead O, Vittinghoff E, Serufilia A, Allen S. Sexual coercion, physical violence and HIV infection among women in steady relationships in Kigali, Rwanda. *AIDS Behav*. 1998;2:61–73.
90. Kiarie JN, Farquhar C, Richardson BA, et al. Domestic violence and prevention of mother-to-child transmission of HIV-1. *AIDS*. 2006;20:1763–1769.
91. Dunkle KL, Jewkes RK, Brown HC, Gray GE, McIntryre JA, Harlow SD. Gender-based violence, relationship power and risk of HIV infection in women attending antenatal clinics in South Africa. *Lancet*. 2004;363:1415–1421.
92. Silverman JG, Decker MR, Saggurti N, Balaiah D, Raj A. Intimate partner violence and HIV infection among married Indian women. *JAMA*. 2008;300:703–710.
93. Jewkes R, Dunkle K, Nduna M, et al. Factors associated with HIV sero-status in young rural South African women: connections between intimate partner violence and HIV. *Int J Epidemiol*. 2006;35:1461–1468.
94. Jenny C, Hooton T, Bowers A, et al. Sexually transmitted disease in victims of rape. *N Engl J Med*. 1990;322:713–716.
95. Glaser JB, Schachter J, Benes S, Cummings M, Frances CA, McCormack WM. Sexually transmitted disease in post-pubertal female rape victims. *J Infect Dis*. 1991; 164:726–730.
96. Menick DM, Ngoh F. Serum prevalence of HIV infection among sexually abused children in Cameroon. *Med Trop*. 2003;63:155–158.
97. Decker MR, Silverman JG, Raj A. Dating violence and sexually transmitted disease/HIV testing and diagnosis among adolescent females. *Pediatrics*. 2005;116: e272–e276.
98. Dude AM. Spousal intimate partner violence is associated with HIV and other STIs among married Rwandan women. *AIDS Behav*. 2009; Epub ahead of print.
99. Kim A, Malele F, Kaiser R, et al. HIV infection among internally displaced women and women residing in river populations along the Congo River, Democratic Republic of Congo. *AIDS Behav*. 2009;13:914–920.
100. Fonck K, Leye E, Kidula N, Ndinya-Achola J, Temmerman M. Increased risk of HIV in women experiencing physical partner violence in Nairobi. *AIDS Behav*. 2005;9:335–339.
101. Decker MR, Seage GR III, Hemenway D, et al. Intimate partner violence functions as both a risk marker and risk factor for women's HIV infection: findings from Indian husband-wife dyads. *J Acquir Immune Defic Syndr*. 2009;51:593–600.
102. Gruskin L, Gange SJ, Celentano D, et al. Incidence of violence against HIV-infected and uninfected women: findings from the HIV Epidemiology Research (HER) study. *Urban Health*. 2002;79:512–524.

103. Maman S, Mbambo JK, Hogan NM, et al. HIV-positive women report more lifetime partner violence: findings from a voluntary counseling and testing clinic in Dar es Salaam, Tanzania. *Am J Public Health*. 2002;92:1331–1337.
104. Cohen M, Deamant C, Barkan S, et al. Domestic violence and childhood sexual abuse in HIV infected women and women at risk for HIV. *Am J Public Health*. 2000;90:560–565.
105. Ntaganira J, Muula AS, Masaisa F, Dusabeyezu F, Siziya S, Rudatsikira E. Intimate partner violence among pregnant women in Rwanda. *BMC Women's Health*. 2008;8:17.
106. World Health Organization, Global Coalition on Women and AIDS. *Intimate Partner violence and HIV/AIDS*. Geneva, Switzerland: Global Coalition on Women and AIDS/WHO; 2004.
107. Dunkle KL, Jewkes RK, Nduna M, et al. Perpetration of partner violence and HIV risk behavior among young men in rural Eastern Cape, South Africa. *AIDS*. 2006;20:2107–2114.
108. Decker M, Seage G III, Hemenway D, et al. Intimate partner violence perpetration, standard and gendered STI/HIV risk behavior and STI/HIV diagnosis among a clinic-based sample of men. *STI*. 2009; Epub ahead of print.
109. Raj A, Santana C, La Marche A, Amaro H, Cranston K, Silverman JG. Perpetration of intimate partner violence associated with sexual risk behaviors among adult men. *Am J Public Health*. 2006;96:1873–1878.
110. Silverman JG, Decker MR, Kapur NA, Gupta J, Raj A. Violence against wives, sexual risk and sexually transmitted infection among Bangladeshi men. *Sex Transm Infect*. 2007;83:211–215.
111. Raj A, Reed E, Welles SL, Santana MC, Silverman JG. Intimate partner violence perpetration, risky sexual behavior, and STI/HIV diagnosis among heterosexual African American men. *Am J Mens Health*. 2008;2:291–295.
112. Shanks L, Schull MJ. Rape in war: the humanitarian response. *Can Med J*. 2000;163:1148–1149.
113. Spiegel PB. HIV/AIDS among conflict-affected and displaced populations: dispelling myths and taking action. *Disasters*. 2004;28:322–339.
114. Mills EJ, Singh S, Nelson B, Nachega J. The impact of conflict on HIV/AIDS in Africa. *Int J STD AIDS*. 2006;17:713–717.
115. World Health Organization. *Violence against Women and HIV/AIDS: Sexual Violence in Conflict Settings and the Risk of HIV*. Geneva, Switzerland: The Global Coalition on Women and AIDS; 2004.
116. Rehn E, Sirleaf EJ. *Women, War, Peace: The Independent Experts' Assessment on the Impact of Armed conflict on Women and Women's Role in Peace-Building*. New York: UNIFEM; 2002.
117. Mills EJ, Singh S, Nelson B, Nachega J. The impact of conflict on HIV/AIDS in Africa. *Int J STD AIDS*. 2006;17:713–717.
118. Save the Children. *HIV and Conflict: A Double Emergency*. London: Save the Children; 2002:1–32.
119. Jones B. *The Silent Emergency: HIV/AIDS in Conflicts and Disasters*. London: UK NGO AIDS Consortium; 1999. http://www.reliefweb.int/rw/lib.nsf/db900sid/LGEL-5SED74/$file/ukcon-silent-dec02.pdf?openelement. Accessed October 27, 2009.
120. Anema A, Joffres MR, Mills E, Speigel PB. Widespread rape does not directly appear to increase the overall HIV prevalence in conflict-affected countries: so now what? *Emerg Themes Epidemiol*. 2008;5:11.
121. Supervie V, Halima Y, Blower S. Mass rape in conflict-affected countries could substantially increase the incidence of HIV: implications for prevention and intervention.

Paper presented at 6th International AIDS Society Conference on HIV Pathogenesis, Treatment and Prevention; July 19–22, 2009; Cape Town, South Africa.
122. Kalichman SC, Simbayi LC. Sexual assault history and risks for sexually transmitted infections among women in an African township in Cape Town, South Africa. *AIDS Care*. 2004;16:681–689.
123. Hoffman S, O'Sullivan LF, Harrison A, Dolezal C, Monroe-Wise S. HIV risk behaviors and the context of sexual coercion in young adults' sexual interactions: results from a diary study in rural South Africa. *Sex Transm Dis*. 2006;33:52–58.
124. Wingood MG, Diclemente RJ. Child sexual abuse, HIV sexual risk, and gender relations of African American Women. *Am J Prev Med*. 1997;13:380–384.
125. Dunkle KL, Jewkes RK, Brown HC, et al. Transactional sex among women in Soweto, South Africa: prevalence, risk factors and association with HIV infection. *Soc Sci Med*. 2007;59:1581–1592.
126. Saul J, Moore J, Murphy ST, Miller LC. Relationship violence and women's reactions to male- and female-controlled HIV prevention methods. *AIDS Behav*. 2004;8:207–214.
127. Koenig MA, Zablotska I, Lutalo T, Nalugoda F, Wagman J, Gray R. Coerced first intercourse and reproductive health among adolescent women in Rakai, Uganda. *Int Fam Plan Perspect*. 2004;30:156–163.
128. Wingood GM, Diclemente RJ. The effects of an abusive primary partner on the condom use and sexual negotiation practices of African-American women. *Am J Public Health*. 1997;87:1016–1018.
129. Lang DL, Salazar LF, Wingood GM, Diclemente RJ, Mikhail I. Associations between recent gender-based violence and pregnancy, sexually transmitted infections, condom use practices and negotiation of sexual practices among HIV-positive women. *J Acquir Immune Defic Syndr*. 2007;46:216–221.
130. Díaz-Olavarrieta C, García SG, Feldman BS, et al. Maternal syphilis and intimate partner violence in Bolivia: a gender-based analysis of implications for partner notification and universal screening. *Sex Transm Dis*. 2007;34:S42–S46.
131. Temmerman M, Ndinya-Achola J, Ambani J, Piot P. The right not to know HIV test results. *Lancet*. 1995;345(8955):969–970.
132. Maman S, Mbwambo J, Hogan NM, Kilonzo GP, Sweat M. Women's barriers to HIV-1 testing and disclosure: challenges for HIV-1 voluntary counseling and testing. *AIDS Care*. 2001;13:595–603.
133. Go VF, Sethulakshmi CJ, Bentley ME, et al. When HIV prevention messages and gender norms clash: the impact of domestic violence on women's HIV risk in slums of Chennai, India. *AIDS Behav*. 2003;7:263–372.
134. Ratner PA. The incidence of wife abuse and mental health status in abused wives in Edmonton, Alberta. *Can J Public Health*. 1993;84:246–49.
135. Ellsberg M, Caldera T, Herrera A, Winkvist A, Kullgren G. Domestic violence and emotional distress among Nicaraguan women: results from a population-based study. *Am Psychol*. 1999;54:30–36.
136. Fikree FF, Bhatt LI. Domestic violence and health of Pakistani women. *Int J Gynaecol Obstet*. 1999;65:195–201.
137. McCauley J, Kern DE, Kolodner K, et al. The "battering syndrome": prevalence and clinical characteristics of domestic violence in primary care internal medicine practices. *Ann Intern Med*. 1995;123:737–746.
138. Golding JM. Intimate partner violence as a risk factor for mental disorders: a meta-analysis. *J Fam Viol*. 1999;14:99–132.
139. Campbell JC, Soeken K. Women's responses to battering over time: an analysis of change. *J Interpers Viol*. 1999;14:21–40.

140. Walker LE. Post traumatic stress disorder in women: diagnosis and treatment of battered women syndrome. *Psychotherapy*. 1991;28:21–29.
141. Silva C, McFarlane J, Soeken K, Parker B, Reel S. Symptoms of posttraumatic stress disorder in abused women in a primary care setting. *J Womens Health*. 1997;6:543–552.
142. Counts DA. Female suicide and wife abuse in cross cultural perspective. *Suicide Life Threat Behav*. 1987;17:194–204.
143. Bergman B, Brismar B. Suicide attempts by battered wives. *Acta Psychiatr Scand*. 1991;83:380–384.
144. Kupoor S. *Domestic Violence against Women and Girls*. Florence, Italy: Innocenti Research Center; 2000. Innocenti Digest Series No.6 June 2000.
145. Sandis E. United Nations measures to stop violence against women. *Ann N Y Acad Sci*. 2006;1087:370–383.
146. Wathen N. MacMillan HL. Interventions for violence against women. *JAMA*. 2003;289:589–600.
147. Sullivan CM, Bybee DI. Reducing violence using community-based advocacy for women with abusive partners. *J Consult Clin Psychol*. 1999;67:43–53.
148. Parker B, McFarlane J, Soeken K, Silva C, Reel S. Testing an intervention to prevent further abuse to pregnant women. *Res Nurs Health*. 1999;22:59–66.
149. Pronyk PM, Hargreaves JR, Kim JC, et al. Effect of a structural intervention for the prevention of intimate-partner violence and HIV in rural South Africa: a cluster randomized trial. *Lancet*. 2006;368:1973–1983.
150. Jewkes R, Nduna M, Levin J, et al. Impact of stepping stones on incidence of HIV and HSV-2 and sexual behavior in rural South Africa: cluster randomized controlled trial. *BMJ*. 2008;7:337.
151. World Health Organization. *Guidelines for the Management of Sexually Transmitted Infections*. Geneva, Switzerland: WHO; 2004. http://www.who.int/hiv/pub/sti/pub6/en/. Accessed October 10, 2009.
152. Centers for Disease Control and Prevention. *Sexually Transmitted Infection Treatment Guidelines*. Atlanta, GA: CDC; 2006. http://www.cdc.gov/std/treatment/. Accessed October 10, 2009.
153. Jaffe P, Suderman M, Reitzel D, Killip S. An evaluation of a secondary school primary prevention program on violence in intimate relationships. *Violence Vict*. 1992;7:129–146.
154. Walther DJ. Wife abuse prevention: effects of information on attitudes of high school boys. *J Primary Prev*. 1986;7:84–90.
155. Krajewski SS, Rybarik MF, Dosch MF, Gilmore GD. Results of a curriculum intervention with seventh graders regarding violence in relationships *J Fam Violence*. 1996;11:93–112.
156. Baker G, Ricardo C, Nascimento M. *Engaging Men and Boys in Changing Gender-Based Inequity in Health: Evidence from Program Interventions*. Geneva, Switzerland: World Health Organization; 2007.

This page intentionally left blank

CHAPTER 20

Issues in Contraception and Abortion: The Debate Rages

Stanley K. Henshaw, PhD

Few issues in public health have fostered as much controversy as contraception and abortion. Passions run high on both sides of the debate; religion, politics, and policy regularly clash. The public interference with one of the most private acts (or not) of conceiving has been a reality in America for decades. Women's reproductive health has been a lightning rod for policy makers, and, in some cases, irrespective of the potential harm or danger to the woman's health. Strong moral sentiments, legal constraints, and economic barriers have restricted or limited many women from obtaining safe and effective methods of controlling their fertility. Obtaining contraceptives and seeking an abortion to prevent an unintended pregnancy continues to be a major problem both in the United States as well as in other countries worldwide. The battle Margaret Sanger waged during the first half of the 20th century seems to be never-ending. Major advances have been made since her time, but the evidence clearly shows that many women still have problems obtaining legal, safe birth control, including abortion.

This chapter provides an epidemiological discussion of the issue and explores some of the barriers that inhibit or even prevent women from controlling their own fertility.

UNINTENDED PREGNANCY

Unintended pregnancies[1] occur in all societies. In developed countries, where couples want about two children on average, women spend most of their reproductive lives trying to avoid pregnancy. Unintended pregnancies are inevitable, given the failure rates of contraceptive methods and the difficulty of using them perfectly at every exposure. Even in countries where the norm is to have as many children as

1. A pregnancy is considered unintended if the woman says that at the time she became pregnant she wanted no more children or did not want to become pregnant until later, or if the pregnancy ended in induced abortion.

possible, unwanted pregnancies occur in circumstances in which sexual activity is proscribed. Compared with other developed countries, unintended pregnancy is particularly high in the United States. In 2001, data from the 2002 National Survey of Family Growth (NSFG) combined with the number of abortions indicate that 49 percent of U.S. pregnancies were unintended.[1]

Many unintended pregnancies come to be wanted, and some women report being happy to find themselves accidentally pregnant. Attitudes toward pregnancy form a continuum from extremely unwanted to welcome, even among women who did not plan to get pregnant. Some women decide whether a pregnancy is wanted only after it occurs and the degree of social support for a birth and its other implications become apparent.

A majority of unintended pregnancies in the United States, however, are unwanted. Data from 2001 show that 48 percent of pregnancies (excluding miscarriages) ended in induced abortion,[1] and undoubtedly, many of the women who continued their pregnancies would have preferred not to give birth.

In 2001, the unintended pregnancy rate in the United States, including unintended pregnancies that miscarried, was 51 pregnancies per 1,000 women age 15 to 44 (see table 20.1). The rate varied widely according to demographic characteristics of the women. The age-group with the highest rate was 18 to 19 years of age, with a rate of 108 per 1,000, meaning that almost 11 percent of these young women became unintentionally pregnant that year. The rate declined with age, reflecting reduced fecundity and more effective contraceptive practice, including sterilization especially among the older women.

The rate of unintended pregnancy was highest among unmarried cohabiting women (138 per 1,000) and also was higher among other unmarried women than among those who were married, in part because of the younger age of unmarried women. The rate was almost three times as high among women with incomes below the federal poverty level as among those with income twice the poverty level, and it was strongly associated with low education. Non-Hispanic black women were almost three times as likely as non-Hispanic white women to have an unintended pregnancy, with rates of unintended pregnancy among Hispanic women falling between the two.

These findings understate the problems couples have with unintended pregnancy because they include in the denominator women who are not sexually active, are pregnant or seeking pregnancy, and are infecund or surgically sterilized and thus unable to become pregnant. Rates are much higher when these women are deleted from the denominator. Rates of unintended pregnancy based on women who are actually at risk of unintended pregnancy are shown in the second column of table 20.1. The overall rate was 109 pregnancies per 1,000 women at risk, which means that 11 percent of these women experienced a pregnancy in 2001. Among teenagers, 17 percent became pregnant, as did 16 percent of women ages 20 to 24. The rate of unintended pregnancy was highest among poor women (27 percent), cohabiting women (23 percent), women who had not graduated from high school (22 percent), and black women (22 percent). Overall, 35 percent of births resulted from unintended pregnancies (not shown in table 20.1). The unintended birthrate was 22 per 1,000 women and ranged as high 53 per 1,000 women ages 18 to 19, 54 per 1,000 unmarried women, and 58 per 1,000 poor women.

Table 20.1
Unintended Pregnancy Rate of All Women and Exposed Women, and Unintended Birth Rate, by Demographic Characteristics, 2001

	Unintended pregnancy rate*	Unintended pregnancy rate among exposed women+	Unintended birth rate*
Characteristic			
All women	51	109	22
Age at pregnancy outcome			
<15	3	n.a.	1
15–19	67	174	34
15–17	40	n.a.	21
18–19	108	n.a.	53
20–24	104	157	46
25–29	71	113	32
30–34	44	87	20
35–39	20	51	6
≥40	6	21	3
Marital status at pregnancy outcome			
Currently married	32	69	20
Unmarried	67	140	24
Cohabiting	138	229	54
Unmarried and not cohabiting	52	117	18
Income as a percentage of the poverty level			
<100%	112	266	58
>100–199%	81	187	35
≥200%	29	58	11
Education‡			
Not HS graduate	76	221	40
HS graduate/GED	54	119	25
Some college	47	91	16
College graduate	26	48	10
Race/ethnicity			
White non-Hispanic	35	73	17
Black non-Hispanic	98	223	35
Other non-Hispanic	48	176	20
Hispanic	78	176	40
Place of residence			
Nonmetropolitan area	47	0	23
Metropolitan area	51	0	22

Sources: Finer and Henshaw 2006; Mosher et al. 2004.
Note:
n.a. = not available.
* Per 1,000 women aged 15–44 or in age group.
‡ Among women aged 20 and older.
+ Per 1,000 women exposed to risk of unintended pregnancy, that is, women who are sexually active, not pregnant or seeking pregnancy, and able to become pregnant.

CONTRACEPTIVE USE

Primary prevention of unintended pregnancy is contraception. The rate of unintended pregnancy depends on the amount of exposure to sexual intercourse and on the extent and effectiveness of contraceptive use. According to NSFG data, 11 percent of sexually active women who were not seeking pregnancy were using no method to prevent conception in 2002 (see table 20.2), an increase from 7 percent in 1995.[2] The 11 percent of women not using any contraceptive method accounted for 52 percent of unintended pregnancies in 2001 and 46 percent of the abortions.[1] As of 2002, sterilization was the method most commonly used in the United States: 24 percent of women at risk of unintended pregnancy[2] used female sterilization, and 8 percent used male sterilization. Next most popular was oral contraception (27 percent), followed by condom use (16 percent). A small proportion of couples used other methods. Young women were most likely to use pills and condoms or no method, while sterilization was most common among women approaching the end of their reproductive years (see table 20.2).

All of the contraceptive methods are fairly to extremely effective when used correctly, but typical use in the United States does not approach clinical effectiveness for methods that require attention on the part of the user. About 8 percent of women can expect to become pregnant on the pill in the first year of use; among condom users, the proportion is 15 percent. Even users of the three-month injectable Depo-Provera have failures because some women find it difficult to return for each injection. Aside from sterilization, the only methods that require no effort by the user are intrauterine devices (IUDs) and implants, which are used by relatively few women, despite the problems couples have with other reversible methods.

So why do women not use contraception when in most cases they clearly do not want to become pregnant? A 2001 survey of 10,683 women having abortions at 100 facilities in the United States asked for reasons of nonuse.[3] The large majority, 83 percent, had used contraception in the past, and many of those who had not were teenagers. Only 6 percent never thought about using a method, so lack of information was not the problem in most cases. Many of the nonusers had only recently stopped using contraception. According to an earlier survey of abortion patients, 59 percent of the nonusers had stopped using within three months of becoming pregnant, including 32 percent who had stopped for only one month.[4] These were women who became pregnant during a gap in contraceptive use. A survey of 1,978 U.S. women found that 15 percent of contraceptive users reported a gap in use while at risk of unintended pregnancy within the past 12 months.[5]

Many of the nonusers having abortions in 2001 (32 percent) said they had not used a method because of problems they had had with contraception in the past, fear of side effects, or other concerns about methods.[6] These women may have stopped using a method because of dissatisfaction but had delayed in starting a new method. Some 28 percent of the nonusers said they did not think they would get pregnant at that time, and 26 percent, possibly some of the same women, had not expected to have sex at the time. In some cases, they may have taken a calculated

2. Here "risk of unintended pregnancy" includes contraceptive sterilization to reflect the measures taken to prevent pregnancy.

Table 20.2
Percent Distribution of Sexually Active Women Not Seeking Pregnancy by Contraceptive Method, According to Age, and Use-Failure Rate, United States, 2002

Method	15–44	15–19	20–24	25–29	30–34	35–39	40–44	First-year failure rate in typical use (percent)
Female sterilization	24.1	–	3.2	13.6	24.9	37.2	45.7	0.5
Male sterilization	8.2	–	0.7	3.7	8.4	12.7	16.7	0.15
Pill	27.2	44.2	46.0	33.7	28.6	16.8	10.0	8
Implant, Lunelle, or patch	1.2	1.1	1.3	2.2	1.2	0.6	0.3	0.05–8
Three-month injectable	4.8	11.6	8.8	5.8	3.8	1.9	1.4	3
Intrauterine device (IUD)	1.9	0.3	1.6	3.3	2.9	1.3	1.1	0.2–0.8
Condom	16.0	22.5	20.2	18.4	15.5	14.1	10.5	15
Periodic abstinence	1.3	–	1.2	0.9	1.4	1.8	2.1	25
Withdrawal	3.6	–	4.5	7.0	3.4	3.1	1.3	27
Other	1.2	2.1	0.4	0.9	0.7	0.6	2.0	8–29
No method	10.7	18.3	12.1	10.5	9.2	9.8	8.8	85

Source: Mosher, WD et al. Centers for Disease Control and Prevention. National Center for Health Statistics. Use of Contraception and Use of Family Planning Services in the United States: 1982–2002. Advance data. 2004; (350); 1–36.

risk. Twelve percent had problems obtaining contraception, including 8 percent who had financial problems, 2 percent who did not know where to get a method, and 2 percent who ran out and neglected to get a prescription filled. Ten percent had partners who did not want to use a method.

Of the 54 percent of abortion patients who had been using contraception during the month they became pregnant, 51 percent had been using condoms, 25 percent oral contraceptives, 14 percent withdrawal, 4 percent periodic abstinence, and 6 percent other methods. Fourteen percent of the condom users said they had used the method correctly at every exposure, 49 percent said they had not used the method consistently, and 42 percent said a condom had slipped or broken. (Some said both that their use had been inconsistent and that a condom broke or slipped.) The most frequent reasons for inconsistent use were believing the risk of pregnancy was low (41 percent), not having a condom available (29 percent), and not expecting to have sex (26 percent). Twelve percent cited attitudes and behavior of partners as reasons for inconsistent use.[3]

It is standard practice for primary care providers, obstetrician-gynecologists, and abortion providers to discuss contraception with their female patients of reproductive age and to encourage effective use. More emphasis on counseling seems difficult to achieve and unpromising. Evaluation studies of programs to provide more intensive contraceptive counseling of abortion and family planning clients have found that they have little impact on subsequent contraceptive use or unintended pregnancies. Sterilization is already the most common method used in the United States, and most unintended pregnancies occur to women who want children in the future or are not sure they will never want more children.

An approach to reducing unintended pregnancy that has been gaining support is to encourage use of long-acting reversible contraceptive methods. In the United States, three such methods are approved for use: a copper-T IUD (ParaGard), a levonorgestrel IUD (Mirena), and an implant (Implanon). These methods have extremely low first-year failure rates (0.8 percent, 0.2 percent and 0.05 percent, respectively), so virtually all the unintended pregnancies caused by incorrect and inconsistent use of reversible methods are prevented. In addition, many of the pregnancies that occur in gaps between use are prevented because women are less likely to discontinue the methods. Unlike other methods, discontinuation requires a visit to a clinician and cannot occur by sheer inertia. Tabulations of the National Survey of Family Growth show that about half of women are still using their IUDs after three years, and other studies have found that about four-fifths of women are using these methods after one year, a higher percentage than are still using other reversible methods.[7] One study, for example, found that only 29 percent of U.S. women started on the pill consistently refilled their prescriptions for 12 months.[8]

Long-acting reversible methods have significant potential to reduce the rate of unintended pregnancy in the United States, where only about 2 percent of women seeking to avoid pregnancy were using them in 2002. Although their use has increased in recent years, the rate is undoubtedly still much lower than in most other developed countries and many developing countries. IUD use is as high as 27 percent in Norway. More typical are the Baltic states (14 percent), France (17 percent), Germany (10 percent), and the United Kingdom (11 percent).[9]

In the United States, there are two major barriers to IUD use. The first stems from the attitudes and practices of health care professionals. Many providers have

reservations about the method because of experiences with the Dalcon Shield, which caused pelvic inflammatory disease (PID) and infertility for some users. At the time the problem was discovered in the 1970s, all IUDs were considered risky, and at one point in the 1980s, none were on the market in the United States. Fear of IUDs among health professionals as well as the public is still widespread despite research demonstrating that the infection risk was limited to the Dalcon Shield.

Many professionals are unaware of recent research that has found IUDs to be appropriate for patients who were excluded in the past. The WHO eligibility criteria classify the benefits of IUDs as outweighing the risks for teenagers, nulliparous women, and women with a history of PID.[10] They also permit insertion immediately postabortion and postpartum, although expulsion rates may be slightly higher. However, a survey of family planning clinicians in California found that a majority considered an IUD inappropriate in these cases.[11] They also found that 39 percent of the respondents do not offer IUDs at all. The reasons most frequently cited are inadequate reimbursement for insertion, lack of training, low patient interest, the risk of the procedure, and the risk of litigation. Among all respondents, 40 percent were not "very comfortable" inserting the ParaGard and 60 percent were not "very comfortable" inserting the Mirena.

Many protocols make it difficult or impossible to insert an IUD on the same day as an abortion or family planning visit. When negative sexually transmitted infection (STI) screening results are required, clients usually have to return on another day when test results are available. Reports indicate, however, that approximately 35 percent of women who said that they prefer an IUD fail to return for insertion.[12] One study has found that same-day insertion with follow-up treatment for women whose STI tests come back positive results in no higher a rate of pelvic infection than the more conservative protocol with insertion only after a negative STI test.[13]

Cost is the second major barrier to most contraceptive methods, and especially long-acting reversible ones, in countries where insurance coverage is inconsistent and in developing countries where family planning services may be underfunded. In the United States, many women lack health insurance and others have insurance that excludes coverage of some contraceptive methods, such as the IUD. A woman without insurance must pay an estimated $588 including insertion for a copper-T IUD, $679 for a levonorgestrel IUD, and $749 for a hormonal implant.[14] Although these methods are cost-effective over the long run when compared with other methods, their upfront cost is a significant obstacle for most women. Some physicians discourage use of long-acting methods because women who have them removed after a short time will have wasted so much money. Oral contraceptives, which can cost as much as $53/month plus the charge for an office visit, are also expensive for self-paying women in the United States.[14] Manufacturing the devices and pills costs less than 1 percent of the usual retail price. In contrast, the health plans of most other developed countries cover the cost of contraceptive services and supplies.

Women who have insurance also face barriers. Private plans usually require copayments that may be substantial. For example, some plans charge $5 per month for drugs. The Mirena levonorgestrel-releasing IUD is classified as a drug, and since it is approved for five years, the copayment is $5 times 60 months, or $300. Some plans require the woman to obtain a prescription from a clinician, purchase the device at a pharmacy, and return to the clinician for insertion. Providers who do not

offer the method may feel that insurance reimbursement rates are too low. The women at greatest risk of unintended pregnancy are the ones most likely to be deterred by these requirements and financial barriers.

Insurance procedures also affect women who use oral contraceptives. Most insurance plans allow the purchase of only one month's supply of pills at a time, which means that a woman has a narrow window in which to return to the pharmacy each month. Discontinuation and gaps in use frequently result. A study in California found that women who received 13 cycles of pills at one visit were less likely to have a pregnancy test over the next year and more likely to obtain enough pills for 15 months of continuous use than were women who received three cycles or one cycle of pills. The study also found that dispensing 13 cycles of pills at one time was less expensive for the insurance plan, even though some pills were wasted.[15]

Although IUD insertion at the time of an abortion is accepted practice, few abortion providers do so. In most cases, women's insurance does not cover contraceptive services provided by an abortion clinic, or the clinic cannot be sure it will be reimbursed. When the abortion is covered by insurance, a second procedure (IUD or implant insertion) will not be covered on the same day or will be covered at a reduced rate. Red tape is involved for a clinic to stock IUDs because distributors will not sell them directly to physicians for resale to patients.

Overcoming the obstacles to IUD insertion at the time of abortion would have important benefits in reducing unintended pregnancies and abortions. Women having abortions are one of the highest-risk groups for additional unintended pregnancies and abortion. A study that offered same-day IUD insertion to all abortion patients conducted a record review to see how many had subsequent abortions in the following 12 months. Almost 10 percent of patients who selected a contraceptive method other than an IUD had a second abortion in the same facility within a year, compared with about 4 percent of patients who received an IUD at the time of the abortion. The pregnancies among the latter group occurred after the IUDs were removed or expelled.[16] Even more unintended pregnancies were undoubtedly prevented in the second and third years after the abortion.

Unintended pregnancy occurs in all countries but is especially prevalent in the United States. Primary prevention could be improved by reducing barriers to contraception and encouraging use of long-acting reversible methods.

ABORTION

In all societies, some women rely on abortion to resolve unwanted pregnancies. Descriptions of methods of inducing abortion have been found in writings as old as the 28th century BCE from ancient Egypt, China, and Rome. Even in traditionally high-fertility cultures where becoming pregnant is normally a cause for celebration, women still need and obtain abortions under some circumstances, for example, when pregnancy occurs under disapproved circumstances such as outside marriage or when pregnancy would threaten the woman's health. Abortion was common in some traditional cultures—for example, among the Siraya, an Austronesian society in Taiwan in the 17th century, where women married as teenagers but were forbidden to have children until they were in their 30s. All pregnancies before age 30 were

ended by abortion, which was done by massage.[17] In some other cultures, abortion is highly disapproved.

Among the countries of the world, the legal status of induced abortion ranges from complete prohibition to elective abortion at the request of the pregnant woman.[18] At of the end of 2007, 40 percent of the world's population lived in the 56 countries where abortion was permitted without restriction as to reason. These countries include China, the countries of the former Soviet Union, the United States, Canada, and many European countries. Some of these countries have gestation limits beyond which they permit abortion only for health or other reasons. Fourteen countries containing 20 percent of the world's population allow abortion only for health or socioeconomic reasons such as inadequate income, substandard housing, unmarried status, and contraceptive failure. Great Britain, India, and Japan are in this category. Threat to a woman's mental or physical health, which might or might not include socioeconomic conditions insofar as they affect health, is legal justification for abortion in 23 countries with 4 percent of the world's population.

Ten percent of the world's population lives in the 35 countries that permit abortion only if the woman's physical health is threatened by the pregnancy, and 26 percent live in the 68 countries that permit abortion only if the woman's life is endangered or prohibit abortion completely. These include most Muslim countries (Tunisia and Turkey, which allow abortion on request, are notable exceptions), almost two-thirds of Latin American countries, and a majority of the countries of Africa. Some of these countries also allow abortion in cases of rape and fetal abnormality.

Several of the above restrictions are subject to widely varying interpretations, depending on the country. A statute authorizing abortion to avert a threat to the pregnant woman's mental health may be interpreted strictly or may allow most women to obtain abortions. Social indications usually are interpreted broadly to allow almost any woman to terminate a pregnancy within gestation limits, as in Great Britain, India, and Japan.

The abortion statutes of some countries are not strictly enforced, and occasional abortions on medical grounds are quietly tolerated in almost all countries. It is well known that in some countries with restrictive laws, abortions can be obtained openly and without interference from the authorities when performed by private physicians, as in South Korea and parts of South America. On the other hand, in restrictive countries that allow abortion for rape and life endangerment, it may be impossible for women to terminate pregnancies legally even for these reasons because of physicians' ideological opposition, ignorance of the law, or fear of prosecution.

Abortion was highly restricted in almost the entire world in the first half of the 20th century. Then starting in the 1950s, a wave of liberalization swept the developed world and much of the developing world. The trend toward liberalization has continued up to the present, though at a slower pace. Between 1998 and 2007, 16 countries liberalized their laws while two increased their restrictions. During this period, abortion became legal on request in Nepal, Portugal, Switzerland, and parts of Australia and Mexico. No country imposed restrictions where none existed previously.[18]

Legal restrictions, however, have had limited success in eliminating abortions. The abortion rates in Africa and Latin America, where abortion is predominantly illegal, are estimated at 29 and 31 per 1,000 women age 15 to 44, respectively, whereas the rates in Europe and North America, where abortion is generally

legally available, are similar or lower at 28 and 21 per 1,000 women age 15 to 44, respectively.[19] However, many of the areas where abortion is highly restricted are developing countries, which tend to lag in use of contraception.

Abortion rates vary widely among the countries where the procedure is legal. The highest rate ever recorded in a country with complete statistics was 252 abortions per 1,000 women between the ages of 15 and 44 in Romania in 1965. In a sense, abortion became one of the more common forms of birth control in this country. The lowest rate where abortion was unrestricted as to reason and services were readily available was 5.1 per 1,000 in the Netherlands in 1989.[20] Many of the abortions in the Netherlands were obtained by immigrants; the abortion rate among Dutch-born women was 3.5 per 1,000 in 1992.

Abortion rates range from around 20 to 21 per 1,000 women in the countries with the highest rates, including Australia, Sweden, and the United States; to 15 to 17 per 1,000 women in countries such as Canada, England and Wales, and France with intermediate rates; and to 7 to 11 per 1,000 women in countries with the lowest rates, including Belgium, Germany, and Italy.[21] Rates are higher in Russia and in most of the other former Soviet Union countries. Data from household surveys place current rates highest in Georgia (135 per 1,000 women), Azerbaijan (116 per 1,000 women), and Armenia (81 per 1,000 women). Rates in other developing countries where abortion is legal may be high, as in Cuba (57 per 1,000 women) or low, as in Tunisia (7 per 1,000 women).

In the United States, there were 1,206,000 abortions in 2005 (a rate of 19.4 per 1,000 women ages 15 to 44 years). In that year, 22.4 percent of pregnancies, excluding miscarriages and stillbirths, were terminated by abortion. Although the U.S. rate is at the high end of Western Hemisphere countries, it has fallen steadily since 1981, when it was 29.3 per 1,000 women.[6] Rates also have fallen in recent years in most Eastern European and former Soviet Union countries. No consistent trend has emerged in other developed countries.

EPIDEMIOLOGY OF ABORTION IN THE UNITED STATES

A woman's probability of having an abortion closely follows the pattern of unintended pregnancy as shown in table 20.1. The highest abortion rate, 40 abortions per 1,000 women in the age-group, occurred among women 20 to 24 years old (see table 20.3). Approximately 18 percent of abortions were obtained by teenagers, with the abortion rate considerably higher among women ages 18 to 19 than among younger teenagers. The abortion rate has fallen significantly among teenagers in recent years, more than among older women.[22] Yet, if one looked at the data in a slightly different way, using the ratio of abortions to births plus abortions (not shown) a somewhat different pattern emerges. In this case, the abortion ratio is high among teenagers; approximately 33 percent of pregnancies (excluding miscarriages) are terminated by abortion. The abortion ratio declines to 16 percent at ages 30 to 34, and then rises with age to 26 percent.[22]

Non-Hispanic white women account for 33 percent of U.S. abortions, but their abortion rate is relatively low, 11 per 1,000 in 2004, which is in the mid-range of the rates in other Western Hemisphere countries. The rate among black women (52 per 1,000 women) is more than four times as high, and that of Hispanic women (28 per

Table 20.3
Percentage Distribution of Abortions and Abortion Rate per 1,000 Women Ages 15 to 44, by Selected Characteristics, United States, 2000[a]

Characteristic	Percent	Rate Population
Total[a]	100	20
Age years[a]		
<15[b]	1	3
15–17	6	12
18–19	11	32
20–24	33	40
25–29	23	30
30–34	15	18
35–39	8	10
≥ 40[c]	3	3
Race/ethnicity[a]		
White non-Hispanic	33	11
Black non-Hispanic	37	52
Other non-Hispanic	8	23
Hispanic	22	28
Education[d]		
Not HS graduate	13	23
HS graduate/GED	30	20
Some college	41	26
College graduate	16	13
Marital status		
Married	17	8
Previously married	16	29
Never married	67	35
Cohabiting[e]		
Unmarried, cohabiting	25	55
Unmarried, not cohabiting	58	29
Number of live births		
0	39	19
1	27	32
2 or more	34	18
Religion[f]		
Protestant	43	18
Catholic	27	22
Other	8	31
None	22	30
Income as a percentage of the poverty level		
$<100\%$	27	44
100–199%	31	38
200–299%	18	21
$\geq 300\%$	25	10
Has Medicaid coverage		
Yes	24	57
No	76	18

(Continued)

Table 20.3 *(Continued)*

Characteristic	Percent	Rate Population
County of residence		
Metropolitan	88	24
Non-metropolitan	12	12
Place of residence		
Nonmetropolitan area		
Metropolitan area		

Sources: Jones, Darroch, and Henshaw 2002, and Ventura et al. 2008.
Note:
a. Total, age, and race/ethnicity are for 2004.
b. Denominator is women age 14.
c. Denominator is women ages 40–44.
d. Among women aged 20 and older.
e. Based on single women only.
f. Limited to women over 17.

1,000 women) lies between that of blacks and whites. Among both black and Hispanic women, low income and high rates of unintended pregnancy explain the high abortion rates. Asian and Native American women have a moderate abortion rate; this category includes a mixture of ethnic groups with widely varying abortion patterns.

Occasional large-scale surveys of abortion patients have provided demographic information that is not available from abortion reporting forms. The most recent study was conducted by the Guttmacher Institute in 2000 and 2001 and involved 10,683 abortion patients in 100 U.S. abortion facilities.[3] Patients were asked to complete self-administered questionnaires at the time of the abortion. Most of the information in the rest of this section was taken from these survey results. Although this information is several years old, the demographic characteristics of women having abortions are relatively stable over time.

Women of all education levels have occasion to seek abortion services, but college graduates have a lower abortion rate than less-educated women. Approximately 41 percent of abortions are obtained by women with some post–high school education, but who are not college graduates. Never-married women obtain the majority of abortions (67 percent); married women account for only 17 percent. The abortion rate is higher among never-married women (35 per 1,000 women) than among previously married (29 per 1,000 women) or currently married women (8 per 1,000 women). The high rate among never-married women results partly from their young age compared with the other marital groups. Women living with a partner to whom they are not married account for 26 percent of abortions but only about 10 percent of women in the population. Their abortion rate is almost two times that of other unmarried women. Thus, cohabiting is one of the strongest risk factors for abortion as well as for unintended pregnancy.

A majority (61 percent) of women having abortions have had at least one birth, and a third have had two or more. When age is taken into account, women who have children are substantially more likely than women without children to have an

abortion, and the highest abortion rate is found among women with four or more children. Such women may have difficulty using contraception and thus may have unplanned children as well as abortions. A large majority of abortion patients with children are unmarried (76 percent), and more than half (56 percent) of the unmarried abortion patients have children.

The abortion rate among Protestants (18 per 1,000 women) in 2000 is somewhat lower than that of all women (21 per 1,000 women), while that of Catholics (22 per 1,000 women) is about the same. Women of other religions, including Russian and Greek Orthodox as well as Islam and others, and those who claim no religious identification appear to have somewhat higher abortion rates (30 to 31 per 1,000 women). The rate among women who name no religion is somewhat uncertain, because answers to questions on religious identification vary according to the context and wording of the question.

The higher rate of Catholics compared with Protestants is confirmed by a comparison of their rates after excluding women from minority groups with high abortion rates, namely blacks and Hispanics. As expected, excluding black and Hispanic women reduces the abortion rates of both Protestants and Catholics but that of Catholics remains higher. Possible reasons for the higher rate among Catholics include that Catholics use less effective methods of contraception, are more opposed to childbearing outside of marriage, and are concentrated in cities and geographic areas with high abortion rates.

Household income is strongly associated with abortion utilization. Women whose income is below the federal poverty level are more than four times as likely to have an abortion as are those with income three or more times the poverty standard. The high relative abortion rate of low-income women is reflected in the abortion rate among those covered by Medicaid, the government program that provides health insurance for many low-income women. Twenty-four percent of patients say they have Medicaid coverage (although in most states it will not pay for abortion services), while only 9 percent of all U.S. women of reproductive age are covered (as of the year 2000). Thus, the abortion rate of women with Medicaid coverage is three times as high as that of other women.

Women covered by Medicaid have a number of attributes that may contribute to their relatively high risk of abortion: they are disproportionately nonwhite, unmarried, and poor, all characteristics associated with high abortion rates. In addition, many women on Medicaid are covered by that program because of a prior accidental pregnancy that they carried to term and are prone to unplanned pregnancy. The abortion rate of women who live in nonmetropolitan counties is half that of metropolitan women. Nonmetropolitan women have less access to abortion services, and they may be more careful to avoid unwanted pregnancy or are more reluctant to have abortions.

Data show that repeat abortions are common. In 2005, almost half (45 percent) of abortions are obtained by women who have had one or more prior abortions.[23] Women who have had an abortion are at high risk of additional unintended pregnancies and abortions because they are sexually active, probably above average in fecundity, have difficulty using contraception, are likely to be in demographic groups with high rates of unintended pregnancy, and are willing to terminate unwanted pregnancies by abortion.

GESTATIONAL AGE, PROCEDURE, AND SETTING

More than half (62 percent) of all induced abortions in the United States occur at eight weeks or earlier, counting from the first day of the last menstrual period (LMP) or two weeks before the estimated date of conception. This proportion has increased in recent years because of greater acceptance of early abortion by medication (usually mifepristone followed by misoprostol) and early vacuum aspiration. Approximately 11 percent of abortions are performed past 12 weeks, including 1.4 percent past 20 weeks.[23] In most developed countries other than England and Wales, somewhat fewer abortions take place after 12 weeks, probably because women respond more promptly to unwanted pregnancies and because restrictions in some countries make later abortions more difficult to obtain.[20] An additional factor is that most other countries provide universal health insurance that covers abortion services. In contrast, women in the United States may be delayed by difficulty gaining access to abortion services and acquiring money to pay for the procedure.[3, 24]

In all countries with relevant statistics, teenagers obtain abortions later in gestation on average than do older women. In the United States in 2005, 26 percent of abortions obtained by women under age 15 were past 12 weeks LMP as were 16 percent among women 15 to 19 years old, compared with 10 percent among women age 20 and over. Abortions generally occur earlier with age until age 39, after which a few women are delayed because they mistake pregnancy for the menstrual changes of menopause.[23] The delay among younger women probably reflects their inexperience in recognizing the symptoms of pregnancy, their reluctance to accept the reality of their situation, their lack of knowledge of where to seek advice and services, and their hesitation to confide in adults. In addition, teenagers may have more difficulty paying for abortions, and minors may be affected by parental consent or notification requirements.

Approximately 88 percent of abortions in the United States in 2005 were accomplished by suction or sharp curettage (primarily suction), which includes dilation and evacuation. Before nine weeks LMP, vacuum aspiration represented the most frequently used method, although mifepristone or methotrexate followed by a prostaglandin is increasingly common and accounts for at least 15 percent of these early abortions.[23] The proportion may be higher; according to the Guttmacher survey of abortion providers, 161,000 early medical abortions were provided in 2005, accounting for about 21 percent of abortions before nine weeks.[3] Provision of suction abortion before seven weeks LMP represents another recent trend, because use of sensitive pregnancy tests and vaginal ultrasound have reduced the risk of failing to end an early pregnancy or to detect an ectopic pregnancy. The proportion of U.S. abortions occurring before seven weeks LMP increased from about 16 percent in 1995 to 30 percent in 2005.[23]

Of U.S. abortions past 12 weeks LMP, the vast majority (96 percent) are performed by dilation and evacuation. Even after 20 weeks, this method was used for 85 percent of abortions in the United States.[23] In many other developed countries, second-trimester abortion is commonly performed by labor induction, using prostaglandin administered by various routes.

Both first- and second-trimester abortions can be provided safely outside of hospitals, in clinics and physicians' offices.[25, 26] The proportion of U.S. abortions

performed in hospitals has declined from more than half in 1973 to 5 percent in 2005.[3] The number of hospitals where abortions are performed has dropped sharply, as has the average number of abortions per hospital provider. A tabulation of data on approximately 300,000 abortions in 14 states in 1992 indicates that, even after 20 weeks, 83 percent were performed outside of hospitals. Near universal agreement as to the safety of second-trimester abortion outside of hospitals is further demonstrated by the finding that in 2001 about 55 percent of abortion clinics offered the service at 18 weeks or later.[27] In most European countries other than Great Britain, however, most second-trimester abortions are performed in hospitals, which often are considered the appropriate setting for labor induction.

PUBLIC HEALTH IMPACT OF LEGALIZATION OF ABORTION IN THE UNITED STATES

In the United States, abortion on request was legalized in Alaska, Hawaii, New York, and Washington in 1970, then nationally by the Supreme Court decision in *Roe v. Wade* in 1973. These changes had consequences for the number of pregnancy terminations, maternal mortality and morbidity, and child health. Before 1970, although abortions were restricted in all states, illegal abortions were common. Women with resources and connections were able to obtain relatively safe abortions from physicians. Although most abortions were performed illegally, some women were able to have their pregnancies terminated in hospitals by obtaining a certification from psychiatrists and hospital committees that the pregnancy posed a danger of suicide. Other women went to Puerto Rico[3] or Europe for abortion services. Women without resources, however, had to rely on primitive and generally unsafe methods such as inserting objects into their uteruses or seeking help from individuals without medical training. Many serious complications and even deaths resulted.

A number of studies have tried to estimate the number of abortions that took place annually before 1970. A study based on the change in the birth rate in New York City after 1970 concluded that 70 percent of the legal abortions in 1971 replaced abortions that would have taken place anyway illegally.[28] From a survey in North Carolina in 1967, researchers estimated that 829,000 abortions were occurring annually in the country as a whole, which is about 80 percent of the number of legal abortions that occurred in 1975, when legal abortion services were available in all states.[29] A study that analyzed the change in number of births nationally concluded that roughly 600,000 abortions would have occurred without legalization.[30]

In the United States as well as in other countries, legalized abortion reduced mortality from induced abortion. Over the decade from 1958 through 1967, more than 3,400 women died from induced abortion, almost all of which resulted from illegal abortion.[31][4] The number of abortion deaths rose during the 1950s and reached at least 430 in 1961, then declined during the 1960s partly because more

3. Abortion was illegal in Puerto Rico but was readily available from physicians.
4. The numbers taken by Hilgers from the National Center for Health Statistics have been adjusted to reflect deaths *associated* with abortion as well as those *attributed* to abortion, to make the results comparable to number of deaths associated with abortion as reported by the Centers for Disease Control and Prevention for 1972 and later years.

physicians became willing to terminate pregnancies. Legalization of abortion resulted in a sharp drop in the number of deaths, from about 251 in 1966 to 14 in 1976, a reduction of 94 percent. In recent years, the number of deaths has ranged from 4 to 11 per year according to the Centers for Disease Control and Prevention.[23]

To the extent that legal abortion replaced childbirth, it also reduced maternal deaths because it was safer than continued pregnancy and delivery. During the period 1991–1999, mortality associated with live birth was 7.1 deaths per 100,000 live births (excluding deaths from miscarriage and ectopic pregnancy, which might not be prevented by abortion), compared with 0.6 deaths per 100,000 induced abortions.[32] One researcher calculated that between 1970 and 1980, on the order of 1,500 women's lives were saved by the replacement of unwanted and mis-timed births and illegal abortions by legal abortions.[33] The saving might have been greater since after legalization of abortion many more women with serious medical conditions were able to terminate potentially life-threatening pregnancies.

For each death from illegal abortion, many other women also suffered complications. Complete statistics on the number of such complications during the 1960s are unavailable, but hospital studies found marked decreases in the number of women treated for abortion complications. For example, in municipal hospitals in New York City, for each 1,000 births there were 234 admissions for incomplete spontaneous and induced abortions in 1969 and 130 such admissions in 1971 after the abortion restrictions were repealed. At least five studies in other groups of hospitals found similar decreases.[34]

Because legalization made abortion more accessible to the youngest and oldest women of childbearing age and those in poor health, one effect was to reduce the rate of infant mortality and premature birth. A study by economists associated with the National Bureau of Economic Research found that the increase in the abortion rate was the most important factor in explaining the reduction in neonatal mortality between 1964 and 1977. The abortion rate dominated other public policies, including Medicaid, subsidized family planning services, and maternal and infant care projects, in explaining the mortality decline of both white and black infants.[35, 36] Another economist found that abortion also reduced the rate of low birth weight and preterm births.[37]

Prenatal screening followed by abortion when necessary has allowed couples to avoid the birth of children with congenital abnormalities. The result is a marked decrease in the number of such births. The number of children born in 2001 with Down syndrome was 3,654, about half the number that would be expected in the absence of screening, 7,262.[38] Many women choose to continue pregnancies when abnormalities are identified, but for some of those who end such pregnancies, the birth of a disabled child would be devastating psychologically, socially, and economically.

Prenatal screening with the option of abortion has allowed couples who know they are at risk of giving birth to children with genetic abnormalities to have children without fear that they will be unhealthy. For example, each child of couples who are carriers of Tay-Sachs disease has a 25 percent chance of having a genetic abnormality that causes death by age five. With prenatal testing and abortion, these couples can have a full and healthy family and avoid having a child who inevitably will die at an early age. Without the availability of prenatal testing and second-trimester abortion, many such couples would choose not to have children.

To the extent that abortion, like contraception, allows couples to have children when they are wanted, the children benefit. Studies of children born after abortions were denied have found that compared with wanted children, those born unwanted are more likely to have sought and received mental health care later in life and to have engaged in criminal behavior.[39–41] Other benefits of legal abortion (found after 1970–1973) were that fewer infants were abandoned in hospitals,[42] infant homicide was reduced,[43] and fewer children were born to unmarried mothers.[30]

Reviews of the numerous studies of the effects of first-trimester vacuum aspiration abortion on a woman's future reproductive capacity conclude that no risk has been demonstrated.[44, 45] Rates of infertility, ectopic pregnancy, miscarriage, and congenital malformation are similar among women who have had an abortion and women who have not. Among studies of preterm delivery and low birth weight, a minority have found an association with abortion history, but these have not controlled for all possible confounding factors such as having a subclinical gynecological infection at the time of delivery.

Studies of women's mental health immediately after abortion and one or two years later have found little difference from that of women generally.[46, 47] A minority of women has adverse reactions after unintended pregnancy ending in abortion, but no evidence indicates that such reactions are more common than among women who give birth after being unable to terminate an unwanted pregnancy. Few studies have been able to compare women who had abortions with women who were refused abortions. An exception is a large study in Britain, which found no difference between the two groups in the rate of psychiatric morbidity.[48] A task force of the American Psychological Association recently concluded that "the best scientific evidence indicates that the relative risk of mental health problems among adult women who have an unplanned pregnancy is no greater if they have an elective first-trimester abortion than if they did deliver the pregnancy."[49]

BARRIERS TO ABORTION SERVICES

Abortion services in most developed countries are easily accessible, but in the United States, despite the large number of women who need abortion care, abortion services continue to be less available than for other common medical conditions. A significant but unknown number of women in the United States continue unwanted pregnancies because of lack of access to an abortion provider. A problem for many women in the United States is distance from a provider. In 2005, 8 percent of women who were able to obtain abortion services traveled more than 100 miles to a provider and 19 percent traveled 50 to 100 miles, according to providers' estimates. Ninety-two percent of women who lived in nonmetropolitan counties had no local provider, whereas 37 percent in metropolitan areas had none. The number of facilities where abortions are performed has been declining since 1982, and fell from 2,042 in 1996 to 1,787 in 2005.[6]

Cost also represents a barrier for women of limited means, who constitute a majority of abortion patients. In 2001, 57 percent of women obtaining abortions had family income less than two times the federal poverty level.[3] In that year, 74 percent of all women having abortions paid in cash; of the remainder, half had Medicaid coverage in states where Medicaid pays for abortion, and half had private insurance that was accepted by the provider.[27] In states where Medicaid does not

cover abortion, 91 percent of women paid out of pocket. The average charge for an abortion at 10 weeks in 2005 was $413.[3]

Harassment by antiabortion activists adds to the difficulty women experience in accessing abortion services and the challenges of providing services. In 2000, 80 percent of large nonhospital facilities (400 or more abortions a year) in the United States experienced antiabortion picketing.[27] Picketing and antiabortion violence impede access to women who might be intimidated by aggressive protesters. The stigmatization of abortion undoubtedly affects many women, although this factor is difficult to measure. Legal restrictions on abortion also impede access to abortion, especially in the United States where states regulate abortion services within constitutional limits defined by the Supreme Court.

Following are the major restrictions that are now in effect in some or most states.

Limited Medicaid Coverage for Abortion

The "Hyde Amendment," in effect in various forms since 1977, prohibits the use of federal Medicaid funds for abortions except in cases of life endangerment, rape, or incest. Although 17 states have policies to pay for medically necessary abortions with state Medicaid funds,[50] a majority of low-income women needing abortions live in states where public financial support for abortion is unavailable. Private charities provide assistance to some women in these states. Most European countries have national health plans that provide free or low-cost abortion care for all or most women.[51]

The best studies of the impact of Medicaid abortion restrictions indicate that about one-fourth of women who would have Medicaid-funded abortions instead continue their pregnancies when this funding is unavailable.[52] One of these studies examined abortion and birth rates in North Carolina, where the legislature created a special fund to pay for abortions for poor women. In several years, the fund was exhausted before the end of the fiscal year, so financial support was unavailable to women whose pregnancies occurred after that point. The analysis found that about one-third of the women who would have had abortions when support was available instead carried their pregnancies to term when the abortion fund was unavailable.[53] A study in Michigan, Ohio, and Georgia found that 18 to 22 percent of women who would have had abortions continued their pregnancies after Medicaid funding was cut off.[54] A study done in Texas found that 35 percent of women continued their pregnancy; in Illinois, the figure was 24 percent.[55]

One can assume that women who are unable to overcome the barrier posed by lack of financial support for an abortion are those with the fewest financial resources. These are also likely to be women who are not in a favorable position to take responsibility for care and support of children. Women with borderline motivation to prevent childbirth also may be affected by lack of financial support for abortion. The ban on public funding for abortion appears to be the restriction with the greatest impact on abortion accessibility.

Parental Notification or Consent for Minors

European countries vary in their parental consent policies and vary in their definitions of "minor." In the United States, 24 states require the consent of one or both

parents, and an additional 11 states require parental notification before an abortion can be performed for a minor, defined as a woman under age 18 in all states except Delaware (under age 16) and South Carolina (under age 17).[56] In upholding state parental consent statutes, the U.S. Supreme Court required that an alternative to parental involvement be provided, so that parents do not have an absolute ability to block the abortion. States requiring parental involvement therefore allow a minor to petition a judge for authorization to have an abortion without consulting parents. The petition should be granted if the minor is mature enough to make an informed decision or if the abortion is in her best interest. In all or parts of several states, however, judicial authorization is rarely an option because judges reject all petitions.

Although the ostensible purpose of these laws is to encourage minors to consult their parents before having an abortion, little evidence indicates that this has been accomplished in states with a functioning judicial bypass option.[57] In the absence of any law, about 60 percent of minors say one or both parents are aware that they are having abortions. The proportion is higher among younger teenagers, reaching as high as 90 percent of those under age 14.[58]

A review of studies of the impact of parental involvement laws found a few effects.[59] The clearest impact of the laws is that minors cross state lines to seek abortion services in states with less severe requirements. There was no effect on minors' abortion rates in several states, but in one their abortion rate fell slightly with a corresponding increase in their birth rate. One study found evidence that some 17-year-olds wait until their 18th birthday to end a pregnancy, therefore delaying the abortion to the second trimester. Several studies found that parental involvement requirements caused minors to have their abortions later in gestation, but others identified no effect. Little evidence indicates that the laws resulted in fewer pregnancies to minors.

Although public opinion polls suggest that parental involvement requirements have wide public support, the results of referenda on the issue have been mixed. Parental involvement laws have been rejected by voters three times in California and once in Oregon, while they were narrowly approved in Colorado and Florida.

State-Mandated Counseling and Waiting Periods

An increasing number of states, 23 as of 2009, require that women receive state-directed counseling, that is, that they be given certain specific information before an abortion can be performed. This information frequently includes the risks of abortion and childbirth; a description of fetal development; the assistance available to women for prenatal care, childbirth, and infant care; and a list of agencies that provide information or services designed to help women carry their pregnancies to term. Five states require that women be told that having an abortion may increase the risk of breast cancer, although medical authorities are in agreement that no such relationship exists. Nine states require that women be told that a fetus may feel pain, and seven require that women be given information on the possible negative psychological impacts of abortion.[60]

Twenty states require that the women receive the information 24 hours before the abortion can be performed. Seven of these require that the information be delivered in person, which means that in most cases the woman must make two trips to

the abortion facility. Although proponents of these requirements say they are needed to help women make informed decisions, the implicit purpose is to discourage abortion. Evidence for this is that no similar requirements apply to women seeking prenatal care, although the decision to continue a pregnancy carries more risk and responsibility than the decision to terminate a pregnancy.

Abortion providers report that the required counseling and waiting periods have little effect on women's abortion decisions, although they object in principle to the assumption that women need special constraints when making important decisions and that the state can override the judgment of medical professionals about the information needed by their patients. The two-trip requirement, however, appears to be a significant burden that prevents some women from obtaining desired abortion services. A study of abortion statistics from Mississippi and neighboring states found that approximately 11 percent of the women who would have had abortions were prevented from doing so by a two-trip requirement. In addition, many women went to other states to avoid the requirement, and abortions were delayed on average.[61, 62]

Other Restrictions

In the early period after legalization, some jurisdictions, including many European countries, required abortions to be performed in hospitals. Experience in the 1970s demonstrated that the procedure could be performed safely in nonhospital clinics and physicians' offices, where 95 percent of U.S. abortions take place in the 21st century, including the large majority of second-trimester procedures. Despite the extraordinary safety record of nonhospital facilities, some states require that abortions past a certain point in gestation be performed in hospitals, and some require all providers of more than a minimal number of abortions to be licensed as outpatient surgical facilities or the equivalent. The latter requirement appears to apply to medication abortions as well as surgical.

Similarly, many states allow abortions to be performed only by licensed physicians, although studies have shown that physician assistants and other midlevel practitioners can safely perform early vacuum aspirations [63, 64] and are doing so in several states. In an area with a shortage of physician abortion providers, midlevel practitioners potentially can increase the availability of services.

In the United States, the federal government and several states prohibit a procedure called "partial-birth abortion," a term without accepted medical definition but referring to certain techniques used in the second trimester. Providers have responded to the law by modifying the methods used to perform many of the abortions past about 16 weeks of pregnancy because it is unclear from the law exactly what is prohibited.[65]

Other state restrictions on abortion include the prohibition of coverage in private or public insurance plans except at extra cost, burdensome requirements for disposing of fetal remains, prohibition of abortion services in public facilities, and laws allowing health professionals to refuse to participate in abortion services.

CONCLUSION

The intermingling of politics and religion and contraception has had a long and turbulent history in the United States. Despite the fact that safe and effective

contraception and abortion services are available, many find it hard if not impossible to obtain such services. Perhaps because of the hostile climate, unintended pregnancies, which occur in all societies, are especially prevalent in the United States. The evidence is quite clear. Rates of unintended pregnancy are highest among young women, those who are cohabiting, poor women, minorities, and women with low education. Primary prevention is contraception, which in theory has a high rate of effectiveness. Studies have shown that in practice, women and couples in the United States have difficulty using reversible methods consistently at every exposure and experience gaps in use. For example, although oral contraceptives are extremely effective, 8 percent of women become pregnant during the first year of use. Each year, 11 percent of women at risk of unintended pregnancy experience an accidental pregnancy, about half because they were not using contraception and half because of contraceptive use failure. Greater acceptance of long-acting reversible contraception, which is more widely used in Europe, could potentially reduce the rate of unintended pregnancy and abortion. The copper-T and progesterone IUDs and hormonal implant are available in the United States, but there are significant service delivery and cost barriers to their use.

Abortion to resolve an unintended pregnancy is increasingly accepted worldwide, and its legalization has many public health benefits, in part because it virtually eliminates unsafe illegal abortion. In the United States, however, states have imposed a variety of restrictions since *Roe v. Wade*. Barriers to the use of the most effective contraceptive methods and to abortion services are most harmful to low-income and minority women, who have the highest rates of unintended pregnancy and the fewest resources to overcome the barriers. Much more needs to be done to enable those who wish to control their fertility to do so. The medical and public health consequences of an unsafe abortion, for example, are crystal clear, as are the ramifications of bearing an unwanted pregnancy.

REFERENCES

1. Finer LB, Henshaw SK. Disparities in rates of unintended pregnancy in the United States, 1994 and 2001. *Perspect Sex Reprod Health*. 2006;38(2):90–96.
2. Mosher WD, et al. Centers for Disease Control and Prevention, National Center for Health Statistics. Use of contraception and use of family planning services in the United States: 1982–2002. *Advance Data*. 2004;(350):1–36.
3. Jones RK, Darroch JE, Henshaw SK. Patterns in the socioeconomic characteristics of women obtaining abortions in 2000–2001. *Perspect Sex Reprod Health.* 2002;34(5):226–235.
4. Henshaw SK, Kost K. Abortion patients in 1994–1995: characteristics and contraceptive use. *Fam Plann Persp*. 1996;28:140–147,158.
5. Frost JF, Singh S, Finer LB. US women's one-year contraceptive use patterns: 2004. *Perspect Sex Reprod Health*. 2007;39(1):48–55.
6. Jones RK, et al. Abortion in the United States: incidence and access to services, 2005. *Perspect Sex Reprod Health*. 2008;40(1):6–16.
7. Trussell J. Contraceptive efficacy. In: Hatcher RA, et al., eds. *Contraceptive Technology*. 19th rev. ed. New York: Ardent Media; 2007.
8. Nelson A, Westhoff C, Schnare SM. Real-world patterns of prescription refills for branded hormonal contraceptives: a reflection of contraceptive discontinuation. *Obstet Gynecol*. 2008;112(4):782–787.

9. Sonfield A. Popularity disparity: attitudes about the IUD in Europe and the United States. *Guttmacher Policy Rev.* 2007;10(4):19–24.
10. World Health Organization. *Medical Eligibility Criteria for Contraceptive Use*. Geneva, Switzerland: World Health Organization; 2004.
11. Harper C. Challenges in translating evidence to practice: the provision of intrauterine contraception. *Obstet Gynecol*. 2008;111:1359–1369.
12. Reeves MF, Smith KJ, Creinin MD. Contraceptive effectiveness of immediate compared with delayed insertion of intrauterine devices after abortion. *Obstet Gynecol.* 2007;109(6):1286–1294.
13. Goodman S, et al. Increasing intrauterine contraception use by reducing barriers to post-abortal and interval insertion. *Contraception*. 2008;78:136–142.
14. Trussell J, et al. Cost effectiveness of contraceptives in the United States. *Contraception*. 2009;79(1):5–14.
15. Foster DG, et al. Number or oral contraceptive pill packages dispensed, method continuation, and costs. *Obstet Gynecol*. 2006;108(5):1107–1114.
16. Goodman S, et al. Impact of immediate post-abortal insertion of intrauterine contraception on repeat abortion. *Contraception*. 2008;78:143–148.
17. Shepherd JR. *Marriage and Mandatory Abortion among the 17th Century Siraya.* Arlington, VA: American Anthropological Association; 1995.
18. Boland R, Katzive L. Developments in laws on induced abortion: 1998–2007. *Int Fam Plann Perspect.* 2008;34(3):110–120.
19. Sedgh G, et al. Induced abortion: estimated rates and trends worldwide. *Lancet*. 2007;370:1338–1345.
20. Henshaw S, et al., eds. *Readings on Induced Abortion*; vol 2. *A World Review 2000*. New York: The Alan Guttmacher Institute; 2001.
21. Sedgh G, et al. Legal abortion worldwide: incidence and recent trends. *Int Fam Plann Perspect*. 2007;33(3):106–115.
22. Ventura SJ, et al. Estimated pregnancy rates by outcome for the United States, 1990–2004. *Natl Vital Stat Rep*. 2008;56(15).
23. Gamble SB, Centers for Disease Control and Prevention, et al. Abortion surveillance: United States, 2005. *MMWR Surveill Summ*. 2008:57(SS–13).
24. Finer LB, et al. Reasons US women have abortions: quantitative and qualitative perspectives. *Perspect Sex Reprod Health*. 2005;37:110–118.
25. Cates W, Grimes DA. Deaths from second trimester abortion by dilatation and evacuation: causes, prevention, facilities. *Obstet Gynecol*. 1981;58:401–408.
26. Grimes DA, Cates W, Selik RM. Abortion facilities and the risk of death. *Fam Plann Perspect*. 1981;13:30–32.
27. Henshaw SK, Finer LB. The accessibility of abortion services in the United States, 2001. *Perspect Sex Reprod Health*. 2003;35:16–24.
28. Tietze C. Two years' experience with a liberal abortion law: its impact on fertility trends in New York City. *Fam Plann Perspect*. 1973;5(1):36–41.
29. Abernathy JR, Greenberg BG, Horvitz DG. Estimates of induced abortion in urban North Carolina. *Demography*. 1970;7(1):19–29.
30. Levine PB, et al. *Roe v Wade* and American fertility. *Am J Public Health.* 1999;89(2): 199–203.
31. Hilgers TW, O'Hare D. Abortion related maternal mortality: an in-depth analysis. In: Hilgers TW, Horan D, Mall D, eds. *New Perspectives on Human Abortion*. Frederick, MD: University Publications of America, Inc.; 1981:69–91.
32. Grimes DA. Estimation of pregnancy-related mortality risk by pregnancy outcome: United States, 1991 to 1999. *Am J Obstet Gynecol*. 2006;194:92–94.

33. Tietze C. The public health effects of legal abortion in the United States. *Fam Plann Perspect.* 1984;16(1):26–28.
34. Institute of Medicine, National Academy of Science. *Legalized Abortion and the Public Health.* Washington, DC: National Academy of Sciences; 1975.
35. Grossman M, Jacobowitz S. Variations in infant mortality rates among counties of the United States: the roles of public policies and programs. *Demography*. 1981;18(4):695–713.
36. Corman H, Grossman M. Determinants of neonatal mortality rates in the US: a reduced form model. *J Health Econ.* 1985;4(3):213–236.
37. Joyce T. The impact of induced abortion on black and white birth outcomes in the United States. *Demography*. 1987;24(2):229–244.
38. Egan JF, et al. Down syndrome births in the United States from 1989 to 2001. *Am J Obstet Gynecol*. 2004;191(3):1044–1048.
39. David HP, Dytrych Z, Matejcek Z. Born unwanted: observations from the Prague study. *Am Psychol.* 2003;58(3):224–229.
40. Kubicka L, et al. Children from unwanted pregnancies in Prague, Czech Republic revisited at age thirty. *Acta Psychiatr Scand*. 1995;91:361–369.
41. Forssman H, Thuwe I. One hundred twenty children born after application for therapeutic abortion refused. *Acta Psychiatr Scand.* 1966;42:70–87.
42. Lanham JT, Kohl SG, Bedell JH. Changes in pregnancy outcome after liberalization of the New York State abortion law. *Am J Obstet Gynecol*. 1974;118(4):485–492.
43. Lester D. Letter to the editor: *Roe v Wade* was followed by a decrease in neonatal homicide. *JAMA*. 1992;267(22):3027–3028.
44. Atrash HK, Hogue CJR. The effect of pregnancy termination on future reproduction. *Baillières Clin Obstet Gynecol*. 1990;4(2):391–405.
45. Hogue CJR, Cates W, Tietze C. The effects of induced abortion on subsequent reproduction. *Epidemiol Rev*. 1982;4(1)66–94.
46. Adler N, et al. Psychological responses after abortion. *Science Magazine*. April 6, 1990:41–44.
47. Major B, et al. Psychological responses of women after first-trimester abortion. *Arch Gen Psychiatry*. 2000;57(8):777–784.
48. Gilchrist AC, et al. Termination of pregnancy and psychiatric morbidity. *Br J Psychiatry*. 1995;167(2):243–248.
49. American Psychological Association. Report of the APA Task Force on mental health and abortion. http://www.apa.org/releases/abortion-report.pdf. Published August 13, 2008. Accessed September 18, 2009.
50. The Guttmacher Institute. *State Policies in Brief: State Funding of Abortion under Medicaid.* New York: The Guttmacher Institute; 2009. http://www.guttmacher.org/statecenter/spibs/spib_SFAM.pdf.
51. IPPF European Network. *Abortion Legislation in Europe*. Brussels, Belgium: IPPF European Network; 2009. http://www.ippfen.org/NR/rdonlyres/8382308F-B663-471A-A8B99B214695F5BA/0/Pub_AbortionLegislationEuropeupdate_Dec2008.pdf.
52. Henshaw SK, et al. *Restrictions on Medicaid Funding for Abortions: A Literature Review*. New York: The Guttmacher Institute; 2009. http://www.guttmacher.org/pubs/MedicaidLitReview.pdf.
53. Cook PJ, et al. The effects of short-term variation in abortion funding on pregnancy outcomes. *J Health Econ.* 1999;18:241–257.
54. Trussell J, et al. The impact of restricting Medicaid financing for abortion. *Fam Plann Perspect*. 1980;12(3):120–130.
55. Chrissman M, et al. Effects of restricting federal funds for abortion: Texas. *MMWR.* 1980;29(22):253–254.

56. The Guttmacher Institute. *State Policies in Brief: Parental Involvement in Minors' Abortions.* New York: The Guttmacher Institute; 2009. http://www.guttmacher.org/statecenter/spibs/spib_PIMA.pdf.
57. Blum W, Resnick MD, Stark TA. The impact of a parental notification law on adolescent abortion decision-making. *Am J Public Health.* 1987;77(5):619–620.
58. Henshaw SK, Kost K. Parental involvement in minors' abortion decisions. *Fam Plann Persp*. 1992;24:196–207.
59. Dennis A, et al. The impact of laws requiring parental involvement for abortion: a literature review. New York: The Guttmacher Institute; 2009.
60. The Guttmacher Institute. *State Policies in Brief: Mandatory Counseling and Waiting Periods for Abortion.* New York: The Guttmacher Institute; 2009. http://www.guttmacher.org/statecenter/spibs/spib_MWPA.pdf.
61. Joyce T, Henshaw SK, Skatrud JD. The impact of Mississippi's mandatory delay law on abortions and births. *JAMA*. 1997;278(8):653–658.
62. Joyce TJ, et al. *The Impact of State Mandatory Counseling and Waiting Period Laws on Abortion: A Literature Review.* New York: The Guttmacher Institute; 2009.
63. Goldman MB, et al. Physician assistants as providers of surgically induced abortion services. *Am J Public Health*. 2004;94(8):1352–1357.
64. Freedman MA, et al. Comparison of complication rates in first trimester abortions performed by physician assistants and physicians. *Am J Public Health*. 1986; 76(5):550–554.
65. Haddad L, et al. Changes in abortion provider practices in response to the Partial-Birth Abortion Ban Act of 2003. *Contraception*. 2009;79(5):379–384.

CHAPTER 21

Disease Prevention through Vaccination: The Science and the Controversy

Tony Rosen, MPH, MD

> I shall never have smallpox for I have had cowpox. I shall never have an ugly pockmarked face.[1]

INTRODUCTION

Each one of us is constantly warding off the potential for infection or disease. After all, the world is filled with countless microbes, fortunately most of which are harmless and some even beneficial. But there are plenty of microbes that can hurt us (pathogens, from the Greek word for disease, "pathos"). Harmful bacteria, for example, may cause disease through infection of their host or by the release of powerful toxins. Viruses, inert by themselves, have the ability to invade the cells of other life forms. As those cells duplicate, so does the virus. While one's ability to ward off disease is usually strong, there are instances where the microbes overwhelm the body and produce illness. In an effort to protect against many infectious diseases, we rely on vaccines with the intent of conferring protection and immunity.

Vaccines have been a vehicle for disease prevention and eradication for hundreds of years. The development and widespread distribution of safe, effective, and affordable vaccines has done more for disease prevention over time than nearly any other medical or public health intervention. Of the 10 greatest public health achievements over time, certainly immunization against disease ranks at or near the top of the list.[2] Millions of lives have been saved because of the widespread use of vaccines to prevent or eradicate diseases such as measles, diphtheria, pertussis, tetanus, polio, and, of course, smallpox. The eradication of smallpox worldwide is undoubtedly one of the most spectacular public health initiatives of all time.

Vaccination not only protects an individual from disease, but also has the dual role of protecting the community at large from disease outbreaks. For disease to

---

This chapter was originally published as Tony Rosen "Disease Prevention Through Vaccination: The Science and the Controversy," in Finkel, M.L. (ed). *Truth, Lies, and Public Health: How We Are Affected When Science and Politics Collide*. Westport, CT: Praeger. ©2007.

spread, there must be a pool of susceptible people in whom the bacteria or virus can grow. Ironically, those who elect not to vaccinate themselves or their children are actually benefiting from those who are vaccinated. This concept is referred to as "herd immunity." When a disease spreads from one human to another, it requires both an infected person to spread it and a susceptible person to catch it. Herd immunity works by decreasing the number of susceptible individuals, even those who refuse to be vaccinated, and when this number drops low enough, the disease will disappear from the community because there are not enough people to continue the catch-and-infect cycle. The greater the proportion of vaccinated individuals, the more rapidly the disease will disappear. Once-common diseases such as pertussis, polio, smallpox, and measles have all but disappeared thanks to the large numbers of individuals who are vaccinated against these diseases. Periodically, however, there have been mini-outbreaks of disease for which there are vaccinations. For example, there are outbreaks of measles (a particularly contagious, potentially serious disease) in the United States as well as around the world, primarily as a result of a pool of unvaccinated children. Those who are not vaccinated are at high risk of contracting this disease.[3]

Because microbes know no foreign boundary, diseases in one part of the world can quickly and easily spread to other parts of the globe. As such, a unified global vaccination policy is needed; how to achieve such a noble and important goal, however, often is not easily accomplished. Economics, politics, and social constraints can and do play important roles in disease-eradication programs. The success of immunization policies depends on, and is linked with, interrelated factors, including vaccine safety (quality control and monitoring), adequate vaccine supply (to avoid vaccine shortages), effective delivery systems to ensure that the vaccines get to those in need (more of an issue in the developing world), financial incentives and legal protection for the vaccine manufacturers, and educational efforts to inform the public about the benefits and risks of vaccinations. Indeed, perhaps most of all, there is a need to focus on the public's fears about the safety of vaccination and their willingness to be immunized.

In addition to the scientific challenges to vaccine development, social, ethical, economic, legal, and political issues individually and collectively have served to curtail and in some cases to derail efforts to immunize populations. Vociferous antivaccination movements frequently clashed with the government's authority to immunize for the "common good." Historically, antivaccinationists have protested against what they consider the intrusion of their privacy and bodily integrity. One of the potent symbols of the early antivaccine movement was the limp "Raggedy Ann" doll, which was created in 1915 by a man whose daughter died shortly after being vaccinated in school without parental consent. The medical authorities blamed a heart defect, but the parents blamed their child's death on the shot. Since that time, there have been reports of deep-seated public fears of vaccinations, as well as protests against compulsory vaccination laws.

The issue of vaccine safety periodically makes front-page news, usually after an unfortunate event in which someone or many individuals were harmed in some way allegedly as a result of being vaccinated. Proponents of vaccination would be the last to say that vaccination is risk free, but they would be the first to argue that the small risks outweigh the dangers of not being vaccinated. To lose ground to the tremendous achievements realized by vaccines because of the public's mistrust could be potentially serious. Are the antivaccinationists off base, or are their concerns valid? How should

the public health and medical communities respond? What role should government have to legally enforce vaccination policy? This chapter focuses on the history of vaccines and immunization and the new challenges that must be addressed to ensure against a resurgence in vaccine-preventable diseases locally, nationally, and globally.

WHAT ARE VACCINES AND HOW DO THEY WORK?

The doctrine holding that infectious diseases are caused by the activity of microorganisms within the body is referred to as the germ theory of disease, also called the pathogenic theory of medicine, which states that microorganisms are the cause of many diseases. Although highly controversial when first proposed in the 19th century, it is now a cornerstone of modern and clinical microbiology. Put simply, disease-causing organisms, be they viruses, microbes, or bacteria, attack the body and produce illness. The immune system, if working correctly, prevents illness by destroying disease-causing microorganisms that threaten the body.

Vaccines, from the Latin word "vacca," or cow, trigger one's immune system's infection-fighting ability and memory without exposure to the actual disease-producing germs. Instead, the person is injected with a dead or much weakened (and not dangerous) version of the pathogen. Vaccines stimulate the body's immune system by triggering an immune response; the immune system goes into high gear to destroy the invader. The immunity one develops following vaccination is similar to the immunity acquired from natural infection. For some diseases, several doses of a vaccine (a booster) may be needed for a full immune response. For others, one shot is sufficient.

One's body can become immune to bacteria or viruses by either developing a natural immunity to the disease or by vaccine-induced immunity. *Natural immunity* develops after one has been exposed to an organism, and one's immune system develops a defense (from antibodies and memory cells) to prevent one from getting sick again from that particular type of virus or bacterium. *Vaccine-induced immunity* results after one receives a vaccine, which makes the body think that it is being invaded by a specific organism and the immune system reacts by destroying, the "invader" and preventing it from infecting the person again. The immunity one develops following vaccination is similar to the immunity acquired from natural infection. The goal is the same: to stimulate an immune response without causing disease.

Briefly, the human immune system works because antigens (proteins from the foreign microorganism) stimulate an immune response leading to the synthesis of antibodies (proteins that attack and destroy viral or bacterial particles). "Memory cells" are produced in an immune response, and these cells remain in the bloodstream ready to mount a quick protective immune response against subsequent infections with the particular disease-causing agent.[4] If the infection was to occur again, the memory cells would respond to inactivate the disease-causing agents, and the individual would not likely become sick.

Vaccines traditionally have been classified into three broad categories: live attenuated, whole-killed, and subunit vaccines.

- Live weakened vaccines use live viruses that have been weakened (attenuated). The result is a strong antibody response that establishes lifelong immunity, but live, attenuated vaccines carry the greatest risk because they can mutate to the virulent form at

any time, Because the pathogen is alive, it has the potential to multiply within the human body. Examples include vaccines for measles, mumps, and rubella and for chickenpox.

- Inactivated vaccines use killed or inactivated bacteria or viruses. Examples included the typhoid vaccine and the Salk poliomyelitis vaccine. Toxoid vaccines use bacterial toxins that have been rendered harmless to provide immunity to the specific toxin. Examples included diphtheria and tetanus vaccines.
- Acellular and subunit vaccines are made by using only part of the virus or bacteria. Advances in biotechnology and genetic engineering techniques have made it possible to produce subunit vaccines in which genes that code for appropriate subunits from the genome of infectious agent are isolated and placed into bacteria or yeast host cells, which then produce large quantities of subunit molecules by transcribing and translating the inserted foreign DNA (Deoxyribonucleic acid). Subunit vaccines cannot cause the disease. Examples include hepatitis B and *Haemophilus influenza* type B vaccines. A booster every few years is often required to continue effectiveness.

IT ALL STARTED WITH COWPOX

No discussion of vaccines can be considered complete without a discussion of smallpox and Edward Jenner, a country doctor in England who is credited with performing the world's first vaccination in 1796.[5] The eradication of smallpox is probably the world's greatest success story. For thousands of years, epidemics swept across continents, decimating populations and at times changing the course or history. The Crusaders brought smallpox back with them from the Holy Land. The Conquistadors carried it to the New World. This disease destroyed the Incan and Aztec empires. In the American colonies, smallpox helped decimate the indigenous peoples, including Pocahontas who died of smallpox in 1617 after visiting London. Rich and poor, famous and unknown, smallpox did not discriminate. Queen Mary II of England, Emperor Joseph I of Austria, King Luis I of Spain, Tsar Peter II of Russia, and King Louis XV of France are a few of the heads of state who died from smallpox. The disease, for which no effective treatment was ever developed, killed as many as 30 percent of those infected. Between 65 percent and 80 percent of survivors were marked with deep-pitted scars (pockmarks), most prominent on the face. George Washington, for example, survived a bout with smallpox but was severely scarred.

Long before the causes of this disease were known and understood, many tried to protect the population from this disfiguring and deadly disease. The Chinese may have begun intentionally infecting themselves with smallpox virus as early as the 10th century, trying to prevent the disease by exposing uninfected individuals to the pus and fluid from a smallpox lesion. The thinking was that the dried pus would confer protection to the individual. This practice, called variolation, was also used hundreds of years later in other parts of the world. Specifically, in the early 18th century, Lady Mary Wortley Montagu, the wife of the British ambassador to Constantinople, who as a young girl contracted smallpox and whose brother died of the disease, popularized variolation upon her return to England. Because of Lady Montagu's efforts, the Princess of Wales in 1722 was persuaded to have her two children inoculated against smallpox. Although the physiological effects of variolation varied, ranging from a mild illness to death, its effectiveness was evident. Smallpox mortality and morbidity rates were lower in populations that used variolation than in those that did not.[6]

Across the Atlantic Ocean, smallpox was threatening Boston. Clergyman Cotton Mather and Dr. Zabdiel Boylston in Massachusetts practiced variolation in an attempt to inoculate residents of this city. Although inoculations were illegal in the American colonies, their efforts helped prevent a wide-scale smallpox epidemic. They documented that the smallpox case fatality rate was much lower among those inoculated than those not inoculated.[6]

Although Jenner was not the first to experiment with inoculation against smallpox, his efforts, which most certainly would be considered to be unethical by 21st-century standards, are acknowledged to mark the beginning of widespread vaccination. Jenner observed that milkmaids who had cowpox (a mild disease) rarely developed smallpox (a serious and potentially fatal disease). This observation prompted him to experiment and ultimately devise the first vaccine to protect individuals from this dreaded disease. Jenner's experiment on eight-year-old James Phipps spared the boy from developing smallpox, but still Jenner's peers did not readily accept his findings. Rebuffed by the Royal Society of London, Jenner was undeterred and completed more experiments and self-published his findings in 1798. His results were so compelling that thousands of people elected to protect themselves by infecting themselves with cowpox.[7] Though it took several years until Jenner's theories about vaccination were accepted by the professional societies, by 1800, more than 100,000 people had been vaccinated against smallpox worldwide. Vaccination was made compulsory in Bavaria, Denmark, Sweden, and, by the mid-19th century, in Great Britain. Massachusetts was the first U.S. state to make vaccination compulsory in 1809.[8]

By the mid-20th century, 150 years after the introduction of vaccination, an estimated 50 million cases of smallpox occurred in the world each year, a figure that dropped to around 10 to 15 million per year by 1967 because of successful vaccination efforts. In 1967, when the World Health Organization (WHO) launched an intensified plan to eradicate smallpox from the earth, the "ancient scourge" threatened 60 percent of the world's population, killed every fourth victim, scarred or blinded most survivors, and eluded any form of treatment.[9] A massive, worldwide outbreak search and vaccination program was initiated. Through the success of this global eradication campaign, smallpox was finally limited to the Horn of Africa and then to a single last natural case, which occurred in Somalia in 1977, although one fatal laboratory-acquired case occurred in the United Kingdom in 1978. The global eradication of smallpox was certified by a commission of eminent scientists in December 1979, based on intense verification activities in countries, and subsequently was endorsed by the WHO in 1980. Three known repositories of the virus were left: one in Birmingham, England, which was later destroyed after an accidental escape from containment caused many deaths, and two still remaining for possible antibioweaponry (stored under extremely strict conditions at the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, and at the State Research Center of Virology and Biotechnology in Koltsovo, Russia).

MILESTONES IN VACCINE HISTORY

It was almost 100 years after Jenner's seminal work that vaccination moved beyond smallpox. The French chemist, Louis Pasteur, developed what he called a

Table 21.1
Highlights in Vaccine Development

1905	U.S. Supreme Court upholds state law mandating smallpox vaccination.
1944	Pertussis vaccine recommended for universal use in infants.
1947	DPT (trivalent diphtheria/pertussis/tetanus) recommended for routine use.
1955	Salk inactivated polio vaccine licensed.
1961	Sabin oral, live-virus polio vaccine licensed.
1963	Measles vaccine licensed.
1971	MMR (trivalent measles/mumps/rubella) licensed.
1972	United States ended routine use of smallpox vaccine.
1977	Smallpox eradicated worldwide.
1986	Vaccine Injury Compensation Act passed. Recombinant Hepatitis B vaccine licensed (recommended for all newborns and children in 1991).
1988	Vaccine Injury Compensation Program funded.
1999/2000	Joint statement by the U.S. Public Health Service, the American Association of Family Practitioners, the American Association of Pediatrics urged manufactures to remove the preservative thimerosal as soon as possible from vaccines routinely recommended for infants.

Source: National Vaccine information. www.900shot.com/timeline.html.

rabies vaccine in 1885, but technically what he produced was a rabies antitoxin that functioned as a postinfection antidote.[10] By the 20th century, advances in the science of virology bacteriology and immunology led to a better understanding of how the human body defends itself against invading microorganisms. Development of viral vaccines and bacteria-based vaccines flourished: the development of vaccines against more than 20 diseases has affected disease morbidity and mortality. Since 1980, more than 15 new or improved vaccines have been approved as a result of advances in molecular biology and genetics, which led to new and improved subunit vaccines that promise to offer increased safety and efficacy. (See table 21.1 for a listing of some of the highlights in history of vaccine development and table 21.2 for a listing of currently recommended childhood vaccinations.)

PERTUSSIS AND DIPHTHERIA

Although the causative agent of pertussis (whooping cough) was isolated in 1907, it was not until the late 1920s that the first whole-killed pertussis vaccine was introduced.[11] Pertussis, particularly serious among infants, is a contagious respiratory disease caused by the *B. pertussis* bacterium and spread by coughing or sneezing. Toxins produced by *B. pertussis* can cause high fever, convulsions, brain damage, and death.

Diphtheria, also caused by a bacterium, is a contagious and potentially life-threatening infection that usually attacks the throat and nose. In more serious cases, it can attack the nerves and heart. Although he survived smallpox, George Washington may have died of diphtheria. In the mid-1930s, a vaccine against pertussis and diphtheria was developed and was later modified in 1947 to include tetanus (the diphtheria, pertussis, and tetanus [DPT] vaccine). In the 21st century, the DPT shot is among the first that an infant receives after birth. A child needs five DPT shots,

Table 21.2
Recommended Vaccinations

By age 6	Measles, mumps, rubella, polio, chicken pox, DPT (diphtheria, tetanus, pertussis), Hib (meningitis), PVC (pneumonia), rotavirus (diarrhea), hepatitis A and B, flu (annually).
By age 18	Meningococcus, cervical cancer (girls only; an HPV vaccine for boys is being developed), flu (annually).
Ages 18+	Flu (annually), tetanus and diphtheria (every 10 years), measles, mumps, rubella, chicken pox (for those not previously infected), pneumococcal pneumonia (boost after age 65).

Source: Centers for Disease Control and Prevention 2009.
Note: The CDC has updated its recommended list of vaccines several times over the past 15 years. Each state rather than the CDC decides which vaccines to make compulsory for entry into school.

given at specified intervals, to ensure complete protection. Unfortunately, after the DPT booster vaccine became widely used, serious adverse reactions including convulsions, brain damage, and even death were noted in a tiny percentage of children who were vaccinated. In particular, the pertussis component of the DPT vaccine was identified as causing problems in some children. As a result, children with a history of convulsions or neurological disease were strongly advised not to be vaccinated.[12–14] Although serious acute neurologic illness was a rare event, the Institute of Medicine (IOM) was mandated by Congress to study the issue. In 1991, the IOM issued its report and concluded that the evidence was insufficient to indicate a causal relation between DPT and neurologic damage.[15] The National Childhood Encephalopathy Study (NCES) also found that children who experienced rare but serious acute neurologic disorders within seven days of receiving DPT were no more or less likely to experience documented chronic nervous system dysfunction or to have died within 10 years of the acute disorder than children who had not received DPT within seven days before the onset of the disorder.[16] In sum, no special characteristics were associated with acute or chronic nervous system illnesses linked to DPT exposure. But, the public's trust was shaken.

POLIO

Probably no disease created as much fear as polio. Paralysis and death were the major hazards of this disease. Probably the most famous polio victim in the United States, perhaps even in the world, was President Franklin D. Roosevelt, who hid the extent of his disability from the public throughout his presidency. Polio was one of the most dreaded childhood diseases of the 20th century. The first clinical description of polio dates to 1789 when a British physician provided the first description of the disease (debility of the lower extremities). The first known large epidemic occurred in 1916, killing 6,000 people and leaving 27,000 more paralyzed.[17] In retrospect, isolation and quarantine were not effective means of controlling the disease. A race to develop an effective polio vaccine began in the 1930s, and unfortunately early clinical trials failed in that many individuals ended up infected with polio, which clearly was not the intent of the vaccine developers.

Widespread epidemics of polio were documented after World War II, with an average of more than 20,000 cases a year occurring between 1945 and 1949. In 1952, there were 58,101 cases of polio in the United States, the most ever counted. By the mid-20th century, "polio hysteria" fueled fear across the country.

The difficulty in developing a polio vaccine stemmed from the fact that this disease is caused by three strains of virus. Understanding the polioviruses took decades, with much of the research funded by the March of Dimes Foundation, a grassroots organization founded with the help of President Roosevelt. In the late 1940s, Dr. Jonas Salk began to use the newly developed tissue cultures method of cultivating and working with the poliovirus. The first safe and effective vaccine, the Salk-injected vaccine, used killed poliovirus. During the 1950s, massive trials of the Salk vaccine, unprecedented in medical history, were conducted and led to a nationwide mass immunization campaign promoted by the March of Dimes. This effort led to a significant drop in the number of new cases of polio in the United States, and in 1955, the inactivated polio vaccine was licensed for use in the United States. While the vaccine helped stop polio in its tracks, there were problems with the vaccine related to the incomplete inactivation of some virus particles. This was soon corrected.

During this time, Dr. Albert Sabin, a bitter rival of Salk, also was working on a polio vaccine. His vaccine used live, attenuated (weakened) virus rather than killed poliovirus. Whereas the Salk vaccine required injections, the Sabin vaccine was oral. Field trials of this vaccine proved the Sabin oral vaccine to be effective; the oral, live-virus polio vaccine was licensed in 1961. Because live vaccine contains a weakened type of poliovirus that could in theory mutate into more virulent forms (albeit exceedingly rare, but not unheard of), it is not given to people with impaired immune systems. The oral vaccine was superior in terms of ease of administration, and it also provided long-lasting immunity. Both vaccines have advantages and disadvantages with regard to safety and cost, and both are used throughout the world.

The discovery and use of the polio vaccines nearly eliminated polio in the United States, and in 1994, this disease was declared eradicated in all of the Americas. While both the Salk and the Sabin vaccines proved to be highly effective in preventing the disease, those who had been paralyzed by polio, estimated to be in the hundred of thousands, unfortunately did not benefit from these milestones in polio vaccine development.

MEASLES, MUMPS, AND RUBELLA

As late as the 1950s, and before a vaccine was developed, parents were encouraged to expose their children to diseases like measles, mumps, and chicken pox to develop immunity. With the marketing of an effective vaccine to protect people from these diseases, such thinking was rendered moot. Building on the momentum of success achieved with the oral polio vaccine, a number of live attenuated vaccines were being developed. The most significant of these at the time was the measles-mumps-rubella (MMR) vaccine.

For hundreds of years, measles was so ubiquitous it was thought to be a natural episode of childhood. It was not until the 14th century that the word "measles'' was used, stemming from the word "miser," which was used to refer to the wretchedness

of lepers.[18] Before the development of an effective vaccine, measles was one of the most common childhood diseases in America. Characterized by fever and a rash, measles is a serious disease that is highly contagious and can lead to death; but recovery confers a lifelong immunity. Interestingly, women who have been vaccinated but who never had the disease do not have natural maternal measles antibodies to pass on to their babies, which mean that most babies born in America are vulnerable to getting this disease.

Mumps, a viral disease, used to be common in childhood. Discovery of the mumps virus in 1934 helped researchers gain a better understanding of the symptoms and how this disease is transmitted. Characterized by fever, headache, and inflammation of the salivary glands (making the cheeks swell, producing the signature sign of the disease), this disease rarely leads to death. Recovery confers lifelong immunity.

Rubella (German measles) is considered a mild childhood disease characterized by a pink rash. While similar to measles, the rubella virus is comparatively benign and less infectious. Recovery usually confers lifelong immunity although repeat cases can occur, albeit rarely. Should a pregnant woman get rubella in the first trimester of pregnancy, there is a greater chance of giving birth to a baby with birth defects.

A measles vaccine was licensed in 1963, and a rubella vaccine was licensed in 1969, and the trivalent MMR vaccine was licensed in 1971. Protection is estimated to last for up to 11 years. Despite the availability of the vaccine, however, around 1 million children, predominantly in resource-poor countries, die every year from measles. Even in the United States, outbreaks occur. For example, a measles outbreak in the 1980s and early 1990s showed that there were a significant number of vaccine failures in older children, teenagers, and adults, especially among those who had been vaccinated before 15 months of age. As such, the government recommended that a second MMR booster be given either before a child enters kindergarten or before entering junior high school. Almost all who get the vaccine have no serious adverse reactions from it.

As is often the case, the risks of the vaccine are usually smaller than the risks from the diseases. However, in the mid-1990s, reports of an association between autism and the MMR vaccine were published. There was speculation that the MMR vaccination could cause autism in some children. This finding alarmed both the lay public and the scientific community. Parents refused to have their children immunized, and the IOM was asked to investigate this link (see "Protesting Vaccines: Fact or Myth").

INFLUENZA

Historically, influenza (the flu) epidemics have cased havoc. Charlemagne's army may have been decimated by the flu during an epidemic in 876. The great influenza pandemic of 1918–1919, the 20th century's worst epidemic, killed millions of people. During World War I, the number of American killed by influenza (44,000) almost was equal to the number killed in battle (50,000). As the nation entered World War II, the military made influenza vaccination mandatory.

Influenza is a contagious disease spread by person-to-person contact and caused by the influenza virus. Peak flu season occurs usually from late December through March. There are three basic flu germs, variants of which are popularly designated

according to where they first strike—that is, Hong Kong B, Bangkok A, and so forth. It is important to remember that influenza viruses are constantly changing, so an antibody made against one strain will become less effective against new strains as influenza strains evolve over time. In addition, different types of influenza viruses circulating and different variants within virus types, and the same type of flu virus does not necessarily circulate each year. For instance, during the 2005–2006 flu season, influenza A (H3N2) viruses predominated; however, infection with influenza A (H3N2) virus would not provide protection against influenza B or influenza A (H1N1) viruses. The viruses that cause flu are prone to mutation, making the manufacture of vaccines an annual guessing game of sorts. If a new mutation pops up anywhere in the world, resulting from a major change (antigenic "shift"), it will quickly spread, leaving most people unprotected.

A flu shot can help prevent one from getting sick, but even with the flu vaccine available, each year millions of people get sick, and some tens of thousands die from the flu. When complicated by pneumonia, it is one of the 10 most common causes of death in the United States. Most people who get a flu shot have no serious problem from the vaccination. Those over age 60 should get a flu shot every year.

HEPATITIS

Hippocrates was the first to note epidemics of jaundice, a telltale characteristic of hepatitis. Hepatitis is a gastroenterological disease featuring inflammation of the liver. Most cases of acute hepatitis are due to viral infections. There are many types of hepatitis, and the disease can be contracted in a few different ways. *Hepatitis A* is transmitted by the orofecal route and is contracted through contaminated food or water. This form of hepatitis does not lead to chronic or lifelong disease, and just about everyone who gets hepatitis A has a full recovery. *Hepatitis B* can be contracted from blood, semen, and saliva (making it one of the venereal diseases) and also from tattoos. Hepatitis B can be a serious infection that can cause liver damage; some individuals are not able to get rid of the virus, which makes the infection chronic. Before routine testing of the blood supply, thousands of deaths occurred each year from post-transfusion hepatitis B. Fortunately, this is no longer a risk. *Hepatitis C* is spread the same way as hepatitis B through an infected person's blood and other body fluids as well as from injection drug use. *Hepatitis C* is a chronic infection and often causes liver damage. *Hepatitis D* can only thrive in cells also infected with hepatitis B and is not common. It can be spread through infected blood, dirty needles, and from unprotected sex with a person infected with hepatitis virus. *Hepatitis E* can be contracted from host to host via fecal-oral contact and contamination of water. This type of hepatitis does not occur often in the United States and does not cause long-term damage to the liver. In 1991, a recombinant hepatitis B vaccine was recommended for all newborn infants and children. There is no vaccine for hepatitis C, D, or E.

PNEUMOCOCCAL PNEUMONIA

The *Streptococcus pneumoniae* bacterium (often called pneumococcus) can cause a variety of diseases, from severe, potentially life-threatening pneumonia, meningitis,

and bacteremia to common but less severe otitis media (an ear infection), sinusitis, and bronchitis. These "pneumococcal" diseases, which preferentially attack young children, older adults, and people with immunosuppressive conditions, such as HIV, present significant public health problems.

WHO estimated in 2005 that 1.6 million people die of pneumococcal disease annually, with 63 percent of these deaths occurring in children under age five, and the vast majority occurring in developing countries.[19] In the United States, pneumonia and influenza combined represent the leading infectious cause of mortality in children one to four years old, and the eighth leading cause of death overall.[20] Otitis media is the most common cause of sick visits in preschool-age children in the United States, with 24.5 million clinic visits and $2 to 5.3 billion in annual treatment costs. By three years old, 83 percent of U.S. children will have experienced at least one episode of otitis media.[21] Otitis media is the most common reason for antibiotic prescriptions among U.S. children and thus contributes significantly to growing antibiotic resistance.

Designing an effective vaccine to immunize patients against pneumococcus is challenging, because more than 90 different "serotypes" of the bacterium have been identified. Each of these microorganisms is unique, with slightly different proteins and polysaccharides (complex sugars) on their surface.[22] When the human body produces antibodies to a single bacterium, these typically provide protection against a single serotype. (This is the reason that the same patient may contract multiple ear infections.) Therefore, the recent strategy in pneumococcal vaccine design has been to include in the formulation polysaccharides from multiple serotypes, with a focus on those that most commonly cause severe, invasive disease, to build up broader protection through vaccination.

In 2000, the first of these pneumococcal conjugate vaccines, PVC-7, was released. This vaccine included seven serotypes covering 65 to 80 percent of bacteria associated with invasive pneumococcal disease among young children in Western Hemisphere industrial countries.[23] In the United States, this vaccine was almost immediately recommended for all children under two years old,[24] and it has already enjoyed dramatic success. In the eight years since the introduction of the vaccine, there has been a 69 percent to 91 percent decrease in the rate of invasive pneumococcal disease.[25, 26] Although the decrease has been most dramatic in young children, the vaccine's target group, significant declines have been seen in older children, adults, and the elderly,[27–30] suggesting that the vaccine induces herd immunity. Also, dramatic reductions in otitis media and severe cases of otitis requiring tympanostomy tubes have been reported.[31–34]

Though the pneumococcal vaccine has achieved wide acceptance, it has not yet been introduced into developing countries where the need is greatest. As of August 2008, 26 countries offered PVC-7 to all children as part of national immunization programs or had the vaccine in widespread use.[35] Unfortunately, none of these were low-income countries, which account for more than 97 percent of pneumonia cases in children under five years old.[35] The countries that have implemented the pneumococcal vaccine are not those with particularly high prevalence of childhood HIV infection, where using it should be a high priority, because persons infected with HIV are up to 300 times more likely to have pneumococcal disease than those who are HIV negative.[36]

The international community has acknowledged challenges to implementation of pneumococcal vaccine in these low-income countries, which include the high cost of the vaccine and logistical issues with safe delivery, and potential solutions are being developed. Both hepatitis B [37] and *Hemophilus influenzae* type B [38] vaccines took nearly 20 years to be introduced in many developing countries, and international action has been aggressive to ensure that the pneumococcal conjugate vaccine is adopted more quickly.

The GAVI Alliance (formerly the Global Alliance for Vaccines and Immunizations), an organization that aligns public and private resources to ensure global vaccine access, has made funding available through 2015 for PVC-7 in the 72 countries with the lowest per capita income (less than $1,000 per year).[35] In addition, countries and private donors have offered an advanced market commitment of $1.5 billion for 7 to 10 years to vaccine manufacturers, guaranteeing them a viable market for next-generation pneumococcal vaccines and simultaneously ensuring that GAVI-eligible countries can purchase these vaccines for a low price (currently less than $0.30 per dose).[35] Despite these actions, none of the GAVI-eligible countries had implemented the vaccine as of August 2008. More encouragingly, 11 have applied for GAVI funding and 8 of these applications (Central African Republic, Democratic Republic of Congo, Gambia, Guyana, Honduras, Kenya, Nicaragua, and Rwanda) have been approved, suggesting that many of these countries may introduce universal childhood pneumococcal vaccination soon.

PCV-7 also poses logistical challenges for low-income countries that decide to implement vaccination programs. Most vaccines are supplied in multidose vials that minimize volume and reduce medical waste but that must be transported and stored cold. Unfortunately, PCV-7 is currently available only in single-dose, prefilled glass syringes, which leads to increased transport and storage requirements and larger waste disposal. Pneumococcal vaccine only comes in prefilled, single-dose syringes that are not automatically disabled, raising safety concerns associated with the potential reuse of syringes and needles.[39] Satisfactory solutions to these problems must be developed for pneumococcal vaccine programs to be safe and effective.

In addition to the challenges of expanding vaccination to the developing world, limitations of the current vaccine strategy are already becoming apparent. Widespread pneumococcal vaccination with the PCV-7 vaccine in the United States has led to the increase in disease from serotypes not covered by the vaccine.[40] Particularly worrisome is the emergence of serotype 19A, which causes a virulent otitis media and is resistant to all Food and Drug Administration (FDA)-approved antibiotics for childhood ear infections.[41] To temporarily address this problem, expanded 11- and 13-serotype vaccines, which include the predominant "replacement" serotypes, are in trials and are expected to become available for use. Constant surveillance will be required, however, as every new vaccine will further alter the prominent serotypes in the environment. Ultimately, a pnuemococcal vaccine that is not serotype-specific would be the ideal solution, and several of these "common-antigen" vaccines are currently in development.

Overcoming the challenges to global introduction of the pneumococcal vaccine and working to develop even more effective next-generation vaccines remain urgent public health priorities for the WHO and UNICEF (United Nations Children's Fund).[19] Successful implementation is imperative, as the global use of

pneumococcal conjugate vaccine will prevent an estimated 5.4 to 7.7 million childhood deaths by 2030.[35]

HUMAN PAPILLOMAVIRUS

Human papillomavirus (HPV) is the most common sexually transmitted infections in the United States. Sexually transmitted HPVs, common in adults and sexually active adolescents, more often than not are harmless and come and go without causing any symptoms. However, there is a subset of 19 high-risk HPV types that can lead to the development of cervical cancer and genital warts. Whereas genital warts can cause discomfort and psychosocial trauma, cervical cancer, if not detected in the early stages, can be deadly. Therefore, a vaccine that would protect against these diseases, especially cervical cancer, would be beneficial indeed.

In 2006, the FDA approved the first preventive HPV vaccine marketed by Merck and Co. under the trade name Gardasil. Gardasil, a recombinant vaccine (contains no live virus), is a preventive rather than a therapeutic vaccine and is recommended for women who are between 9 and 25 years old who do not have HPV. The vaccine will not protect a woman it she has been infected with HPV types prior to the vaccination, indicating the importance of getting immunized before potential exposure to the virus (before initiation of sexual activity). A series of three shots over a six-month period was shown to offer 100 percent protection against the development of cervical precancers and genital warts caused by the HPV types in the vaccine. The protective effects of the vaccine are expected to last a minimum of four and a half years after the initial vaccination.

The vaccine represents a significant advance in the protection of women's health. There are, however, a couple of drawbacks to the vaccine that have sparked debate. First, the vaccine is expensive. Second, the vaccine offers no protection against other specific types of HPV that also can cause cervical cancer (there are more than 120 known HPV types, and 27 are known to be transmitted through sexual contact). The vaccine targets two of the most common high-risk HPVs, type 6 and 18, which cause 70 percent of all cervical cancers, and HPV types 6 and 11, which cause about 90 percent of all cases of genital warts. Third, it is unknown whether the vaccine's protection against HPV-16, in particular, is long lasting. Fourth, because the vaccine works only against specific kinds of HPV, regular Pap tests should still be performed. And, fifth, the vaccine is targeted only to females, leaving males to serve as an asymptomatic reservoir for the virus.

Perhaps the most contentious issue of HPV vaccination is the recommendation to vaccinate young girls. Social conservative religious groups have publicly opposed the concept of making HPV vaccination mandatory for preadolescent girls because they fear that this might send a subtle message that sexual intercourse is acceptable, thus detracting from their abstinence-based position. Other critics question *mandating* the vaccine for young girls.[42, 43] They argue that the vaccine does not address a public health threat as serious as polio, measles, or other childhood illnesses and contend that the vaccine is too new for its long-term safety and effectiveness to be known. Adverse events related to Gardasil use have been reported, raising public concern, but the CDC and FDA recently determined that the serious events were not due to vaccination and that Gardasil was safe.[44, 45]

Many parents are extremely uncomfortable at the notion of vaccinating their young daughters against a sexually transmitted disease. But the reality is that the vaccine will not work after a woman has been infected, so the thinking is that it is preferable to have the young girl vaccinated before she becomes sexually active. Not surprisingly, there is heated debate as to whether the vaccinations should be required or recommended. Proponents argue that the objections are not strong enough to forgo the protection against a potentially dangerous disease.

HPV vaccination has been a contentious and complex issue in state legislatures as well. As of this writing, legislators in 41 states and the District of Columbia have introduced legislation to require, fund, or educate the public about the HPV vaccine, and 19 have enacted this legislation.[46] Texas was the first state to require vaccinating girls ages 11 and 12, through an executive order by Gov. Rick Perry, a conservative Republican.[47] This requirement was short-lived, though, as it was revealed that Merck had made campaign contributions to the governor and hired his former chief of staff as a lobbyist. Ultimately, within months, the Texas legislature overturned the requirement by a vote of 181 to 3.[48] In Illinois, when a legislator who had had HPV introduced a vaccination bill, a conservative group's blog speculated that she had been promiscuous.[49] Initially, Merck actively lobbied states to consider legislation, but this practice was ended in February 2007 amid significant criticism.[50] Currently, the only state with an active vaccination mandate for girls in middle school is Virginia. The Virginia legislation has also been dogged by controversy, as Merck recently announced it would invest heavily in a plant in Elkton, Virginia, to make Gardasil and other drugs.[48] Both Texas and Virginia offered broad freedom for parents to decline the shot, not requiring a medical reason to do so.[48] Whether other states will follow Virginia's lead remains to be seen. What is clear is that this new vaccine has been shown to have the ability to protect females from a serious and potentially deadly disease.

THE IMPORTANCE OF VACCINATION AND RESISTANCE TO IT

The marketing of Gardasil and similar prototypes in the pipeline illustrates that pharmaceutical companies and biotech companies are engaged in vaccine research. With the advancement of molecular biology and genetics, vaccine development continues to grow at an exciting rate. New and improved subunit vaccines that promise to offer increased safety and high efficacy are being studied. Additionally, novel strategies for vaccine delivery, especially the elimination of needles, as well as the combination of multiple vaccine components to different pathogens into a single vaccine delivery (of note, the MMR and the DPT vaccines) hold great promise. Yet, resistance to immunization among some groups of individuals persists. Indeed, vaccine development has had its share of political drama and controversy over time.

The American antivaccination movement, for example, focused on the concept of "inalienable rights," and its proponents argued that vaccination opposed the laws of nature and religious laws. Aggressive campaigns to repeal vaccination requirements were held in numerous states. Some antivaccination activists went so far as to argue that smallpox was not contagious, and a few tried to prove this by intentionally exposing themselves to the disease, usually with disastrous results. Taking a lead role in the antivaccination movement were the patent medicine manufacturers,

who feared (probably correctly) that vaccination laws would ruin their business. Whenever a new vaccine was introduced, including diphtheria and typhoid fever in the 1920s, polio in the 1950s, and measles, rubella, pertussis, hepatitis, and most recently HPV in the later part of the 20th century, groups opposed to vaccination would appear on the scene.

VACCINATION AND SCHOOL POLICY

The public health initiative to create school laws requiring vaccination began in the 1960s and I970s, after the polio epidemic of the 1950s. By 1963, 20 states required immunization as a requirement for school entrance, and this number grew to 29 by I970.[8] Many of these laws were created and enforced to protect against measles, in particular. Data showed that states with school immunization laws had 40 to 51 percent lower rates of measles than states without such laws.[51] These findings were compelling and provided the impetus for the remaining states to enact and enforce school immunization laws. As these laws are state based, variations exist in requirements and enforcement. By 2006, all states allowed medical exemptions, 48 had a provision for religious exemptions, and 19 permitted "personal belief" exemptions.[52] "Personal belief" exemptions refer to religious, philosophical, and any other undetermined exemptions that are not medical. Interestingly, a study looking into the effect of such exemptions on disease outbreaks found that states with personal belief exemptions" had a 27 percent higher rate of new pertussis cases than states without such an exemption.[53] Moreover, enforcement of school vaccination laws varied significantly at the local school level. Schools with simplified or inexplicit exemption claim procedures, as well as schools allowing philosophical exemptions, had increased exemption rates and higher risk of disease outbreaks.

PROTESTING VACCINES: FACT OR MYTH

Antivaccination movements often can have a significant effect on public health, primarily as a result of outbreaks of vaccine-preventable diseases. But, are the fears and concerns of the antivaccination movement unwarranted? What is the trade-off between benefit and risk? While 19th- and early-20th-century fears of vaccination might have been based on anecdotal horror stories, vaccine safety is a real and constant concern. New vaccines and vaccine combinations that provide a wider array of protection from diseases often require more injections, which, in rare cases, may lead to serious reactions. As with other pharmaceutical products, vaccines can produce side effects ranging from local injection-site soreness or redness to low-grade fevers to more serious adverse events. Therefore, safety concerns are not entirely off base. Public health experts believe that the system of routine childhood immunizations rests on a tenuous foundation of public support. Primarily because the success of immunization programs depends on parents' beliefs that vaccinating their children is safer than not doing so, it is imperative that parents and the public understand the risks as well as the benefits of vaccination. Parents should understand the consequences of choosing not to vaccinate. For some diseases, for example, pertussis, unvaccinated children are much more likely to contract the disease compared with children who are vaccinated.

Yet, all it takes is the hint of vaccine-safety controversy to scare off people from getting vaccinated. For example, in 1976, there was a scare that the swine influenza vaccine was associated with a severe paralytic illness called Guillain-Barre Syndrome (GBS). During the 1976–1977 swine influenza vaccination campaign, for example, 1,300 cases of GBS were reported to the CDC.[54] According to the CDC's vaccine information sheet on the influenza vaccine, if there is a risk of GBS from this influenza vaccine, it is estimated at one or two cases per million persons vaccinated, much less than the risk of severe influenza. Nevertheless, at that time, fear of developing GBS after a flu shot heightened distrust of getting an influenza vaccination. The overwhelming majority of people who get the influenza vaccine have no serious problem from it.

Also in the 1970s, when a suggested connection between the DPT vaccine and neurological damage in children was alleged, acceptance of this vaccine plummeted, resulting in a widespread resurgence of pertussis, especially in Great Britain where parents refused to have their children immunized.[55] Many parents who chose not to vaccinate their children doubted the reliability of vaccination information from authorities, believing that doctors overestimate protection and underestimate dangers of vaccines.[56] Furthermore, vaccines have been so effective that many parents, thankfully, have never seen cases of diseases against which vaccines protect, which in a sense diminishes the vaccine's perceived value and creates apathy.[57] A survey showed that 25 percent of those polled felt that children receive more vaccines than needed.[58] Between 1990 and 2000, for example, vaccines against four diseases (*Haemophilus influenza* type B, hepatitis B, chicken pox, and pneumococcal disease), entailing 10 to 12 injections, were added to the immunization schedule. Parents felt that children were becoming "pediatric pin cushions."[59]

The antivaccination movement, both in the United States and abroad, has been facilitated by the Internet, a fertile breeding ground for dissemination of information both correct and incorrect. Several studies have evaluated the information posted on the Internet, and the results are troubling. One study found that almost 43 percent of online sites about the MMR vaccine were negative and contained inaccurate and unbalanced information.[60] The most frequently cited incorrect information was that vaccines cause other illnesses such as neurologic disorder, multiple sclerosis, autism, asthma, and sudden infant death syndrome. Other common bits of information were that vaccines contain potentially large amounts of contaminants and mercury.[61]

The media, perhaps unintentionally, also fuels antivaccination sentiments. The power of the media to influence vaccination policy is illustrated by the impact of a British television documentary that aired in 1974 showing children allegedly harmed by vaccines. In the United States, too, a 1982 television special on the DPT vaccine included interviews with families alleging that their children were brain damaged after being vaccinated. By insinuating a cover-up, the media played into the fears of the antivaccination movement. Though heavily criticized by physicians and scientists, the show won an Emmy.

More recently, in 1998, British scientists, led by Andrew Wakefield, published an article in *Lancet* suggesting a connection between the MMR vaccine and autism.[62] Before the study (which was based on only 12 cases) could be adequately evaluated by the scientific community, the lay press in both the United Kingdom and the United States picked up on the article. The authors of the article eventually

retracted the assertion of a link between the MMR vaccine and autism, but the public's confidence in the MMR vaccine was certainly shaken. Immunization rates for MMR fell despite the fact that British and American governments advised parents not to reject MMR vaccinations. Wakefield's "research" was shoddy at best. His license to practice medicine was revoked as a result of the scandal.

Another controversy that engaged the antivaccination movement was the use of thimerosal, a mercury-containing organic compound, as a preservative to extend the shelf life of some vaccines. It was suggested that thimerosal in childhood vaccines could contribute to, or cause, a range of neurodevelopmental disorders in children, including attention-deficit/hyperactivity disorder. The critics argued that the ethylmercury-based preservative could cause serious side effects when administered to young children who have relatively undeveloped immune and neurological systems.

These concerns provided the impetus for the passage of the FDA Modernization Act of 1997, which called for a review and risk assessment of mercury-containing food and drugs. The FDA's Center for Biologics Evaluation and Research investigated the issue and found that some children could have exceeded the federal guidelines for single-dose mercury exposure, but the results were inconclusive. A 2004 IOM report on the subject concluded that the evidence did not support a causal relationship between thimerosal-containing vaccines and autism, although a congressional investigation did find evidence that thimerosal posed a risk.[63] The actual amount of thimerosal present in vaccines for children is listed, usually labeled as "trace" or nil. Currently, adolescent and adult tetanus vaccine and certain influenza vaccines still contain thimerosal.

ENSURING VACCINE SAFETY AND MONITORING: CHECKS AND BALANCES

The challenge is to reduce the number of injections and minimize the side effects without compromising effectiveness and patient acceptability. Before the FDA can license any vaccine, however, it must be assessed for safety and efficacy. Postlicensure studies continue to monitor vaccine safety. Given the potential for problems and adverse effects from vaccines, Congress passed the 1986 National Childhood Vaccine Injury Act, which was spearheaded by parents who were troubled by a putative link between vaccination and neurological problems. Essentially, the act was designed to reduce the potential financial liability of vaccine makers due to vaccine injury claims and established a no-fault system for litigating claims against vaccine manufacturers. Mounting potential liabilities totaling in the tens of billions of dollars posed financial threats to the pharmaceutical companies who produced vaccines. Vaccine makers indicated that they would cease production if this protection under the law was not enacted. The argument was that public health safety depended on the financial viability of pharmaceutical companies whose ability to produce sufficient supplies of vaccines could be imperiled by civil litigation on behalf of vaccine injury victims.

The act also mandated that all health care providers and manufacturers report certain adverse events following vaccinations to the Vaccine Adverse Event Reporting System (VAERS). Through VAERS, jointly operated by the FDA and the CDC to monitor the safety of licensed vaccines, experts look for patterns and any unusual trends that may raise questions about a vaccine's safety once it is used more widely in the population. The FDA continuously reviews and evaluates individual reports,

in addition to monitoring overall reporting patterns. The FDA monitors reporting trends for individual vaccine lots. Most reports come from health care providers, but anyone can report an unexpected event after vaccination to VAERS. VAERS' role is to generate new hypotheses about the cause of adverse events. For example, in August 1998, a vaccine against the rotavirus became available and infants were immunized. Within a few months, VAERS received reports that 15 infants developed a rare intestinal condition shortly after receiving the rotavirus vaccine. Although the number was very small in comparison to the number of infants who received the injection, analysis of the VAERS reports and other data suggested that the vaccine could be associated with an increase in the risk of this rare complication and in October 1999 the vaccine was discontinued pending further study.

The Clinical Immunization Safety Assessment Centers provide an additional level of scrutiny for selected patients whose symptoms or diagnoses may represent a new adverse event. The Vaccine Safety Datalink provides data from a variety of sources, including immunization records, hospital discharge records, and mortality data.

In 1988, a National Vaccine Injury Compensation Program was created. The program is a federal "no-fault" system designed to compensate those individuals or families of individuals who have been injured by childhood vaccines. A claim may be made for any injury or death thought to be the result of a vaccine covered under the program. The U.S. Department of Health and Human Services, the U.S. Court of Federal Claims, and the U.S. Department of Justice administer the program. But, the program stipulated that all claims against vaccine manufacturers could not be heard in state or federal court, but had to be heard in the U.S. Court of Federal Claims, often referred to as the "vaccine court." Cases are heard without juries, and awards damages are typically far below damage awards rendered in other courts.

Recently, the vaccine court has come under criticism,[64] with controversy surrounding the highly publicized case of Atlanta child Hannah Poling. In March 2008, it became public that the court had settled a case brought by Poling's parents, who claimed that she had been diagnosed with autism after receiving five vaccinations against nine infectious diseases in 2000. In the settlement, government lawyers concluded only that the vaccines aggravated an underlying mitochondrial disorder the child already had, which affects the ability of the body's cells to produce energy.[65] The CDC was quick to emphasize that compensating the Poling family was not an admission that vaccines may cause autism, but vaccine critics immediately claimed vindication, and many parents expressed newfound concern. Scientists have suggested that the vaccine court has reduced the required burden of proof that vaccines caused harm, ruling in favor of families even if the majority of epidemiologic evidence refutes their claims.[66] This debate continues, as the vaccine court is currently hearing the cases of 4,800 additional autistic children whose parents claim that thimerosal caused their disease, and the court was scheduled to hear cases about whether the MMR vaccine alone can cause autism.[64] In February 2009, the vaccine court ruled in the first three of these cases that neither the vaccine nor thimerosal was responsible for the children's autism.[67]

In 2006 the U.S. Fifth Circuit Court of Appeals ruled that plaintiffs suing three manufacturers of thimerosal could litigate in either state or federal court. This ruling was significant for the fact that the Fifth Circuit Court concluded that thimerosal is

not a vaccine but a preservative and that manufacturers cannot share in the protection afforded by the no-fault system of the National Childhood Vaccine Injury Act. Recently, the Georgia Supreme Court became the first to allow a state case against a thimerosal manufacturer to go to trial.[68]

Globally, the WHO has taken steps to ensure vaccine safety by establishing in 1999 the Global Advisory Committee on Vaccine Safety.[69] This committee is charged with advising the WHO on vaccine-related safety issues to enable WHO to respond promptly to issues of vaccine safety. The committee assesses the implications of vaccine safety worldwide and has weighed in on all of the important vaccine controversies, including MMR and autism, the safety of the mumps vaccine, thimerosal-containing vaccines, and the safety of influenza vaccination for pregnant women.

NEW CHALLENGES

The successful implementation of mass immunization programs and the subsequent eradication or reduction of smallpox, polio, measles, pertussis, meningococcal meningitis, diphtheria, mumps, rubella, and tetanus are among the most notable public health achievements of the 20th century. Yet, the path to the eradication of diseases by means of vaccination has not always been smooth. Efforts to develop an effective vaccine against tuberculosis (TB) so far have not been completely successful. The BCG vaccine is used in many countries, but not in the United States. The parasites responsible for malaria continue to challenge those working on a vaccine for this deadly disease that kills more than 1 million people worldwide each year and infects more than 300 million children a year.[70] Attempts to develop an HIV vaccine to target the retrovirus that causes this disease as well as a vaccine against malaria have so far ended in failure.

Other challenges remain. Pharmaceutical firms and biotech companies have little incentive to develop vaccines because the revenue potential is low, regulatory barriers are high, and the exposure to litigation is high should it be shown that adverse events are associated with the vaccine. For example, Warner Lambert (now Pfizer) stopped making Fluogen vaccine for influenza in 1998 primarily because of regulatory obstacles and financial loss. Some opine that this led to the flu vaccine shortage in the United States in 2004.[71] Although the number of vaccines administered has risen dramatically in recent decades, this increase probably is due to government mandates rather than economic incentives. Researchers and policy makers are calling for a different approach to motivate vaccine producers, including offering tax credits or guaranteed purchase, as well as other mechanisms to ensure an adequate vaccine supply and a financial return.[72]

Providing vaccines to the world is a necessary public health challenge that cannot be lost. Despite the monumental successes in vaccine development, the burden of infectious disease remains an important global concern. Fragmented delivery systems and difficulties in tracking and verifying immunization coverage, too, need to be addressed. Minimizing the difficulties in producing, distributing, and administering vaccines and ensuring the safety of the vaccine products should be every government's top priority for disease prevention.

Tremendous success has been achieved in eradicating smallpox, polio, and other childhood diseases, diarrheal disease remains one of the leading causes of

death in the 21st century, especially among poor children primarily in the developing world. In the developed world, children with severe viral diarrhea usually can be saved by the timely administration of intravenous fluids. In poor countries, they usually die. The rotavirus shots already are routine for babies in the Untied States and other developed countries. But, for those who are unvaccinated, the disease can and all too often does turn life-threatening and deadly. Seemingly intuitive, managing to vaccinate every child is not easy. The shots are expensive, the vials must be refrigerated (quite difficult in areas without electricity), and the ability to transport the vaccine to those in greatest need is more often than not challenging. Despite the difficulties, in early 2009, the WHO issued a recommendation that shots against rotavirus be given to every child in the world. Even if only a handful of countries implements a vaccination program, that would be a step forward in addressing a disease that kills 500,000 children a year. It is incumbent on all governments to ensure that the means and the resources be made available to build on the progress already made to eradicate vaccine-preventable diseases and thereby eliminate unnecessary human suffering worldwide.

REFERENCES

1. Barquest N, Domingo P. Smallpox: the triumph over the most terrible of the ministers of death. *Ann Internal Med.* 1997;127:627.
2. US Centers for Disease Control and Prevention. Ten great public health achievements in the twentieth century, 1900–1999. http://www.cdc.gov/od/co/media/tengpha.htm.
3. Parker AA. Implications of a 2005 measles outbreak in Indiana for sustained elimination of measles in the United States. *New Engl J Med.* 2006;355:1184.
4. Okonek BAM, Peters PM. Vaccines: how and why. http://www.accessexcellence.org.
5. Baxby D. *Vaccination: Jenner's Legacy.* Berkeley, UK: Jenner Educational Trust; 1994.
6. Parish HJ. *A History of Immunization.* Edinburgh, UK: Livingstone; 1965.
7. Gross CP, Sepkowitz K. The myth of the medical breakthrough: smallpox, vaccination, and Jenner reconsidered. *Int J Infect Dis.* 1998;3:54–60.
8. Salmon DA, et al. Compulsory vaccination and conscientious or philosophical exemptions: past, present, and future. *Lancet.* 2006;367(9508):436–442.
9. World Health Organization. Smallpox. http://www.who.int.
10. Hansen B. America's first medical breakthrough: how popular excitement about a French rabies cure in 1885 raised new expectations for medical progress. *Am Hist Rev.* 1998;103: 373–418.
11. Payette PJ, Davis HL. History of vaccines and positioning of current trends. *Curr Drug Targets Infect Disord.* 2001;1:241–247.
12. Cody CL, et al. Nature and rates of adverse reactions associated with DPT and DT immunizations in infants and children. *Pediatrics.* 1981;68:650–660.
13. Hennessen W, Quest U. Adverse reactions after pertussis vaccination. International Symposium on Immunization: benefits vs. risk factors. *Brussels Dev Biol Stan.* 1979;43:95–100.
14. Hinman AAR, Koplan J. Pertussis and pertussis vaccine: reanalysis of benefits, risks, and costs. *JAMA.* 1984;251:3109–3113.
15. Institute of Medicine. *DPT Vaccine and Chronic Nervous System Dysfunction: A New Analysis.* Washington, DC: National Academy of Sciences; 1999.
16. Alderslade R, et al. The National Childhood Encephalopathy Study: a report on 1000 cases of serious neurological disorders in infants and young children from the

NCES research team. In: Department of Health and Social Security, ed. *Whooping Cough: Reports from the Committee on the Safety of Medicines and the Joint Committee on Vaccination and Immunisation.* London: Her Majesty's Stationery Office: 1981.
17. Blume S, Geesink I. Essay on science and society: a brief history of polio vaccines. *Science.* 2000;288(5471):1593–1594.
18. Biddle W. *A Field Guide to Germs.* New York: Henry Holt; 1995.
19. World Health Organization. Pneumococcal conjugate vaccine for childhood immunization: WHO position paper. *Wkly Epidemiol Rec*. 2007;82:93–104.
20. Kung HC, et al. Deaths: final data for 2005. *Natl Vital Stat Rep*. 2008;56:1–120.
21. Grijalva CG, et al. National impact of universal childhood immunization with pneumococcal conjugate vaccine on outpatient medical care visits in the United States. *Pediatrics*. 2006;118:865–873.
22. Grijalva CG, Griffin MR. Population-based impact of routine infant immunization with pneumococcal conjugate vaccine in the USA. *Expert Rev Vaccines*. 2008;7:83–95.
23. World Health Organization. Pneumococcal conjugate vaccine for childhood immunization: WHO position paper. *Wkly Epidemiol Rec*. 2007;82:93–104.
24. Preventing pneumococcal disease among infants and young children: recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep*. 2000;49:1–35.
25. Whitney CG, et al. Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. *N Engl J Med*. 2003;348:1737–1746.
26. Hennessy TW, et al. Impact of heptavalent pneumococcal conjugate vaccine on invasive disease, antimicrobial resistance and colonization in Alaska Natives: progress towards elimination of a health disparity. *Vaccine*. 2005;23:5464–5473.
27. Direct and indirect effects of routine vaccination of children with 7-valent pneumococcal conjugate vaccine on incidence of invasive pneumococcal disease: United States, 1998–2003. *MMWR*. 2005;54:893–897.
28. Lexau CA. The changing epidemiology of pneumococcal pulmonary disease in the era of the heptavalent vaccine. *Curr Infect Dis Rep*. 2008;10:229–235.
29. Lexau CA, et al. Changing epidemiology of invasive pneumococcal disease among older adults in the era of pediatric pneumococcal conjugate vaccine. *JAMA*. 2005;294:2043–2051.
30. Haber M, et al. Herd immunity and pneumococcal conjugate vaccine: a quantitative model. *Vaccine*. 2007;25:5390–5398.
31. Fireman B, Black SB, Shinefield HR, Lee J, Lewis E, Ray P. Impact of the pneumococcal conjugate vaccine on otitis media. *Pediatr Infect Dis J*. 2003;22:10–6.
32. Palmu AA, et al. The seven-valent pneumococcal conjugate vaccine reduces tympanostomy tube placement in children. *Pediatr Infect Dis J*. 2004;23:732–738.
33. Poehling KA, et al. Population-based impact of pneumococcal conjugate vaccine in young children. *Pediatrics*. 2004;114:755–761.
34. Poehling KA, et al. Reduction of frequent otitis media and pressure-equalizing tube insertions in children after introduction of pneumococcal conjugate vaccine. *Pediatrics*. 2007;119:707–715.
35. Progress in introduction of pneumococcal conjugate vaccine: worldwide, 2000–2008. *MMWR* 2008;57:1148–1151.
36. Bliss, SJ et al. The evidence for using conjugate vaccines to protect HIV-infected children against pneumococcal disease. *Lancet Infect Dis*. 2008;8:67–80.
37. Global progress toward universal childhood hepatitis B vaccination, 2003. *MMWR* 2003;52:868–870.

38. Progress toward introduction of Haemophilus influenzae type b vaccine in low-income countries: worldwide, 2004–2007. *MMWR*. 2008;57:148–151.
39. Immunization Strategic Advisory Group of Experts. Meeting of the Immunization Strategic Advisory Group of Experts, April 2008: conclusions and recommendations. *Wkly Epidemiol Rec*. 2008;83(22):193–208.
40. Hicks LA, et al. Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998–2004. *J Infect Dis*. 2007;196:1346–1354.
41. Pichichero ME, Casey JR. Emergence of a multiresistant serotype 19A pneumococcal strain not included in the 7-valent conjugate vaccine as an otopathogen in children. *JAMA*. 2007;298:1772–1778.
42. Javitt G, et al. Assessing mandatory HPV vaccination: who should call the shots? *J Law Med Ethics*. 2008;36(2):214, 384–395.
43. Rockoff JD. Mandatory Gardasil vaccination would be unwise, academics say. *Wall Street Journal*. November 11, 2008.
44. Centers for Disease Control and Prevention. Reports of health concerns following HPV vaccination. http://www.cdc.gov/vaccinesafety/vaers/gardasil.htm. Published 2008.
45. Centers for Disease Control and Prevention. Information from FDA and CDC on Gardasil and its safety. http://www.cdc.gov/vaccinesafety/vaers/FDA_and_CDC_Statement.htm. Published 2008.
46. National Conference of State Legislatures. HPV vaccines. http://www.ncsl.org/programs/health/HPVvaccine.htm. Published 2008.
47. Blumenthal R. Texas is first to require cancer shots for schoolgirls. *New York Times*. February 3, 2007.
48. Rosenthal E. Drug makers' push leads to cancer vaccines' rise. *New York Times*. August 19, 2008.
49. Saul S, Pollack A. Furor on rush to require cervical cancer vaccine. *New York Times*. February 17, 2007.
50. Pollack A, Saul S. Merck to halt lobbying for vaccine for girls. *New York Times*. February 21, 2007.
51. Measles and school immunization requirements: United States. *MMWR*. 1978;27:303–304.
52. Omer SB, et al. Nonmedical exemptions to school immunization requirements. Secular trends and association of state policies with pertussis incidence. *JAMA*. 2006;296:1757–1763.
53. Salmon DA, et al. Exemptions to school immunization requirements: the role of school-level requirements, policies, and procedures. *Am J Public Health*. 2005;95:436–440.
54. Chen R, et al. Investigation of a possible association between influenza vaccination and Guillain-Barre syndrome in the United States, 1990–1991. *Post-Mark Surveill*. 1992;6:5.
55. Kaker JP. The pertussis vaccine controversy in Great Britain, 1974–1986. *Vaccine*. 2003;21:4003–4010.
56. Meszaros JR, et al. Cognitive processes and the decisions of some parents to forgo pertussis vaccination for their children. *J Clin Epidemiol*. 1996;49:697–703.
57. Petousis-Harris H, et al. Family physician perspectives on barriers to childhood immunization. *Vaccine*. 2004;22:2340–2344.
58. Gellin BG, Mailbach EW, Marcuse EK. Do parents understand immunizations? a national telephone survey. *Pediatrics*. 2000;10(5):1097–1102.

59. Offit PA, et al. Addressing parents' concerns: do multiple vaccines overwhelm or weaken the infant's immune system? *Pediatrics.* 2002;109:124–129.
60. Abbott VP. Web page quality: can we measure it and what do we find? a report of exploratory findings. *J Public Health Med.* 2000;22:191–197.
61. Zimmerman RK, et al. Vaccine criticism on the World Wide Web. *J Med Internet Research.* 2005;7:2.
62. Wakefield AJ, et al. Ileal-lymphoid-nodular hyperplasia, nonspecific colitis, and pervasive developmental disorder in children. *Lancet.* 1998;351(9(30):637–641.
63. Subcommittee on Human Rights and Wellness, Committee about Reform, US House of Representatives. *The Congressional Record*. May 21, 2003.
64. Offit PA. Inoculated against facts. *New York Times*. March 31, 2008.
65. Harris G. Deal in an autism case fuels debate on vaccine. *New York Times*. March 8, 2008.
66. Parker-Pope T. Will a 9-year-old change the vaccine debate? *New York Times*. April 12, 2008.
67. Harris G. Court hears more claims of vaccine-autism link. *New York Times*. May 13, 2008.
68. Court allows family to sue vaccine maker. *Augusta Chronicle*. October 7, 2008.
69. Folb PI, et al. A global perspective on vaccine safety and public health: the Global Advisory Committee on Vaccine Safety. *Am J Public Health.* 2004;94: 1926–1931.
70. McNeil DG. Malaria vaccine proves effective. *New York Times.* October 15, 2004.
71. Grady D. Before shortage of flu vaccine, many warnings. *New York Times.* October 17, 2004.
72. Stern AM, Markel H. The history of vaccines and immunization: familiar patterns, new challenges. *Health Aff.* 2005;24:611–621.

This page intentionally left blank

About the Editor

MADELON L. FINKEL, PhD, is professor of clinical public health and director of the Office of Global Health Education at the Weill Cornell Medical College in New York City. She also is course director in the Department of Public Health, coordinating the epidemiology/biostatistics, evidence-based medicine, and health care systems courses as well as the public health clerkship. Dr. Finkel's interests focus on women's health issues. Research projects include a cervical cancer screening program in rural India, reduction of preterm birth in Brazil, and better management of hypertension in Peru. She has served as consultant to numerous organizations, including law firms, pharmaceutical companies, and health care organizations. She serves on several boards of organizations whose missions are to promote public health. Her book *Understanding the Mammography Controversy: Science, Politics, and Breast Cancer Screening* was published by Praeger (2005). Praeger also published her book *Truth, Lies, and Public Health: How We Are Affected When Science and Politics Collide* (2007).

This page intentionally left blank

About the Contributors

Ronald D. Adelman, MD, is professor of medicine and co-chief of the Division of Geriatric Medicine at Weill Cornell Medical College in New York City. He is the director of the Irving Sherwood Wright Center on Aging and co-director of Cornell University's Center for Aging Research and Clinical Care.

Matthew Robert Anderson, MD, MSc, is an associate professor in the Department of Family and Social Medicine at the Albert Einstein College of Medicine and is an attending physician in the Department of Family and Social Medicine at the Montefiore Medical Center, Bronx, New York.

Melissa J. Bjelland, PhD, is a research associate at the Employment and Disability Institute in the School of Industrial and Labor Relations of Cornell University. Her background is in the fields of labor economics, health economics, and applied economics. The focus of her research is evaluation of the impact of changes in policies that advance the employment of individuals with disabilities.

John Bongaarts, PhD, is vice president and distinguished scholar of the Population Council in New York City. Dr. Bongaarts is a member of the U.S. National Academy of Sciences and the Royal Dutch Academy of Sciences. His research focuses on population issues, including the determinants of fertility, population-environment relationships, the demographic impact of AIDS, and population policy.

David H. Bor, MD, is with the Department of Medicine, Cambridge Health Alliance, and the Harvard Medical School, Cambridge, Massachusetts.

Gilbert J. Botvin, PhD, is a professor of public health and a professor of psychiatry at the Weill Cornell Medical College in New York City. He is also chief of the Department of Public Health's Division of Prevention and Health Behavior. Dr. Botvin is an internationally known expert on tobacco, alcohol, and drug abuse prevention who served as president of the Society for Prevention Research from 2001 to 2003. He is the founding editor of the journal *Prevention Science*.

Paula Brown, PhD, is a postdoctoral researcher in the Department of Earth and Atmospheric Sciences of Cornell University in Ithaca, New York. Her research interests focus on climate change with particular emphasis on extreme events.

Susanne M. Bruyère, PhD, CRC, is professor of disability studies, director of the Employment and Disability Institute, and associate dean of outreach at the School of Industrial and Labor Relations of Cornell University in Ithaca, New York. She is project director and co-principal investigator on several federally funded projects focusing on employment, disability policy, and effective workplace practices.

María Teresa Canto, DDS, MS, MPH, is program director, Community Network Program Disparities Research Branch of the Center to Reduce Cancer Health Disparities, National Cancer Institute in Rockville, Maryland.

Olveen Carrasquillo, MD, MPH, is associate professor of medicine and chief of the Division of General Internal Medicine at the University of Miami's Miller School of Medicine. His areas of interest include minority health, health disparities, community based participatory research, and access to care.

Gustavo D. Cruz, DMD, MPH, is adjunct associate professor at the Department of Epidemiology and Health Promotion at New York University College of Dentistry. From 1997 to 2001, Dr. Cruz was associate director of the National Institutes (NIH) funded Northeast Research Center for Minority Oral Health where he was also the principal investigator of one of the projects within the center. He is a diplomat of the American Board of Dental Public Health.

Allan J. Davison, MSc, PhD, is a faculty member of the faculty of science at Simon Fraser University in British Columbia, Canada. His research focuses on metabolism and nutrition and has been involved in developing a global health program at Simon Fraser University.

Arthur DeGaetano, PhD, is professor in the Department of Earth and Atmospheric Sciences at Cornell University in Ithaca, New York. He is also the director of the federally supported National Oceanic and Atmospheric Administration (NOAA) Northeast Regional Climate Center.

Michael P. Doyle, PhD, is the regents professor of food microbiology and director of the Center for Food Safety, University of Georgia, Griffin, Georgia.

Madelon L. Finkel, PhD, is professor of clinical public health and director of the Office of Global Health Education at the Weill Cornell Medical College in New York City. Her interests focus on epidemiology of women's health and global public health.

Emily Finkelstein, MD, is an assistant professor of medicine in the Division of Geriatrics at the Weill Cornell Medical College in New York City. She runs the Lipper Gruss Housecall Program, which is a section of the academic house call program at the New York-Presbyterian Weill Cornell Medical Center.

Sandro Galea, MD, MPH, DrPH, is professor of epidemiology at the Mailman School of Public Health at Columbia University.

Michele G. Greene, DrPH, is professor of public health in medicine at the Weill Cornell Medical College in New York City, professor in the Department of Health and Nutrition Sciences at Brooklyn College, and professor in the School of Public Health at the City University of New York.

Kenneth W. Griffin, MPH, PhD, is professor of public health at Weill Cornell Medical College in New York City. His research interests focus on the etiology and prevention of substance abuse and HIV risk behaviors among adolescents and young adults.

Joanna F. Haas, MD, MS, FACP, FACPM, FISPE, is vice president of global patient safety and risk management at Genzyme Corp. She is an internist and pharmaco-epidemiologist.

Stanley K. Henshaw, PhD, is a sociologist and senior fellow at the Guttmacher Institute in New York City. He has conducted numerous research studies on abortion and has authored or co-authored more than 70 articles and publications on abortion utilization in the United States and internationally, abortion policies and services, teenage pregnancy, unintended pregnancy, and family planning.

David U. Himmelstein, MD, MPH, was with the Department of Medicine, Cambridge Health Alliance, and the Harvard Medical School, Cambridge, Massachusetts. He is presently professor of public health, CUNY School of Public Health at Hunter College in New York City.

Nathaniel Hupert, MD, MPH, is associate professor of public health and medicine in the Department of Public Health at the Weill Cornell Medical College in New York City. He is also the director of the Preparedness Modeling Unit at the Centers for Disease Control and Prevention in Atlanta, Georgia. Dr. Hupert is a pioneer in the area of mathematical modeling, including modeling bioterrorism disasters.

Arun Karpur, MBBS, MPH, is a faculty extension associate at the Employment and Disability Institute in the School of Industrial and Labor Relations of Cornell University in Ithaca, New York. His research focuses on health disparities for people with disabilities.

Sandeep P. Kishore, MSc, is a master's/doctoral candidate in the Weill Cornell Medical College/Rockefeller/Sloan-Kettering Institute Tri-Institutional MD-PhD program. He successfully advocated for the inclusion of a cholesterol-lowering statin on the World Health Organization's Model List of Essential Medicines in 2007. He currently serves on the Board of Universities Allied for Essential Medicines (UAEM), is the first Lancet-Global Health Education Consortium Prize winner for community service, and received a 2008 Paul & Daisy Soros Fellowship for New Americans.

Tanja Kuchenmuller is a technical officer in the Department of Food Safety, Zoonoses, and Foodborne Diseases at the World Health Organization, Geneva, Switzerland.

Megan Landes, MD, MSc, is an emergency physician at the University Health Network in Toronto, Canada and is an assistant professor in the Department of Family and Community Medicine at the University of Toronto. Her research focuses on the mother-to-child transmission of HIV and its interaction with sexually transmitted infections.

Gal Mayer, MD, MS, is the medical director of the Callen-Lorde Community Health Center in New York City, which targets its services at the needs of the

lesbian, gay, bisexual, and transgender (LGBT) communities. Dr. Mayer is a board certified internist and a certified HIV specialist.

Sonal S. Mehta, MD, is an assistant professor of medicine and attending physician in the Department of Medicine/Division of Geriatrics at the Weill Cornell Medical College in New York City. She is assistant director of the Palliative Care Services at the New York-Presbyterian Hospital.

Nazanin Meshkat, MD, MHSc, is an emergency physician at the University Health Network in Toronto Canada and an assistant professor at the University of Toronto. Previously, she was stationed in Papua New Guinea with Doctors without Boarders. Her area of interest focuses on gender-based violence.

Marilyn D. Michelow, BA, is a medical student at Weill Cornell Medical College and a graduate of Princeton University. She has worked in Sub-Saharan Africa with the United Nations World Food Programme on issues of food security, and is interested in nutrition and equitable health care delivery.

Sarita A. Mohanty, MD, MPH, is with the Department of Medicine, Division of Geriatric and General Internal Medicine at the University of Southern California, Los Angeles, California.

Madhuvanti M. Murphy, DrPH, is an assistant professor of public health at the Department of Public Health, Weill Cornell Medical College, New York City. Her research focuses on adolescent drug use prevention, particularly the implementation of prevention programs in middle schools. She also has an interest in minority adolescent health.

Susmita Pati, MD, MPH, is with the Children's Hospital of Philadelphia and the Leonard Davis Institute of Health Economics at the University of Pennsylvania, Philadelphia, Pennsylvania.

Anita Radix, MD, MPhil, MPH, is the associate medical director of the Callen-Lorde Community Health Center in New York City, a center that primarily serves the lesbian, gay, bisexual, and transgender communities, and people living with HIV/AIDS. She is an internist and has a master's of public health in epidemiology.

Marcus M. Reidenberg, MD, is professor of medicine, professor of pharmacology, and professor of public health at Weill Cornell Medical College in New York City. He was elected chair of the 17th World Health Organization (WHO) Expert Committee on the Selection and Use of Essential Medicines in 2009. The WHO Essential Medicines list advocates that medicines meeting the priority medical needs of a country must be made available to everyone in that country.

Tony Rosen, MPH, MD, is a graduate of the Weill Cornell Medical College in New York. Prior to entering medical school, he earned a master's in public health at the University of California–Los Angeles where he studied epidemiology, concentrating in methodology and questionnaire design.

Jørgen Schlundt, MD, is the director of the Department of Food Safety, Zoonoses, and Foodborne Diseases at the World Health Organization, Geneva, Switzerland.

Victor W. Sidel, MD, is the Distinguished University Professor of Social Medicine at Montefiore Medical Center and Albert Einstein College of Medicine as well as adjunct professor of public health at the Weill Cornell Medical College in New York City. His areas of interest include social medicine and the health and environmental consequences of war.

Michael Silverstein, MD, MPH, is in the Department of Environmental and Occupational Health Sciences at the University of Washington School of Public Health and Community Medicine, Seattle, Washington.

Claudia E. Stein, MD, MSc, PhD, is in the Department of Food Safety, Zoonoses, and Foodborne Diseases at the World Health Organization, Geneva, Switzerland.

Robert V. Tauxe, MD, MPH, is the deputy director for science of the Foodborne, Bacterial, and Mycotic Disease Division of the Centers for Disease Control and Prevention.

David Vlahov, PhD, RN, is senior vice president for research and director of the Center for Urban Epidemiologic Studies at the New York Academy of Medicine in New York City.

Daniel Wattson, MD, received his undergraduate degree in operations research and industrial engineering. He then earned a medical degree and is currently a clinical intern at the Brigham and Women's Hospital in Boston, Massachusetts. His area of clinical interest is radiation oncology, and research interest focuses on Monte Carlo simulation for radiation treatment planning.

Steffie Woolhandler, MD, MPH, was a faculty member of the Department of Medicine, Cambridge Health Alliance and the Harvard Medical School, Cambridge, Massachusetts. She is presently professor of public health, CUNY School of Public Health at Hunter College in New York City.

Wei Xiong, PhD, is an assistant professor of public health in the Division of Outcomes and Effectiveness Research. Dr. Xiong received a bachelor's of science degree in mechanical engineering and a master's of science degree in systems engineering from Tianjin University in the Peoples Republic of China. He then earned a master's of science degree and doctorate in industrial and systems engineering from Rutgers, the State University of New Jersey. His research interests include simulation, optimization, Stochastic processes, and queuing theory, and their applications in public health emergency preparedness and response.

This page intentionally left blank

Index

This page intentionally left blank

Public Health in the 21st Century

This page intentionally left blank

Public Health in the 21st Century

Volume 2: Disease Management

Madelon L. Finkel, Editor

Foreword by David J. Skorton, MD

PRAEGER

AN IMPRINT OF ABC-CLIO, LLC
Santa Barbara, California • Denver, Colorado • Oxford, England

Library of Congress Cataloging-in-Publication Data

Public health in the 21st century / Madelon L. Finkel, editor ; foreword by David J. Skorton.
p. ; cm.
Public health in the twenty-first century
Includes bibliographical references and index.
ISBN 978-0-313-37546-0 (hard copy : set : alk. paper) — ISBN 978-0-313-37548-4 (hard copy : vol. 1 : alk. paper) — ISBN 978-0-313-37550-7 (hard copy : vol. 2 : alk. paper) — ISBN 978-0-313-37552-1 (hard copy : vol. 3 : alk. paper) — ISBN 978-0-313-37547-7 (ebook : set) — ISBN 978-0-313-37549-1 (ebook : vol. 1) — ISBN 978-0-313-37551-4 (ebook : vol. 2) — ISBN 978-0-313-37553-8 (ebook : vol. 3)
1. Public health. 2. World health. 3. Health policy.
I. Finkel, Madelon Lubin, 1949– II. Title: Public health in the twenty-first century.
[DNLM: 1. Public Health. 2. Disease Management. 3. Health Policy. 4. World Health. WA 100]
RA424.8.P83 2011
362.1—dc22 2010039064

ISBN: 978-0-313-37546-0
EISBN: 978-0-313-37547-7

15 14 13 12 11 1 2 3 4 5

This book is also available on the World Wide Web as an eBook.
Visit www.abc-clio.com for details.

Praeger
An Imprint of ABC-CLIO, LLC

ABC-CLIO, LLC
130 Cremona Drive, P.O. Box 1911
Santa Barbara, California 93116-1911

This book is printed on acid-free paper ♾

Manufactured in the United States of America

To my husband Arnold,
whose ideas, insight, and, most of all,
support and love are so important to me.

This page intentionally left blank

Contents

This page intentionally left blank

Foreword

As we enter the second decade of the 21st century, myriad issues compete for the world's attention, from the continuing stresses of the deep and widespread recession, to personal and national security, to climate change. But no issue looms larger than that of public health. Within this general rubric fall multiple issues critical to the individual, the country, and the world. Infectious diseases, including malaria, multiple drug–resistant tuberculosis, pandemic influenza, and HIV/AIDS pose even greater threats because of enormously increased international travel. In the developing world, the traditional diseases of poverty—communicable disease, especially infectious diarrhea and other waterborne diseases, malnutrition, and inadequate maternal and child health care—and displacement and violence, which are the sequelae of political instability, increasingly are being joined by the ailments of excess such as obesity, diabetes, and atherosclerotic cardiovascular disease, making the burden infinitely greater. Long overdue recognition of the worsening plight of women, particularly but not only in resource-poor environments, compounds the ongoing dilemmas of maternal-child health. Even within rich societies, such as the United States, shocking health disparities stubbornly continue.

Despite these daunting challenges, the tools of public health are more robust than ever. In addition to the traditional tools of medicine and the social sciences, the use of molecular genetics techniques and advanced statistical analysis presents new opportunities for the student and practitioner of public health. This comprehensive work on public health thus appears at a most opportune time.

Including a carefully assembled combination of original work and important recent literature and covering a huge sweep of relevant problems, *Public Health in the 21st Century* succeeds admirably in bringing together much of the broad field into one work that should find its place as a reference for public health workers and academics as well as policy makers and those in the private sector, whether health care providers, insurers, or drug or device manufacturers. Dr. Madelon Finkel, an experienced and recognized expert in several aspects of public health and, importantly, in the pedagogy of public health, has assembled a most impressive group of writers on a huge variety of public health topics, covering everything from global population health, to special needs cohorts, to

health care policy, to the often-ignored topic of public health teaching strategies and tactics. Readers from across the spectrum of public health concerns will find thought-provoking material of great value.

I commend Professor Finkel and her many colleagues on bringing to fruition a work that undoubtedly will receive wide use.

David J. Skorton, MD
President, Cornell University

Acknowledgments

This three-volume set could not have been produced without the contributions of the authors who so generously took the time to research and write their respective chapter. Most of the authors are my friends and colleagues who gladly agreed to accept my invitation to be included in the effort. I thank each of the authors for their time, effort, and especially their friendship.

My editorial assistant, Sophia Day, was tremendously helpful in organizing the huge volume of material and keeping track of missing information. Editorial reviews of many of the chapters were graciously and professionally done by Dr. Rebecca Finkel, an author and former editor who also happens to be my accomplished daughter. Technical computer work and support was provided by Jean Policard of the Department of Public Health at Weill Cornell Medical College. His assistance was invaluable to me.

Many thanks to my editor, Debbie Carvalko of Praeger, who invited me to write this multivolume text. Her support and faith in my being able to deliver the goods on time was reassuring.

This page intentionally left blank

Introduction

Madelon L. Finkel, PhD

Compiling topics for inclusion in a multivolume text on public health at first seemed like a simple task. Because so many significant advances have been made in disease prevention and health promotion, and so many public health initiatives have been put in place over the years to improve health and well-being, deciding which topics to select proved more difficult than anticipated. Which ones should be included? Which ones are the most relevant, the most important to highlight? Narrowing the focus, but being as comprehensive and inclusive as possible, seemed the most prudent way to proceed. And, therein lay the problem. How was I to select from such a wide array of public health issues to produce a comprehensive text on current public health topics? In an effort to be both comprehensive and inclusive, I endeavored to select as many important and timely subjects as possible for these three volumes. For fear of overwhelming the readers with chapters on every conceivable public health issue, a careful selection was made to highlight topics that represent and reflect the field of public health's breadth and scope. As such, the three volumes include chapters on topics reflecting advances and progress in knowledge and practice as well as challenges that remain. Naturally, many more topics could have been included. The essays selected for inclusion, many written specifically for this multivolume set and others reprinted from the published literature, represent a broad overview of important public health issues in the 21st century.

Charles-Edward A. Winslow, a bacteriologist and professor of public health at the Yale School of Medicine from 1915 to 1945, proposed a definition of public health as

> the science and art of preventing disease, prolonging life, and promoting physical health and efficiency through organized community efforts for the sanitation of the environment, the control of community infections, the education of the individual in principles of personal hygiene, [and] the organization of medical and nursing service for the early diagnosis and preventive treatment of disease.[1]

To a large extent, his definition has not been changed or amended over the ensuing decades. Public health's focus, then and now, is to safeguard the public's health and to handle threats to public health.

Public health has its roots in antiquity. It has long been recognized that polluted water and air, inadequate waste disposal, overcrowding and a concomitant lack of hygiene, and lifestyle behavior contributed to the spread of disease. Moving away from the miasma theory of disease, which argues that most diseases are caused by miasma (Greek for "pollution," that is, a noxious form of "bad air"), to the germ theory of disease was an important step in disease prevention. The 19th century witnessed so many discoveries and advances in the fields of medicine and public health. Essentially, pre-20th-century efforts focused on the eradication of infectious diseases and improvements in hygiene and sanitation, which led to a dramatic increase in average life expectancy. For example, the science of epidemiology probably dates from Dr. John Snow's identification of polluted public water wells as the source of the 1854 cholera outbreak in London. Hungarian physician Ignaz Semmelweis, for example, successfully reduced infant mortality at a Vienna hospital by instituting a disinfection procedure. His findings were published in 1850, but his work was ill received by his colleagues, who unwisely discontinued the procedure. Disinfection did not become widely practiced until British surgeon Joseph Lister "discovered" antiseptics in 1865, helped significantly by the work of the French chemist and microbiologist Louis Pasteur and German physician and bacteriologist Robert Koch.

The early 20th century expanded the scope and complexity of public health concerns. High rates of infant mortality led to the establishment of maternal and child health programs that emphasized nutrition. The disgraceful state of the food processing industry was notably depicted in Upton Sinclair's book, *The Jungle*. The book dealt with conditions in the U.S. meat-packing industry, causing a public uproar that partly contributed to the passage of the Pure Food and Drug Act and the Meat Inspection Act in 1906. High rates of occupational injuries and occupational-related diseases led to programs for industrial hygiene and occupational health, but it was not until 1970 that the U.S. Occupational Safety and Health Administration (OSHA) was created by Congress under the Occupational Safety and Health Act. Its mission is to prevent work-related injuries, illnesses, and occupational fatality by issuing and enforcing rules called standards for workplace safety and health.

These and other public health efforts contributed substantially to a dramatic decrease in mortality. From 1900 to 1940, for example, mortality rates in the U.S. fell by 40 percent, and life expectancy at birth increased from 47 years to 63 years.[2] No other period in American history showed such a dramatic decline in overall death rates. Nearly all of this decrease can be accounted for by reductions in infectious diseases, which in 1900 accounted for 44 percent of deaths. Contributing to the decrease in infectious diseases was the implementation of clean water technologies, one of the most important public health interventions of the early 20th century. At the turn of the century, waterborne diseases accounted for one-quarter of reported infectious disease deaths in urban areas. By 1936, less than 20 percent of deaths were due to infectious diseases. Perhaps the greatest public health feat was the worldwide eradication of smallpox, a

highly contagious, serious viral disease that was a worldwide scourge. The last case was recorded in 1977 in Somalia, and the eradication was certified by the World Health Organization (WHO) in 1979.

By the mid- to late-20th century, achievements saw a shift in focus from acute infectious diseases to the treatment and prevention of the growing burden of non-infectious, chronic diseases. English physician-researcher Sir Richard Doll and English epidemiologist and statistician Sir A. Bradford Hill pioneered the randomized clinical trial, and together were the first to demonstrate the connection between cigarette smoking and lung cancer. The focus on individual behaviors and risk factors (for example, antismoking campaigns) was an important step in addressing the *social determinants of disease*. The *new* public health sought to address the burden of chronic disease in a more comprehensive way by focusing on the effects of disease on vulnerable populations (for example, the elderly, the young, and the disabled), how health status differs among population groups (health inequalities), and how health care systems are organized and financed. Indeed, the challenges facing modern public health in the 21st century must be broad and inclusive, and focus on improvement in population health through the reduction of preventable diseases, both communicable and noncommunicable.

Looking back over the last century, public health is credited with adding 30 years to the life expectancy of people in the United States over the course of the 20th century; 25 years of this gain are attributable to advances in public health.[3] The Centers for Disease Control and Prevention (CDC) cataloged 10 of what it considered to be the most notable public health achievements based on the opportunity for prevention and the impact on death, illness, and disability.[4] These include (not ranked by order of importance) the following:

- Vaccination programs (as a result of widespread vaccine use, many of the infectious diseases that once killed so many have been almost eliminated);
- Fluoridation of drinking water (fluoride was first added to the public water system in 1945; tooth decay and tooth loss has declined substantially as a result);
- Occupational safety policies (since 1980, the rate of fatal occupational injuries has decreased by 40 percent);
- Access to safe, improved family planning and contraceptive services;
- Control of infectious diseases as a result of antibiotics, clean water, and improved sanitation;
- Food safety (safer and more healthful foods can be attributed to decreased microbial contamination and increased nutritional;
- Recognition of tobacco use as a health hazard;
- Motor vehicle safety (safety belts, child safety seats, motorcycle helmets, and engineering improvements in both vehicles and highways have helped reduce fatal motor vehicle accidents);
- Decline in deaths from coronary heart disease and stroke (lifestyle modifications and pharmaceuticals have led to a decline in deaths for these diseases); and
- Healthier mothers and babies as a result of better hygiene, prenatal health care and nutrition.

The 21st century presents new challenges. Largely preventable infectious diseases, such as tuberculosis (TB), polio, measles, and cholera continue to

plague millions of people around the globe, especially children. HIV/AIDS, which appeared on the scene in the mid-1980s, continues to be a major public health problem, although antiretroviral medications have done wonders in terms of extending life. Malaria, multidrug-resistant TB, and global outbreaks of viral diseases, most recently the H1N1 swine flu pandemic of 2009, continue to challenge public health efforts. That being said, chronic diseases such as diabetes and heart disease are now prevalent around the world. Obesity is not just a problem of the wealthy nations, as the increase in adult and childhood obesity in the developing world threatens to jeopardize progress. Statistics compiled by WHO show that chronic diseases are the largest cause of death in the world today and that global prevalence of all the leading chronic diseases is increasing, with the majority occurring in developing countries. Cardiovascular disease is already the leading cause of mortality in the developing world.[5] The increased burden of chronic diseases in countries that also have a high burden of infectious diseases is creating both a tremendous economic and public health strain. Furthermore, the recognition that health is affected by many factors, including genetics, economics, ethnicity and race, and geography, has necessitated a shift in focus in thinking. Public health in the 21st century must address these health inequalities to reduce the incidence of disease and improve health and well-being. Malnutrition, poverty, lack of access to health care, and so forth threaten to undermine the progress made in disease control and prevention.

In summary, over the past 150 years, much of the focus has been on disease control, understanding sources of contagion, and implementing programs to prevent the spread of disease. As scientific knowledge grew, public health's purview expanded to include maternal and child health care, health education, nutrition, aging of the population, the recognition of the role of behavioral factors in determining health, the impact of violence (domestic, civil, and international), health care disparities, and globalization. Indeed, increased globalization and technological advances have contributed to a worldwide economic, political, and social interdependence. In 1945, the United Nations Conference in San Francisco unanimously approved the establishment of a new, autonomous international health organization, the WHO, which came into being on April 7, 1948. The WHO was established as a specialized agency of the United Nations to serve as a coordinating authority on international public health issues.

Despite the progress made in improving the health, so much still remains to be done, not just in the United States, but also globally. In 2000, for example, 11.1 million children under the age of 5 died from preventable diseases such as diarrhea and acute respiratory infection.[6] These and other primarily preventable diseases kill more people each year than conflicts alone. Worldwide, poverty is one of the most significant causes of preventable mortality. Gender inequality persists and perhaps in some areas of the world actually has increased. Population growth remains a serious concern as the world's population has surged to 6.7 billion, most of the increase occurring within the last century. Environmental degradation and climate concerns have significant health implications. Each has a dimension that necessarily involves public health. As such, public health must be looked at in a global context if it is to be successful in fulfilling its mandate. Microbes have no boundaries, and we have seen over and over again, localized

outbreaks can quickly spread to national epidemics, and even worldwide pandemics.

Global health refers to health problems that transcend national borders and are of such magnitude that they have a global political, social, and economic impact. Assessing and measuring the impact of globalization on population health status should not be done in a vacuum; a global public health perspective needs to be integrated into health, social, and economic policies and programs to be effective. Reducing social and economic deprivation, reducing health inequalities, and improving health status go hand-in-hand. Domestic and international entities whose function and purpose is promulgating public health policy need to work together to achieve common goals. Collectively, progress can be made; individually, the effect is more muted. At their summit in 2000, heads of state of the G8 countries went on record as recognizing health as a global challenge and acknowledging that health is the "key to prosperity" and that "poor health drives poverty."[7] Following up on this challenge, G77 heads of state from 130 developing countries also expressed support for working toward the reduction of disease worldwide.[8] The motives and intent are laudable, but a decade later, the world still finds itself grappling with disease control and health prevention issues. Indeed, a WHO assessment of the capacity of 185 countries to prevent, conduct surveillance, and control disease showed that while health ministries had a high level of awareness of the issues, they had little or no allocation of significant resources to address the problems.[9]

Global nongovernmental organizations (NGOs) have played and continue to play a critical role in building capacity and sustainable development in specific areas of the world, although often the focus is narrow (for example, tobacco, TB, malaria, diet and nutrition, and so forth). Foundations, such as the Bill and Melinda Gates Foundation, have provided extraordinary sums of money and manpower to address pressing global health issues. Their importance and impact cannot be denied or ignored. These private investments in global health far exceed government assistance. The pharmaceutical industry also has the potential for being an effective player in the global health arena, but the industry is constantly criticized for not taking a greater role in the access to life-saving drugs, particularly in developing nations. The arguments are plentiful, ranging from focusing research and development efforts on health issues for rich countries to pricing drugs at unaffordable levels. That being said, despite the challenges that the pharmaceutical industry faces, it has been involved in a number of global health initiatives. Often, this means that companies are donating drugs, cutting prices, and developing partnerships with local governments and NGOs.

The World Bank and the United Nations play a major role in setting priorities for global health. The World Bank recognizes the negative effect of the increasing burden of disease, especially on the poor. Billions of dollars have been provided to countries for disease prevention. These efforts are crucially important as most developing countries have inadequate financing, lack of manpower, and poor infrastructure. Numerous UN organizations are specifically designed to focus on the global burden of disease. For example, the United Nations Population Fund, the United Nations Children's Fund, and many other UN agencies and organizations focus on providing assistance to the poorest

countries in an effort to "make a difference." The Millennium Development Goals (MDGs), also, are an agreed-upon set of goals that were developed in response to the world's main development challenges. They were drawn from the actions and targets contained in the Millennium Declaration that was adopted by 189 nations and signed by 147 heads of state and governments during the UN Millennium Summit held in September 2000. The MDGs are targeted to promote poverty reduction, education, maternal health, and gender equality, and to combat child mortality, AIDS, and other diseases. Poor countries pledged to govern better and invest in their people through health care and education. Rich countries pledged to support them, through aid, debt relief, and fairer trade. The MDGs represent a global partnership that has grown from the commitments and targets established at the world summits of the 1990s. The eight goals are to be achieved by 2015.[10]

Many stakeholders, public and private, are working toward similar goals, but it seems at times as if progress has taken one step forward and two steps back. While tremendous progress has been made over the past century, substantial challenges remain. Capacity development for prevention, treatment, and research remains weak; global economic factors impede progress; and the need for health systems change (delivery, financing, organization, and insurance coverage) remains unmet. The three volumes in this text were formulated to address public health issues from a national and global perspective. Volume 1 focuses on global population health issues, while volume 2 presents chapters on various aspects of determinants of health and disease, and volume 3 examines current public health policy issues, including ethics and human rights, public health education, and challenges we face as we enter the second decade of the 21st century. I made a concerted effort to include authors from around the world. Colleagues from Africa, Australia, Canada, China, Europe, India, Latin America, and New Zealand are well represented in this multivolume text. Their perspective and insight add a global dimension to the set.

VOLUME 1

Section 1 of volume 1 focuses on global population health issues. In their chapter on the global burden of disease, Kishore and Michelow carefully review the salient features of the global burden of disease, including its distribution and changing patterns over time. If current trends continue, diseases such as diarrhea, AIDS, TB, and malaria will become less important causes of morbidity and mortality as heart disease, cancer, diabetes, and traffic accidents increase in prevalence. Although the "burden" of a disease can be defined in a variety of senses, the consensus definition, particularly from the WHO, is a fairly specific one. The global burden of disease (GBD) as defined by the WHO is a comprehensive regional and global assessment of mortality and disability from 136 diseases and injuries and 19 risk factors. While useful, the thinking was that a better measurement of the GBD was needed, one that integrated morbidity, mortality, incidence, and prevalence into a single common metric that can be compared across time, space, and interventions. A new metric, the Disability-Adjusted Life Year (DALY), is a summary measure of population health, measured in units of time (years), combining estimations of both fatal and nonfatal health outcomes

(morbidity and mortality) to provide an estimate of the number of years of fully healthy life lost by an individual with a particular illness or condition. When DALYs are used to estimate the GBD, communicable diseases displace noncommunicable diseases as leading drivers of illness. The authors discuss the explosion of noncommunicable chronic diseases worldwide and the existing burden of communicable diseases, the combination of which poses a significant threat to the public's health. The challenge we face is how to best deal with the double burden of disease.

China and India together account for 37 percent of the world's population, about 6.8 billion. In 2025, India will surpass China in total population. India thus will have the distinction of being the world's most populous country. China's fertility rate is decreasing, whereas India's continues to increase. The United States is the third most populous country. Bongaarts focuses on population projections to the year 2050 for the world and major regions, and then identifies the demographic factors responsible for continued expansion of human numbers. Discussion focuses on policy options for slowing population growth in the developing world, where the growth continues to surpass that in the industrial world. Four main demographic factors contribute to future population growth: continued high fertility, declining mortality, young age structure, and migration. Bongaarts concludes by noting that the unprecedented pace at which the world's population has grown over recent decades has had an adverse impact on social and economic development, on health care, and on the environment. Despite substantial and partially successful efforts to reduce growth in the less developed countries, this expansion of human members is expected to continue at a rapid pace over the next decades with nearly all of this growth occurring in Africa, Asia, and Latin America. He advocates for three key strategies to reduce this growth rate: strengthen family planning programs to provide women with the knowledge and means to regulate their fertility; emphasize "human development," in particular education, gender equality, and child health; and encourage delays in subsequent childbearing.

The effect of urbanization on the public's health is discussed by Galea and Vlahov. The authors focus on the substantial change from how most of the world's population lives, reflecting on how the characteristics of the urban environment affect population health. The key factors affecting health in cities are considered within three broad themes: the physical environment, the social environment, and access to health and social services.

Continuing the theme of adverse effects of rapid population growth and urbanization, Brown and DeGaetano present a scholarly piece on the consequences of climate change on health. Concerns about recent changes in global climates and possible future trends on the health of the world's population are now considered important policy topics. With the election of Barack Obama, who has pledged a new era of leadership and responsibility to reduce the serious negative effects of climate degradation, the United States resumed its leading role in combating climate change and the adverse effects thereof. The United States is the world's largest source of cumulative emissions in the atmosphere, and as such, needs to lead the way for other nations to make a serious effort on climate change. Brown and DeGaetano make the case that climatic changes have, and will continue to have, direct negative health effects from altered weather patterns, but state

also that the indirect effects on agriculture and wider population systems are important factors for the GBD. Global warming (that is, melting of Arctic ice), extreme weather (for example, heat waves, cold spells), flooding (for example, Hurricane Katrina), erosion of ocean coastlines (that is, a result of extreme and heavy precipitation), and drought (for example, dust bowls) are leading to a disruption of food production and to disease. The authors note that exposure to infectious diseases has altered because of changes in temperature, humidity, rainfall, and sea-level rise. Specifically, some evidence of changes in the distribution of mosquito, tick, and bird vectors has been attributed to climate change. Mosquitoes, for example, can transmit diseases, such as malaria, dengue, yellow fever, and Japanese encephalitis, but their sensitivity to weather conditions can inhibit or enhance their efficacy as a vector. Malaria is spread by mosquitoes, which are inhibited from transmitting the parasite in cooler temperatures. Air quality and pollutants are affected by the weather and climate, and can cause negative health effects; the incidence of asthma has soared over the past decades. The authors caution that vulnerability to climate change will depend on responses to prevention, adaptation, adaptive capacity, mitigation, and future advances in disease control. It is clear that doing nothing will only make the situation worse than it already is.

The issue of global health and nutrition is a complex interplay of many factors ranging from politics to economics to food production policies to environmental degradation. Food is a basic human need. With roughly 1 billion humans suffering from overnutrition and a similar number unable to find enough food to subsist, no one seriously disagrees about the urgency of world hunger. One in six individuals does not get enough food to be healthy and to lead an active life. Hunger and malnutrition adversely affect physical and mental development; indeed, one might argue that hunger and malnutrition are leading risks to the health and well-being of individuals worldwide. Davison presents a comprehensive overview of the salient issues and focuses in particular on the interdependence of nutrition, economic development, and health. His discussion of the topic includes an assessment of the MDGs drafted to address the issue of alleviating hunger and malnutrition and a brief overview of some of the programs designed to eliminate global nutrition disparities, including the Millennium Village Projects, the Grameen enterprises and "microcredit" initiatives in resource-poor countries, and the role that foundations play in providing the financial means to reduce poverty and, in turn, to alleviate hunger and malnutrition.

No matter how one defines "health," prevention and treatment of disease is an essential prerequisite for achieving health and well-being. Implicit in this is that the right medicine be available at the time and place of need. Reidenberg presents an overview of the WHO Essential Medicines Program. By definition, an "essential drug" is a drug needed to satisfy the health needs of the majority of the population. The essential drugs concept of purchasing a limited list of essential drugs for a health service and making them generally available has been accepted by 156 countries and most, if not all, donor organizations. The essential medicines idea was developed to help limited resource countries make choices to use their medical resources for the greatest good for the greatest number. Thirty years after the WHO initiated the Essential Medicines Model List, four out of five countries have adopted a national essential medicines list. More than 100 countries have a

national drug policy in place or under development. Furthermore, a network of 83 countries provides global monitoring for adverse drug reactions and as well as for potential safety problems. Regarding pricing, 30 years ago, virtually no publicly available price information was available, and few countries actively encouraged generic substitutions. In the 21st century, at least 33 countries provide such information.

Prevention and safety have long been an integral component of public health. Section 2 of volume 1 presents several essays on the topic. By focusing on ways to control risks, public health works toward making the environment a safer and healthier place in which to live. Silverstein presents an historical account of occupational health and safety in the United States. The Occupational Safety and Health Act of 1970 (OSHAct) declared that every worker in America is entitled to a safe and healthful workplace, and that employers are responsible for work being free from recognized hazards. Now, forty years later, many of the promises of the OSHAct have yet to be met. Silverstein reviews the history of occupational health and safety in the United States and exposes the barriers to OSHA's success (predominantly linked to the statutory design of regulation, inspection, and education) and the challenges that remain in preventing injury and illness at the workplace.

Hupert, Wattson, and Xiong present a sophisticated analysis of the complexity of planning for and responding to public health emergencies. Using the example of a large-scale aerosol anthrax exposure over an urban locale, they explore key determinants of health outcomes and health system surge capacity using several modeling techniques (state transition, queuing network). They suggest that such models can provide valuable insights for forecasting the logistical and staffing needs of large-scale prophylaxis campaigns for a range of intentional and natural disease outbreaks, such as the 2009 influenza A (H1N1) pandemic. While all model-based studies have their potential limitations, they may serve many functions in emergency preparedness and planning that cannot be provided through other means.

Food safety has periodically seeped into the consciousness of the lay public, almost always after a public tragedy involving tainted food. In 2006, there were 1,270 reported U.S. foodborne disease outbreaks, resulting in 27,634 illnesses and 11 deaths. Since then, many other well-publicized incidents have involved the safety of food products, including food recalls (berries from California, grapes from Chile, and so forth), contaminated beef or poultry, and recalls due to contamination (peanuts, almonds, and pistachio salmonella outbreaks occurred in 2009). Furthermore, public concerns over the use of food additives as well as use of pesticides have spurred interest in organically grown food products. Perhaps in response to the recent food outbreaks in the United States, the Food and Drug Administration (FDA) recently named a highly qualified food safety expert to be deputy commissioner for foods at the FDA. The newly created position is the first to oversee all the agency's numerous food and nutrition programs, and setting safety standards for produce is a top priority.

The article on foodborne illnesses by Tauxe, Doyle, Kuchenmuller, Schlundt, and Stein focuses on this important topic. Foodborne diseases are caused by a broad variety of pathogens and toxins. In their comprehensive and scholarly

article, the authors review the epidemiological, microbiological, and public health aspects of foodborne diseases resulting from the ingestion of contaminated foods and food products, and discuss the evolving public health approaches to the global challenges of foodborne infections. The global challenge of safeguarding the world's food supply is complicated by growing international trade, migration, and travel. Through the globalization of food marketing and distribution, contaminated food products can and do affect the health of people in numerous countries at the same time.

Pharmaceutical safety in the United States is under the purview of the FDA. Haas presents an overview of how the benefits and risks of pharmaceuticals are managed and discusses the implications for global drug safety. As more pharmaceutical products are manufactured in countries around the world (including Canada, China, and India), concerns about safety of the products are quite valid. The FDA does not have the money or the manpower to inspect each plant overseas; yet, the final product is distributed and marketed in the United States. Haas provides an excellent historical overview of key issues in drug safety, which led to regulation of the marketing of approved drug products, most notably the Food and Drug Amendment Act of 2007. The act mandated that product label changes for safety were to be imposed and executed promptly. To ensure an acceptable benefit-risk balance, the FDA was empowered to require additional studies or trials, and it could stipulate specific conditions limiting the market availability of a product to ensure its safe use. The FDA was instructed to promptly communicate evolving product safety concerns even if the available information was limited. In addition, the act mandated that virtually all clinical trials, regardless of sponsorship, be registered and that efficacy and safety results be publicly posted in a timely manner. The act created a major new safety information system (the Sentinel System) that would complement spontaneous adverse event reporting. Despite efforts to tighten the mechanisms to ensure drug safety, the system is not fail-safe. The goal for drug safety is to have a flexible and responsive system able to recognize potential risks early, collect information efficiently, and take action that is appropriate in the context of both benefits and risks.

Focusing on the needs of vulnerable populations is an important component of public health. Section 3 of volume 1 highlights health care issues of special population groups. Karpur, Bjelland, and Bruyère from the Employment and Disability Institute of Cornell University highlight the role of public health in improving the health, well-being, and overall quality of life for people with disabilities through the consideration of epidemiological trends in disability prevalence, issues related to health disparities, the legal and regulatory environment affecting access to preventive and curative health services, methods of measuring and tracking the population of people with disabilities, and specific priorities in public health. The Institute of Medicine (IOM) refers to disability as ''the nation's largest public health problem,'' one that affects not only the health of people with disabilities, but also their immediate families and the population at large. Key issues for people with disabilities requiring attention in the U.S. public health system need to be addressed at the global level. It is estimated that there are approximately 650 million people with disabilities in the world with about 80 percent living in developing countries. The authors discuss various models and strategies to improve

health and well-being for people with disabilities, focusing on health disparities for people with disabilities; addressing the unique considerations for youth, women, and the aging with disabilities; and working toward an equitable access to health care, health care insurance coverage, health promotion, and prevention of secondary conditions—universal concerns that public health systems in all countries should be taking into account in the development of their national strategy.

Immigrant health care traditionally has largely been ignored by health policy makers. Yet, in 2009, an estimated 16 million children lived in immigrant families in the United States, representing one of the fastest-growing segments of the population. Clearly, policies and programs are needed to support immigrant parents and children, but the reality is haphazard at best. Mohanty, Woolhandler, Himmelstein, Pati, Carrasquillo, and Bor present compelling data based on the 1998 Medical Expenditure Panel Survey (MEPS) and found that immigrants have less access to health care and less health care use than do U.S.-born individuals. They also found that per capita health care expenditures for immigrants were far lower than expenditures for the U.S.-born. The study convincingly showed that the widely held assumption that immigrants consume large amounts of scarce health care resources is not supported by the data. The authors conclude that the low expenditures of publicly insured immigrants also suggest that policy efforts to terminate immigrants' coverage would result in little savings.

The provision of health care (or lack thereof) to those incarcerated has a long, sordid history. Finkel presents statistical evidence to illustrate the spectrum of health problems in correctional facilities. Inmate health and medical conditions range the gamut from minor (colds or viruses) to the significant (HIV/AIDS, TB). In addition to the communicable diseases, the prevalence of mental health and psychiatric diseases and substance abuse is higher among the prison population than the general population. The public health consequences of not paying attention to the health of prisoners can be quite significant; infectious diseases transmitted or exacerbated in prisons have the potential to become full-blown public health problems when prisoners return to their communities. The scope of this chapter provides an overview of the state of health among prisoners, assesses the provision of health care to those incarcerated, examines the policies regulating care of prisoners, including the challenges governments face in their ability to provide health and medical care to inmates, and discusses the pros and cons of having the private sector (privatization) involved in prison health care delivery.

Lesbian, gay, bisexual, and transgender (LGBT) health care also has received marginal interest and attention among policy makers and in the medical school curriculum. Medical education in the United States, both during medical school and in residency, is often unlikely to include adequate cultural competency related to the care of sexual orientation and gender identity minorities. A survey conducted to assess curricula in U.S. medical schools found that less than 3.5 hours were dedicated to teaching about health issues related to homosexuality. Part of the problem is the paucity of data on population demographics and health status for this population. For public health departments and providers to plan appropriate services for this vulnerable population, it is essential to have reliable data. Until recently, many of the research studies conducted in the LGBT community

were community-based studies using nonprobability sampling techniques. Radix and Mayer discuss the barriers to access to care as well as the health issues prevalent among the LGBT community. Of course, each group has its own set of health care needs, as lesbian health care is necessarily different from gay health care. The authors make the point that LGBT individuals have specific health needs that require targeted and culturally appropriate interventions.

The first ever surgeon general's report on oral health in the United States was published in 2000. The report highlighted a "silent epidemic" of dental and oral diseases, especially among the poor, the elderly, and children. Globally, too, oral disease burden and disability, especially in poor and marginalized populations, is a huge unmet issue. Oral health is much more than the pains of a toothache. Oral diseases such as dental caries, periodontal disease, tooth loss, oral mucosal lesions, and oropharyngeal cancers are major public health problems worldwide. Poor oral health has a profound effect on general health and quality of life. The burden of oral diseases and conditions is greatest among the economically disadvantaged, which include a disproportionately large number of racial and ethnic minorities and underserved populations. The major risk factors for oral disease are known and they are common with other chronic diseases: diet, smoking, alcohol, and risky behaviors. Canto and Cruz provide an epidemiologic overview of the state of oral health care as well as discuss preventive measures initiated to reduce dental caries, including exposure to fluoride (community water fluoridation, for example, has done much to reduce dental caries), use of dental sealants, practice of good oral hygiene, and reduction in sugar intake. The unmet need for dental care is a serious problem that needs to be acknowledged and addressed.

Taking care of the health care needs of the growing geriatric population is complex, challenging, and, to some extent, costly. Adelman, Finkelstein, Mehta, and Greene present an overview of the challenges of providing high-quality care to a rapidly aging population. They examine the medical, psychological, and social components of older age and explore the needs of this heterogeneous cohort. Issues such as dementia and Alzheimer's disease, elder abuse, ageist bias, the risks of polypharmacy, and long-term care issues are discussed.

Section 4 of volume 1 focuses on population-based prevention strategies. Adolescent substance abuse (alcohol, drugs, tobacco) has been well studied over the past decades; yet, the problem remains. Prevention and control programs have received considerable attention over the past decades as well. Botvin, Griffin, and Murphy, leaders in adolescent substance abuse prevention and cessation studies, raise a number of important issues related to adolescent substance abuse, including prevalence trends and types of prevention-based program modalities used by schools, families, and communities. The authors state that the most effective approaches target salient risk and protective factors, are guided by psychosocial theories regarding the etiology of substance use and abuse, and are implemented over many years. Many school-based prevention programs, for example, focus on skill-building in the area of drug resistance as well as life-skills training. While progress has been made in the field of substance abuse prevention, continued efforts must be made in the area of skill-building to prevent adolescents and children from taking drugs in the first place.

The issue of violence has been the subject of numerous reports by private and public organizations over the past decade. The public health consequences of all forms of violence are considerable as violence is associated not just with fatalities, but also with substantial morbidity and costs. It is estimated that in 2006 the health costs of violence (both fatal and nonfatal) in the United States exceeded $70 billion. Anderson and Sidel take a global approach to the discussion of violence and its sequela, and lay out a public health approach to violence prevention. They posit that the goals of public health—to prevent disease, and injury and premature death and to promote healthy living conditions for all—are identical to the goals of violence prevention. The disciplines and methods of public health—analyzing the causes of diseases, injuries, and premature deaths and of poverty and despair and determining methods to counter them—can strengthen efforts to prevent violence. And the ability of public health workers to gain trust both nationally and internationally can bring new skills and vigor to violence prevention.

Women, especially women in resource-poor nations, are an especially vulnerable group in terms of economics and in health care. In the industrial and developing world, gender-based violence (GBV) is endemic. Not only is it a major public health and human rights problem, but also for the victims it can, and most often does, have devastating personal, health, societal, and economic consequences. Meshkat and Landes eloquently delineate the types of GBV ranging from sexual, psychological, and physical, and depict the global burden of the problem. In addressing the issue, it is important to understand that public health initiatives often are bound tightly to existing legal frameworks, and this holds especially true in the case of GBV. It is excellent that GBV is now recognized as a major global public health and human rights issue, but efforts to stem its practice still stymie those involved in the prevention and management of GBV. Much work remains both in the industrial and developing world to ensure the safety and well-being of women of all ages against all forms of GBV.

Few issues in public health have fostered as much controversy as contraception and abortion. Passions run high on both sides of the debate; religion, politics, and policy regularly clash. Henshaw, who has spent his career conducting research in this area, presents a comprehensive statistical report on the issue of contraception and abortion, and explores some of the barriers that inhibit or even prevent women from controlling their fertility. Focusing on unintended pregnancy, contraceptive use, and abortion in the United States, he clearly and concisely presents the statistical evidence showing trends and highlighting the barriers that exist to prevent women of all ages from controlling fertility. Regarding abortion, since the legalization of this procedure in 1973, it is estimated that 35 percent of women in the United States will have had at least one abortion by age 45. Regarding birth rates, recent figures show that in 2007 more babies were born in the United States than in any other year in U.S. history.[11] This increase reflects a larger population of women of childbearing age. Births to teenagers (ages 15 to 17), after declining for many years, increased, reasons for which are poorly understood. Mississippi has the nation's highest teen pregnancy rate, which was 60 percent higher than the national average.

Although tremendous advances have been made in the eradication of once-deadly diseases, the development of vaccines probably is the most significant

reason for the decline in morbidity and mortality from such diseases. Perhaps the world's greatest achievement in this area is the eradication of smallpox. Rosen takes a global look at disease prevention through vaccination, presenting an historical overview and then focusing on the challenges that remain. There is a staggering disparity in vaccination efforts worldwide; millions of children are needlessly dying from preventable infections. Closing the gap will require multinational efforts and significant amounts of manpower and financial resources. The WHO and the Global Alliance for Vaccines and Immunization are deeply involved in coordinating immunization plans, especially in the resource-poor nations.

VOLUME 2

Volume 2 focuses on the determinants of health and disease. Section 1 addresses the treatment and prevention of chronic diseases. Since the mid-20th century, there has been a huge explosion in the number of individuals diagnosed with diabetes mellitus. Endocrinologist Baker's chapter focuses on the global epidemic of diabetes and discusses the public health, medical, and economic implications of dealing with this disease as well as the consequences to patients and to society. Diabetes is a growing and serious disease that affects rich and poor alike. According to the WHO, diabetes is likely to be one of the most substantial threats to human health in the 21st century. In the United States. alone, the direct medical costs of treating diabetics will be $336 billion. This does not take into account the growing proportion of overweight children and teenagers who are at high risk for developing diabetes and does not factor in immigration or the growing population of ethnic minorities who also suffer from diabetes at much higher rates than the U.S. white population. Without significant changes in public or private strategies, the burden of treating diabetics will place a significant strain on an already overburdened health care system. Ironically, and perhaps tragically, diabetes is among the most preventable of major illnesses. Clearly, as Baker discusses, much more needs to be done to stem the epidemic domestically as well as internationally. For a chance of success, prevention efforts must include a partnership between the individual and the health care provider.

Cardiovascular disease, too, is among the leading causes of morbidity and mortality globally. In 2004, according to the WHO, 17.1 million people died from cardiovascular disease, which represents 29 percent of global deaths. By 2010, cardiovascular disease is predicted to be the leading cause of death in developing countries. Kassahun and Borden explore the surge in cardiovascular disease and its risk factors worldwide, the characteristics and implications of this growth, as well as public health initiatives that can stem or even reverse this trend. They discuss the social, environmental, and cultural determinants of cardiovascular health, such as obesity, tobacco use, and access to health care that need to be addressed globally to reduce the incidence of cardiovascular disease.

The surging prevalence of obesity in the United States and around the world is growing faster than that of any other public health condition. This trend is alarming from a medical and an economic perspective. The ever-increasing prevalence of obesity has been accompanied by a host of inherently associated comorbidities. As a result, obesity is fast becoming the major cause of premature death

in the industrial as well as the developing world. Cardiologists Bornstein and Cooper examine the implications of the huge explosion of overweight and obesity in the world in the 21st century. Over the past two decades, the number of overweight and obese adults, adolescents, and children has increased dramatically. Of great concern are the children and adolescents who already have early obesity-related degenerative diseases, such as hypertension, dyslipidemia, metabolic syndrome, and type-2 diabetes mellitus, as well as manifestations of early preclinical atherosclerotic cardiovascular disease that previously has not been observed in this age-group. The economic costs of obesity are examined as are preventive means of addressing the epidemic. For example, health care spending for obese American adults soared 82 percent between 2001 and 2006.[12] Health care costs related solely to obesity could easily total $344 billion in the United States by 2018, or more than one in five dollars spent on health care, if the trends continue. The central message is that if nothing is done to stem the rise in obesity, the economic, medical, and personal consequences will be even more difficult to deal with.

The global burden of asthma is explored by Shirtcliffe and Beasley from the Medical Research Institute of New Zealand. The rising global burden of chronic, noncommunicable diseases over recent decades has been labeled "the neglected epidemic." Over recent decades, asthma has become one of the most common chronic diseases in the world and is now the most common chronic disease of childhood in many countries. The authors present a comprehensive epidemiologic overview of the disease and address the probably causes of the increase of asthma worldwide, including climate change and urbanization. The economic burden of asthma is considerable both in terms of direct medical costs, such as the cost of pharmaceuticals and hospital admissions, and indirect medical costs, such as time lost from work and premature death. The GBD to governments, health care systems, families, and patients is substantial. Indeed, the authors argue that the burden of asthma in many countries is of sufficient magnitude to warrant its recognition as a priority disorder in government health strategies.

Mild to moderate hypertension is generally an asymptomatic disease. It aptly has been called the "silent killer" because it usually produces no symptoms and increases gradually and slowly over the years. People with high blood pressure usually have no idea that they have this problem, and they do not go to the doctor specifically because of elevated blood pressure. The detection and treatment of hypertension is thus a major public health challenge. Cheung and Ong focus on the growing burden of hypertension worldwide, and especially in the United States. Although hypertension is seldom curable, the more practical aim is to control the blood pressure. The authors present an epidemiologic overview of the disease and discuss the known risk factors for hypertension. Medical management and improving compliance with treatment are discussed.

Arthritis, especially osteoarthritis (OA), is a prevalent condition among most of the older population. As the baby boomers age, the number of new cases of OA are likely to increase. Perhaps not surprisingly, there has been a concomitant increase in the number of joint replacements being performed. Lyman and Nguyen focus on OA of the knee and the explosion in the number of total joint replacements being performed. While total knee replacement, in particular, is an elective procedure to treat severe arthritis of the knee, the increase in this surgery is driven

both by the aging of the population as well as the obesity epidemic. Weight loss interventions may be the single most efficacious method of prevention of knee OA, but barring that, surgical intervention is increasingly being used to treat OA and enhance quality of life. The authors discuss the economic burden of OA, which is substantial (direct and indirect costs associated with OA).

Section 2 focuses on advances in cancer screening and the challenges that remain. For years, the dominant view about screening was that early detection and aggressive treatment would lead to increased longevity. Screening for cancer targets healthy, asymptomatic individuals. The purpose is to detect the disease at an early stage to initiate treatment, which hopefully will extend life. A key principle is that the potential benefits of screening should outweigh the harms of testing. The physical, psychological, and economic sequelae of follow-up testing should the screening test be positive needs to be compared with the number of lives saved as a result of screening.

Cancer is the nation's number two killer behind heart disease and accounts for nearly a quarter of annual deaths. The good news, however, is that cancer death rates and the number of new cancer cases in the United States continue to decline.[13] The conclusion drawn was that early detection and new therapies are major contributors to this effect. Almost at the same time as this report was released, a new study on the effectiveness of mammogram screening also was released. An expert panel from the U.S. Preventive Services Task Force recommended that mammography screening to detect breast cancer should be scaled back. This bombshell recommendation, in direct conflict with the recommendations from the American Cancer Society and other medical groups, caused considerable confusion, distrust, and even anger. Studies evaluating the effectiveness of the prostate-specific antigen (PSA) test to screen for prostate cancer also have yielded questionable results, making a clear recommendation for or against this test almost impossible. The American Cancer Society, a staunch defender of most cancer screening, has said that the benefits of detecting many cancers in particular prostate cancer has been overstated. The PSA prostate cancer screening test has not been shown to prevent prostate cancer deaths. The dilemma for breast and prostate screening is that it is not usually clear which tumors need aggressive treatment and which can be left alone.

Some studies focusing on routine early cancer screening found that the screening did not save lives, thus calling into question why screening was being advocated in the first place. In some cases, widespread screening increased the detection and treatment of small, slow-growing tumors that may well never have caused harm. In some cases, the tumor might regress or even disappear. While almost all of the cancer screening tests in wide use are minimally invasive, fairly inexpensive, and generally accepted by the public, none are 100 percent accurate; positive test results require further workup, which often are invasive and costly and usually lead to overtreatment. Screening does come with medical risks. In many cases, disease is not evident, such as in the case of false-positive test results.

Trevena presents a scholarly overview of the benefits and risks of screening and early detection of disease. She then examines the evidence for screening for colorectal cancer. The issue of whether an individual benefits from early detection of cancer is not as straightforward as it may seem. For some diseases, a preclinical

phase may be so short that the disease is not likely to be detected by screening. Or, even if detected, options for cure may not exist. Also, not every preclinical case will progress to clinical disease. Trevena examines the screening options for colorectal cancer, including fecal occult blood testing, flexible sigmoidoscopy, colonoscopy, and a new screening option, CT colonography. She presents the pros and cons of each modality. Some countries recommend that a fecal occult blood test be used while others advocate for colonoscopy. The accuracy of the test, including false-positive results, needs to be weighed against the potential benefit in reducing colorectal mortality.

Elkin and Blinder of the Memorial Sloan Kettering Cancer Center in New York City focus on breast cancer screening. Mindful of the current mammography screening controversy, the authors present a comprehensive overview of the advances made in reducing and preventing breast cancer. Because so many of the risk factors for breast cancer are not modifiable, much attention has been devoted to other means of breast cancer prevention such as understanding the role of hormones in breast cancer etiology. Much of the chapter focuses on the current controversy in mammogram screening. The questions of when screening should be initiated, on whom, and how frequently remain controversial. The authors present a scholarly assessment of the evidence, including a discussion on the realities of false-positive results. The chapter concludes with a discussion of advances in breast cancer treatment, including surgical treatment, radiation therapy, systemic therapy, hormonal therapy, and chemotherapy.

Lung cancer, the leading cause of cancer mortality worldwide, typically exhibits symptoms only after the disease has spread to other organs, unfortunately making it difficult to cure patients with such advanced disease. The overall prognosis of this cancer is poor when compared with other cancers, such as breast or colon, and is dependent on where the cancer is located, the size and type of tumor, and the overall health status of the patient. The two types of lung cancers, small-cell lung cancer and non-small-cell lung cancer, grow and spread in different ways and also have different treatment options. To date, screening for lung cancer is not advocated for these reasons. Yet, we have known for decades that tobacco smoking is the leading cause of this cancer, and evidence is quite clear that if individuals stopped smoking (or never started), the incidence of lung cancer would be greatly reduced. Lung cancer can be prevented. Mazumdar's chapter on lung cancer prevention, screening, and treatment reviews the epidemiology of this cancer as well as focuses on the "effectiveness," "efficiency," and "efficaciousness" of treatment regimens. A national initiative for comparative effectiveness research (CER) for clinical decision making is described. A discussion of CER provides a review of ongoing research and initiatives in this area, and highlights the gaps in information and research. Overall, much works is needed to find a cure for lung cancer and in being able to bring the best possible care to patients of all race, gender, and socioeconomic status.

Controversy over prostate cancer screening and treatment options continues to play out in the lay and professional media. Nguyen and Kattan's chapter reviews the current status of prostate cancer screening and assess its benefits and potential deleterious effects, to determine ways to improve its predictive accuracy and efficacy. To better understand the controversy surrounding prostate cancer

screening and perhaps offer a solution, they provide a review of current screening modalities, assess their accuracy and utility in contemporary medical practice, and suggest future directions for improvement of prostate cancer screening.

Over the past decades, the incidence of skin cancer has increased substantially. The chapter by Berwick, Erdei, Gonzales, Torres, and Flores focuses on the epidemiology and genetics of skin cancer and illustrates the growing public health burden of this particular form of cancer. Advances in screening and treatment are discussed and preventive measures are explored. The incidence of melanoma, a potentially deadly form of skin cancer, has soared over the past few decades perhaps because of an increased interest in screening for the disease. Indeed, the increase might be due to a growing tendency to identify and treat benign lesions as malignant cancers. It is quite difficult, and sometimes impossible, to tell a malignant lesion from a melanocytic nevus, a type of benign mole. The authors discuss how to protect oneself from skin cancer and provide informative information on sunscreens, tanning beds, and genetic susceptibility. Although sun exposure is the major risk factor for skin cancer, it is also necessary for synthesis of vitamin D, necessary for bone and muscle health and a possible protective factor for many diseases, including colon cancer. Given the worry about sun exposure and skin cancer, the question remains: How does one achieve favorable vitamin D levels yet also practice skin cancer prevention?

Cervical cancer, so easily and inexpensively prevented, remains a major killer among women in resource-poor nations. Without screening intervention, morbidity and mortality will continue to increase. Sankaranarayanan, Thara, Ngoma, Naud, and Keita have published groundbreaking research on the topic, and in this chapter they present a comprehensive overview of the issue, including evidence convincingly showing that screening for human papillomavirus (HPV) can yield a significant reduction in the numbers of advanced cervical cancers and deaths from this disease. They review the current status and future prospects for controlling cervical cancer in developing countries in this chapter. Low-tech screening methods (often used because most rural areas cannot realistically conduct Pap smear screening) and a single round of screening for HPV can and does result in a dramatic reduction in the incidence of advanced cervical cancer. There is a huge potential to reduce the cervical cancer burden by means of HPV vaccination. The authors acknowledge that a recommendation for HPV vaccination for adolescence women for a disease that occurs during adulthood is a major paradigm shift in cervical cancer control. Although HPV vaccination holds great promise, and has been licensed for use in more than 100 countries, there are several challenges (notably cost) for its widespread implementation through national immunization programs in high-risk developing countries. Cervical cancer reflects striking global health inequity, resulting in deaths of women in their most productive years in developing countries, with a devastating effect on the society at large. It remains as the largest single cause of years of life lost to cancer in the developing world.

Section 3 of volume 2 focuses on the treatment and prevention of infectious diseases. So much has been written about HIV/AIDS over the past two decades and so much progress has been made in extending life expectancy among those with the disease. Demars takes a broad view of the epidemic, tracing its history and focusing on the global burden of the disease. While recent trend data indicate

that the incidence in Africa has appreciably slowed, dealing with the disease's sociopsychological sequela and ensuring that progress made is not eroded remain challenges both in the industrial world and the resource-poor, hard-hit part of the world.

The WHO estimates that more than 500 million individuals worldwide are infected with the hepatitis B or C virus. Hepatitis viruses are found in every part of the world and often cause infections ranging in severity from acute infections that are asymptomatic to fulminate, chronic infections, which in some instances can lead to cirrhosis and hepatocellular carcinoma or even death. Aden presents a focused discussion on the most prevalent hepatitis viruses (A, B, and C) and explains how these diseases remain an important public health concern in both the developing and the industrial world. Whereas hepatitis B is a more serious type of infection than hepatitis A, hepatitis C infection can result in serious liver damage; hepatitis C is one of the leading causes for liver transplantation. While hepatitis A and hepatitis B are vaccine preventable, no vaccine is available for hepatitis C. Risk factors, population at risk, and treatment modalities are presented.

The ebb and flow of sexually transmitted diseases (for example, chlamydia, gonorrhea, and syphilis) has long been a focus of public health practitioners. All three diseases are preventable, treatable with medication, and, in the early stages, curable. Torrone and Peterman of the CDC present an overview of the topic and focus on syphilis specifically. The authors discuss the challenges of sexually transmitted disease (STD) control focusing on trends, efforts at prevention and control, and the challenges that remain. STD control and eradication is possible, but certainly not easy.

Section 4 of volume 2 addresses the treatment and prevention of mental health illness and disease. The burden of mental health disorders in the United States is substantial with approximately half of the population meeting the criteria for one or more such disorders in their lifetime and almost one-quarter meeting the criteria in any given year.[14] Treatment costs for mental disorders are substantial, rising from $35 billion (in 2006 dollars) to nearly $58 billion, making it the costliest medical condition between 1996 and 2006.[15] The most prevalent class of disorders is anxiety disorders (for example, phobias, panic disorders, and the like) followed by impulse-control disorders, mood disorders (for example, major depressive disorders, bipolar disorders, and the like), and substance abuse disorders (for example, alcohol abuse or dependence, drug abuse or dependence, and the like). The most prevalent type of disorder is major depressive disorders. Most individuals with a lifetime mental disorder had their first onset in childhood or adolescence. Little is known about the epidemiology of child mental disorders and controversy exists about how best to treat children. Contributors to this section focus on specific mental disorders, such as depression, suicide, and substance abuse.

Depressive disorders are prevalent conditions among the general population, and the medical, public health, and economic consequences of depression are considerable. Tedeschini, Cassano, and Fava present an overview of depressive disorders and focus on the recognition, management, and treatment of these diseases. The authors stress that depression is underdiagnosed and undertreated as only half of all Americans with depression receive treatment of any kind. Despite the

availability of numerous effective treatments, many depressive disorders are often misdiagnosed. Several factors contributing to the poor recognition of depression have been identified, ranging from the stigma of depression itself to the relative lack of systematic ascertainment of depressive symptoms by physicians.

Barber and Miller focus on the topic of suicide both within the Untied States as well as globally. In their informative and scholarly piece, they review the salient aspects of the epidemiology of suicide and the challenges posed by a purely clinical approach to its prevention. They outline a public health approach to suicide prevention, with an emphasis on reducing a suicidal person's access to lethal means of suicide. Their thesis clearly illustrates that although suicide is a global problem, a public health approach to prevention is still in its infancy. Public health strategies, such as changing cultural attitudes, increasing social support, improving access to high-quality treatment, and perhaps most important, reducing access to lethal means are measures that can and should be implemented.

Griffin examines the data on substance use and abuse across the life span from early adolescence to late adulthood. There is great diversity in patterns of alcohol, tobacco, and other drug use over the life course, with some individuals abstaining from use throughout their lives and others facing ongoing battles with substance abuse and dependence. The focus is primarily on substance use rather than abuse, because substance use is more prevalent than abuse in the general population and therefore has a greater public health impact. A goal of the chapter is to examine the extent to which substance use can be thought of as a developmental phenomenon not only among young people, but also throughout the life course. The implications of a life span developmental perspective to guide substance use prevention efforts are discussed. Griffin highlights a future challenge: the anticipated increase in substance use problems among the elderly and among the baby boomers, the eldest of whom will be reaching age 65 in a few short years. By taking into account how age-related developmental factors can affect substance use, we may be better able to address these and other new prevention challenges in the future.

VOLUME 3

Volume 3 shifts focus to health policy issues. In section 1, Finkel provides a historical overview of comparative health care systems illustrating why and how other industrial nations moved toward universal health care and why the United States did not. The organization, administration, financing, and delivery of health care in several countries are presented in an in-depth analysis. A critique of how health care is delivered and financed in other countries provides a stark contrast to how health care is provided and paid for in the United States.

Quality and patient safety, in addition to cost management, is an important issue in health care policy. Lazar, Dawson, Hyman, Collins, Regan, Kaplan, Green, Cook, and Graham from the New York-Presbyterian Hospital present an overview of quality assurance, quality metrics, and quality evaluation techniques. Performance improvement management methodologies designed to reduce medical errors and safeguard a safe workplace. In 1999, the IOM published a seminal report entitled *To Err Is Human,* which catalyzed an enormous shift in the understanding of medical errors. The IOM report defined an error as an event in which

there is a failure of a process to achieve the intended outcome, or where an incorrect process of care was selected initially. An adverse event was defined as an injury to a patient caused by medical management rather than the patient's medical condition. The IOM report concluded that medical errors were responsible for as many as 98,000 deaths in the United States annually. Estimated annual costs of these errors were in the range of $17 billion to $29 billion. The report further opined that injuries caused by errors are inherently preventable. Lazar and colleagues state that achieving better outcomes for patients, lowering overall costs, and improving the patient experience will require the continued investment of time and money to spur innovation and create reliable effectiveness, safety, and efficiency in clinical settings. Measurement and continuous performance improvement are the mainstays of a robust organizational quality assurance program.

Until the 1980s, most people with private insurance in the United States were covered by traditional indemnity plans. As remains the case, the vast majority got their coverage through employment-based plans provided as a tax-exempt benefit. These indemnity plans delegated shopping decisions about what care to buy and where to buy it to individual consumers and their physicians and then relied on consumer cost sharing to contain costs. Specifically, plans used deductibles and coinsurance to create financial accountability for purchases; the notion was that responsibility for resulting out-of-pocket payments would create incentives for cost-conscious shopping. By design, health plans were relegated to a passive role of paying the bills, while providers were reimbursed fee-for-service on a cost basis. Such open-ended insurance schemes laid the foundation for rising costs, which the United States is now trying to reign in. Managed care was an attempt to contain costs, but has not succeeded in doing so. White discusses an alternative, consumer-oriented strategy (Consumer Directed Health Plans, CDHPs) to address the concerns and shortfalls of managed care. The basic nation of CDHPs is that by placing consumers at risk for paying for substantial amounts of care with their own money, this simultaneously will restore control over shopping decisions and increase consumers' motivation for cost-conscious shopping, while introducing savings options will mediate the accompanying increase in exposure to financial risk. White provides an in-depth discussion of CDHPs and their potential effectiveness in managing health care cost increases.

The use of health information technology (health IT) has become an exciting and important field in medicine. Ancker, Kern, Patel, Abramson, and Kaushal present a scholarly overview of the present and future uses of health IT. Health IT has been promoted widely as a potential solution to managing the massive amounts of data and information as well as serving as a cost management tool. The authors discuss the various types of health IT systems and explore the barriers to development and implementation of these technologies. Health IT offers particularly exciting possibilities for improving the quality and efficiency of health care delivery by making essential individual-level medical data more readily accessible at the point of care; improving communication among clinicians, patients, and public health agencies; and providing evidence-based clinical decision support to help clinicians practice according to optimal care guidelines.

Section 2 of volume 3 focuses on the difficult issue of health care disparities. Health status and health outcomes vary markedly among racial and ethnic groups.

According to an IOM report *Unequal Treatment: Confronting Racial and Ethnic Disparities* race and ethnicity remain a significant factor in determining whether an individual receives high-quality care and in determining health outcomes. Race has been shown to be a determinant of the characteristics and qualifications of physicians who patients see, the types of hospital to which a patient is admitted, and the types of procedures they will undergo. The explanations are complex.

Boutin-Foster focuses on diversity and the public health implications of a growing racially and ethnically diverse America. She examines the role academic medical centers can and should play in providing care to this multicultural population. An argument is made for the need to bring the issues of cultural diversity to the forefront of medical education. While progress has been made in increasing the proportion of racial and ethnic minorities in the health care field, the racial and ethnic composition of the health care workforce does not match that of the general population. Would systematic biases in treatment be reduced if the composition of the workforce resembles more closely that of the patient population? While no studies have been done to empirically answer this question, given the extent of disparities and unequal treatment (which have been researched), one could assume that it certainly would not hurt.

Section 3 of volume 3 discusses ethics and human rights issues. Atkinson explores why human rights is crucial to the work of public health, and argues that human rights is a necessary framework for public health. Her chapter explores why a human rights framework is crucial to the work of public health. The human rights framework—in concert with traditional medical ethics—articulates certain values and standards that specify how we should conduct ourselves. She presents an argument for an ethical and legal framework for moving forward the global public health agenda. She believes that the human rights framework offers us a reason to believe in the possibility of change.

Bioethicist de Melo-Martin addresses the ethically charged topic of genetic testing and public health. She presents some of the most significant ethical concerns that arise in relation to the use of genetic tests, discusses matters related to the analytic and clinical validity and utility of genetic tests, and explains how these aspects result in ethical quandaries. She then focuses on the concerns that the use of genetic tests, if such tests prove beneficial for the populations' health, might contribute to furthering existing health inequities. Finally, she discusses ethical issues related to obtaining, or omitting, informed consent and to protecting privacy and confidentiality. Ensuring that people are not unjustly discriminated against because of their genetic or health status requires careful attention to issues of privacy and confidentiality; yet, concerns about privacy need to be balanced against the legitimate public health needs. Focusing on these ethical concerns when making public-policy decisions about implementation of genetic testing and screening is necessary if we want to use these medical technologies in ways that will advance the public's health.

As chronic diseases, including cancer, surpass infectious diseases as the primary causes of death, and as individuals are living longer with their diseases, providing timely access to consistently high-quality end-of-life care has become an important international issue. How we manage the dying patient has both medical and ethical concerns. Tickoo and Glare present a comprehensive overview of the

palliative care movement both in the United States as well as in Australia, England, and India. At some point, all humans have to confront the inevitability of end of life. How one prepares for the eventuality of death is a personal and individual matter. What is necessary and important, however, is that end-of-life choices be made clear and available. Providing for end-of-life care is emotionally difficult, thus making it even more imperative that all patients have the option of timely access to palliative care services that are both appropriate and cost-effective.

Section 4 of volume 3 focuses on public health practice and education. Trushin and Bang present an interesting chapter on the role of epidemiology and biostatistics in health news reporting. In their thoughtful piece focusing on the role of uncertainty in science, they provide a comprehensive overview of the mechanisms of research and statistical analysis. The scientific method is based largely on common sense, and statistical thinking involves concepts that are accessible to all: an acceptance of chance and uncertainty, an appreciation of context, an ability to detect logical and factual flaws in information and ideas, and the realization that science is a fluid process whereby new empirical evidence is accumulated every day. The true spirit of science requires a healthy skepticism, which means suspended judgment and the use of reason to evaluate the validity of research results. Science thrives on these qualities, because they lead to a search for knowledge and ensure that the scientific method remains self-correcting.

Evidence-based medicine has been incorporated into the medical school curriculum, but it also has a role in public health practice. The design and use of public health actions that are effective in promoting health and preventing disease underlie the growing field of evidence-based public health (EBPH), which emerged in the 1990s to improve the *practice* of public health. Maylahn, Brownson, and Fielding describe the concepts and principles underlying EBPH, the analytic tools to enhance the adoption of evidence-based decision making, the dissemination and implementation in public health practice, and challenges and opportunities for more widespread use of EBPH, especially through state and local health departments. Unlike solving a math problem, significant decisions in public health must balance science and art, because rational, evidence-based decision making often involves choosing one alternative from among a set of rational choices. By applying the concepts of EBPH outlined in their chapter, the authors concluded that decision making and, ultimately, public health practice can be improved.

The American Association of Medical Colleges (AAMC) is the umbrella organization for U.S. medical schools. The AAMC's position on medical curriculum has far-reaching impact. Maeshiro of the AAMC presents an informative historical overview of the tensions and barriers to integrating the disciplines of clinical medicine and public health. The challenge of incorporating public health content into the standard medical curriculum is not new. Not surprisingly, the roots of this struggle are entwined with the historical events and trends that led to the separation, or "schism" as some have described, between the practice of medicine and the practice of public health in the United States. She relates how over time the disciplines have gradually moved toward an integrated whole both at the medical school curriculum level and at the postgraduate medical training level. The rise of

a specialty in preventive medicine, the development of a residency in preventive medicine, and the subsequent creation of board certification in this area are interrelated. The framework in which medical education exists (for example, accreditation criteria for both medical school and residency training, national examination content) acknowledge the need for physicians to have a population perspective.

Section 5 of volume 3 addresses some of the challenges public health faces as we move into the second decade of the 21st century. Few areas of biomedical science have aroused as much controversy as embryonic stem cell research. With advances in medical research and technology, stem cell research has proliferated around the world. Cauley addresses the stem cell debate, focusing on the medical, ethical, legal, and political aspects of the topic. He provides a scholarly overview of the short history of stem cell research and raises important questions that need to be addressed today and into the future.

Advances in computer science have opened a new area of research for global disease monitoring. McEntee, Castronovo, Jagai, and Naumova from Tufts University provide an overview of a number of advanced computational and analytical techniques that open new opportunities to examine the role of forecasting disease transmission and manifestation. They review applications of various remote sensing (RS) techniques and present the relatively nascent epidemiological applications of this technology. Public health applications of RS data are no longer new; spatial epidemiology is equally important as the strictly environmental applications for which RS was originally intended. This is not surprising because environmental studies and epidemiology are inextricably linked. Each provides information on human health conditions and the corresponding management of environmental resources. Climate and land-use change and variability can be measured remotely and corresponding effects of alterations in natural and built environments can be predicted. Their scholarly and thoughtful presentation of this new field illustrates the tremendous opportunities that can be tapped.

Smith's concluding remarks on thinking creatively about public health in the 21st century is an excellent historical wrap-up of key events in public health over the centuries. He provocatively asks what the public health field needs to do to meet the challenges. How should public health be shaped for the 21st century, both for its own sake as a critical field for the world's well-being and for the sake of the local and global public it serves? The answers to these questions, of course, are multilevel and multifaceted. He advocates that 21st-century public health should begin to look more rigorously at the multiple factors in a society that predict health outcomes. These factors include economics, housing, nutrition, sports and recreation, education, spirituality, family structure, gender relations, childcare, transportation, and whatever other factors make up an integrated human life. Those concerned with the improvement of health on a local, national, and global scale need to work collectively rather than in isolation. Health, after all, is a product of the multiple facets of society, and, as such, requires a multifaceted approach to health promotion, the prevention and treatment of disease, and, most important, the improvement of the quality of life for all people.

This is an exciting time for public health. As public health practitioners continue to work toward improvement in the health and well-being of populations around the world and focus on disease eradication and the prevention and control

of diseases, injury, and disability, this increasingly is being achieved in a global context for the potential benefit of all. It is the aim of this multivolume reference text to identify and analyze the diversity of the work being conducted in the contemporary public health landscape.

The tremendous effort that went into creating this multivolume text could not have been done without the generosity of the contributing authors. My appreciation for their time, their enthusiasm, and their scholarship, and especially their friendship, cannot be underestimated.

Madelon L. Finkel, PhD
New York City

REFERENCES

1. Winslow C-EA. The untilled fields of public health. *Science*. 1920; 51:23.
2. D. Cutler D, Miller G. The Role of public health improvements in health advances: the 20th century United States. http://www.nber.org/papers/w10511. National Bureau of Economic Research working paper w10511. Published May 2004.
3. Bunker JP, Frazier HS, Mosteller F. Improving health: measuring effects of medical care. *Milbank Quarterly*. 1994; 72:225–258.
4. Ten great public health achievements—United States, 1900–1999. *MMWR*. 1999; 48(12): 241–243. http://www.cdc.gov/mmwr/.
5. World Health Organization. *The World Health Report 2003-Shaping the Future*. Geneva, Switzerland: World Health Organization; 2003.
6. Health: Confronting the Challenges. http://www.usaid.gov/our_work/global_health/home/confrontingfactsheet.html.
7. G8. G8 communiqué. http://www.g8.utoronto.ca/summit/2000okinawa/finalcom.htm. Okinawa 2000.
8. Group of Seventy-Seven. Final communiqué adopted by the thirty-fourth meeting of chairmen/coordinators of the chapters of the Group of 77. http://www.g77.org/news/pr062703.htm. June 26–27, 2003; Geneva, Switzerland.
9. Alwan A, MacLean D, Mandil A. *Assessment of National Capacity for Non-Communicable Disease Prevention and Control*. Geneva, Switzerland: World Health Organization; 2001.
10. United Nations Development Programme. Millennium Goals. http://www.undp.org/mdg.
11. National Vital Statistics Report. Births, Marriages, Divorces, and Deaths. Vol. 57, No. 12. March 18, 2009. www.cdc.gov/nchs/data.
12. Trends in health care expenditures by body mass index (BMI) category for adults in the US civilian noninstitutionalized population. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st247/stat247.pdf.
13. Annual report to the nation on the status of cancer: 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. In BK Edwards, eds. *Cancer*. Lawrenceville, GA: Surveillance Research Program, National Cancer Institute.
14. Kessler RC, Wang PS. The descriptive epidemiology of commonly occurring mental disorders in the United States. *Annu Rev Public Health*. 2008; 29:115–129.
15. The five most costly conditions, 1996 and 2006: estimates for the US civilian noninstitutionalize population. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st248/stat248.pdf.

This page intentionally left blank

SECTION 1

TREATMENT AND PREVENTION OF CHRONIC DISEASE

This page intentionally left blank

CHAPTER 1

The Global Burden of Diabetes: Public Health Implications and Challenges

Jason C. Baker, MD

Diabetes mellitus (DM), commonly referred to as diabetes, is a disorder in which the individual has unusually high blood sugar levels and an erratic metabolism. The high blood sugar levels are caused either by low levels of the hormone insulin or by resistance to insulin's effects, plus insufficient levels of secreted insulin. Diabetes symptoms often include excessive urination, excessive thirst, increased fluid intake, and blurred vision. It is a serious and debilitating disease with significant public health, medical, and economic consequences.

Diabetes is not a modern disease. The first known reference to this disease dates to 1552 BCE when Egyptian physician Hesy-Ra described a disorder of "polyuria." In 50 BCE, Indian physician Susruta noted that patients with this disease had a "sweetness" of the urine, and in the first century BCE, ancient Greek physician Aretaeus of Cappadocia coined the term diabetes, meaning "one who straddles" or "siphon." Both terms are appropriate for a disease that causes patients to pass excessive amounts of urine. Centuries later, in 1675, English physician Thomas Willis added the word "mellitus" (Latin for "honey") to the term diabetes, due to the sweet taste of diabetic urine. Despite its long history, it was only in the early 20th century that a clearer understanding of the disease was possible. Based on animal research, it was shown that the pancreas played an important role in the development of this disease. Specifically, the isolation of insulin, a hormone produced in the pancreas, by Banting, Best, Macleod, and Collip in 1921–1922 led to the treatment of DM. In 1922, for the first time, insulin injections were used to treat diabetic patients. Banting (who was knighted for his discovery) and his colleagues were awarded the 1923 Nobel Prize in Physiology or Medicine for this discovery.[1–3]

Since the mid 20th century, there has been a huge increase in the number of individuals diagnosed with DM. This chapter focuses on the global epidemic of diabetes and discusses the public health, medical, and economic implications of dealing with this disease as well as the consequences to patients and to society. Diabetes is a growing and serious disease that affects rich and poor alike.

According to the World Health Organization (WHO), diabetes is likely to be one of the most substantial threats to human health in the 21st century.

OVERVIEW OF THE DISEASE

In nondiabetic individuals, glucose is exquisitely regulated with fasting glucose levels between 70 to 100 (as defined by glucose levels following eight hours of no caloric ingestion). Peak postprandial glucose rarely rises higher than 140, and glucose levels return to a fasting level within two to three hours after food ingestion.[4, 5] A specific test to measure blood glucose levels (HgA1c) provides a snapshot of average daily blood glucose and provides an indication of how well blood sugar levels are being controlled. Normal HgA1c levels range between 4 and 6 percent, with mean values of 5 percent.[6] If the fasting glucose falls between 100 and 125, or the postprandial glucose test results fall between 140 and 199, individuals are classified as prediabetic. When fasting glucose values exceed 126 or a random or postprandial glucose value exceeds 200, an individual is considered to be diabetic.[7]

There are three major types of diabetes: type 2 diabetes (previously known as non-insulin-dependent diabetes) is the most common form of DM, affecting 85 to 90 percent of the world's diabetic population. Also known as late-onset diabetes, it is characterized by insulin resistance and relative insulin deficiency. The disease is strongly genetic in origin, but lifestyle factors such as excess weight, inactivity, high blood pressure, and poor diet are major risk factors for its development. Symptoms may not show for many years and, by the time they appear, significant medical problems may have developed. Insulin injections may be required if oral medication is not sufficient to control blood glucose levels.[8]

Type 1 diabetes (previously known as insulin-dependent diabetes), is an autoimmune disease where the body's immune system destroys the insulin-producing beta cells in the pancreas. This type of diabetes, also known as juvenile-onset diabetes, accounts for 10 to 15 percent of all people with DM. The disease can appear at any age, although more commonly it affects individuals under the age of 40 and is triggered by environmental factors, such as viruses, diet, or chemicals in people genetically predisposed. People with type 1 diabetes must inject themselves with insulin several times a day and follow a careful diet and exercise plan.[8]

Gestational diabetes mellitus (GDM) is diagnosed during pregnancy through an oral glucose tolerance test. It is estimated that between 5.5 and 8.8 percent of pregnant women develop GDM. Risk factors for GDM include a family history of diabetes, increasing maternal age, obesity, and being a member of an ethnic group with a high risk of developing type 2 diabetes. Although the woman's glucose tolerance test levels usually return to normal after the birth, the mother has a significant risk of developing type 2 diabetes during her lifetime. The baby, too, has a higher risk of developing diabetes in later life and also has a higher risk of becoming obese, a major risk factor for diabetes.[8]

Type 2 Diabetes

Type 2 diabetes is a now considered to be a worldwide epidemic. Globally 246 million people are afflicted with this type of diabetes and another 314 million are estimated to be prediabetic. By 2025, the number of diabetics is projected to

soar to a staggering 380 million people, and the number of prediabetics to increase to 500 million.[9–11] According to the WHO, another 7 million people develop diabetes each year. As of 2007, the countries with the largest number of reported cases of diabetes were India (40.9 million), China (39.8 million), the United States (19.2 million), Russia (9.6 million), and Germany (7.4 million).[9–11] The WHO reports that by 2025 the largest increase in diabetes prevalence will be seen in developing countries, especially India and China. In 2007, the five countries with the highest diabetes prevalence in the adult population were Nauru (30.7 percent), United Arab Emirates (19.5 percent), Saudi Arabia (16.7 percent), Bahrain (15.2 percent), and Kuwait (14.4 percent).[9–11]

This disease is the fourth-leading cause of global death. According to the International Diabetes Federation (IDF), in 2007, 3.8 million deaths annually were thought to be secondary to type 2 diabetes, which translates into 6 percent of total global mortality (a percentage on par with HIV/AIDS). Every second, 10 people die from complications attributable to diabetes.[12]

In the United States, the number of individuals with type 2 diabetes is rapidly increasing. As of 2008, 24 million people (accounting for 7.8 percent of the entire U.S. population) were diabetic, an estimated 5.7 million of whom are undiagnosed. Another 57 million people suffer from prediabetes.[10–11] By 2050, if the trend continues, it is estimated that more than 29 million Americans will be diagnosed with type 2 diabetes.[12–17] For individuals born in the year 2000, the estimated lifetime risk for developing type 2 diabetes is 33 percent for males and 39 percent for females.[18] These figures are alarming from a medical and a public health standpoint, and they will be discussed later in this chapter.

The epidemiology of type 2 diabetes clearly shows that the disease affects both the young and old and crosses racial and ethnic lines. Whereas in the past older individuals were more likely to be diagnosed with diabetes, more recently the data show a clear and disturbing trend among younger individuals developing diabetes. Indeed, type 2 diabetes is now being diagnosed at progressively younger ages especially among children and adolescents. While no ethnic group is spared from type 2 diabetes, the Centers for Disease Control and Prevention (CDC) data show that the prevalence of type 2 diabetes in the United States is highest among certain groups such as Hispanic, Mexican, and Latino Americans; African Americans; Native Americans; and Asian and Pacific Islanders.[19]

Type 2 diabetes and obesity are intimately linked, as obesity is the most important risk factor for the development of type 2 diabetes. Not surprisingly, the rising incidence and prevalence of diabetes is directly correlated with the rise in the incidence and prevalence of obesity. Obesity is defined by a Body Mass Index (BMI) >30, with values 25 to 30 deemed "overweight." Evidence suggests that the risk for development of type 2 diabetes increases progressively from a BMI of 21.[20]

According to the WHO, in 2005 approximately 1.6 billion adults (over 15 years of age) worldwide were overweight and 400 million adults were obese. By 2015, the WHO projects that 2.3 billion adults will be overweight and more than 700 million adults will be obese.[20] Children are not spared, with approximately 20 million children under 5 years of age worldwide deemed overweight in 2005. Data from the National Health and Nutrition Examination Surveys (NHANES) (1976–1980 and 2003–2006) demonstrate that the prevalence of childhood obesity

in the United States is dramatically increasing: for children ages 2 to 5 years old, prevalence increased from 5.0 percent to 12.4 percent; for those ages 6 to 11 years, prevalence increased from 6.5 percent to 17.0 percent; and for those ages 12 to 19 years, prevalence increased from 5.0 percent to 17.6 percent.[21–22] The average American is now 23 pounds overweight; in 2005–2006, more than one-third of all U.S. adults (approximately 72 million people) were obese (33.3 percent of all men and 35.3 percent of all women were designated as being obese).[23]

There are marked race-ethnic disparities in obesity prevalence. In 2005–2006, approximately 30 percent of non-Hispanic white adults were classified as obese, while 45 percent of non-Hispanic black adults and 36.8 percent of Mexican American adults were so classified. These racial differences in obesity rates were most pronounced in women. Approximately 53 percent of non-Hispanic black women and 51 percent of Mexican American women ages 40 to 59 are obese in comparison with 39 percent of non-Hispanic age-matched white women. These racial and ethnic differences in obesity rates manifest at an early age with the prevalence rate of obesity higher among adolescent Mexican American boys (22.1 percent) and African American boys (18.5 percent) than among non-Hispanic white boys (17.3 percent).[20–21] Non-Hispanic black girls had the highest prevalence of obesity (27.7 percent) compared with that of non-Hispanic white (14.5 percent) and Mexican American (19.9 percent) girls.[21–22]

The concept of thrifty genes, proposed by geneticist James Neel in 1962, may explain some of the global obesity trends. According to Neel, the "thrifty" genotype would have been beneficial for hunter-gatherer populations, because it would allow them to fatten more quickly during times of abundance and better survive subsequent times of famine. When individuals carrying such genes are faced with the caloric excess seen in modern societies, however, this energy is stored all too efficiently, leading to obesity and concurrent issues such as type 2 diabetes.[24–25] The thrifty genotype hypothesis has been cited to explain the growing levels of obesity and diabetes among ethnic groups newly introduced to Western Hemisphere diets and lifestyles, including Latinos and Hispanics, South Pacific Islanders, Native Americans in the southwestern United States, and Sub-Saharan Africans. Neel postulated that Europeans did not develop thrifty genes because they evolved in areas where famines were less common, which explains why rates of obesity are lower in this population.[24–25]

Complications of type 2 diabetes are considerable and potentially life threatening. Many if not most diabetics develop vascular, cardiac, and neurological damage from the disease. Cancer rates and mortality are also higher in diabetics.[26] According to the WHO, on average, individuals with type 2 diabetes die 5 to 10 years earlier than those who do not have diabetes, primarily due to complications from the disease. Cardiovascular disease, kidney failure, eye complications (including diabetic retinopathy and adult blindness), nontraumatic limb amputations, nerve damage, and stroke are unfortunate and frequent byproducts of this disease.[26–30] Complications arise from the cumulative and caustic effect of hyperglycemia and are further fueled by such conditions as dyslipidemia, obesity, hypertension, and smoking.

Macrovascular damage leads to coronary, cerebral, and peripheral vascular disease. According to the WHO, those with type 2 diabetes are more than twice as likely to have a heart attack or stroke as people who do not have diabetes.[27–31]

Diabetes is a coronary artery disease equivalent, as people with type 2 diabetes are just as likely to suffer a heart attack as people without diabetes who already have had a heart attack. In fact, 50 to 80 percent of deaths among diabetics are a result of cardiovascular disease.[32] Diabetes, too, is a leading cause of nontraumatic lower extremity amputations.[32] Microvascular damage leads to vascular insufficiency leading to retinopathy, nephropathy, and neuropathy. Studies suggest that up to 11 percent of adult diabetics suffer from retinopathy, with 2.5 million people worldwide afflicted by diabetic retinopathy.[27–31] According to the WHO, diabetic retinopathy is the leading cause of blindness in adults 20 to 65 years of age in the developed countries.[20] Diabetic nephropathy is the leading cause of kidney failure in developed countries, with incidence varying between 30 percent and 40 percent in countries such as Germany and the United States.[20]

Globally, 10 to 20 percent of people with diabetes die secondary to complications attributable to renal failure.[29–32] Research has shown that up to 50 percent of diabetics are affected by neuropathy, which is likely an underestimate as it is often missed during the earliest stages of development. Neuropathy is associated with paresthesias, limb ulceration leading to potential limb compromise and eventual amputation, and erectile dysfunction.

Cancer, specifically colorectal, pancreatic, and postmenopausal breast, is more common among diabetics. The excess risk for these cancers in diabetics is 30 percent, 50 percent, and 20 percent, respectively. The mechanism behind this increased cancer rate is unclear, but in part may be related to the high insulin levels seen in diabetics leading to overactivation of the insulin and Insulin-Like Growth Factor-1 (IGF-1) access.[33–36]

Diet and exercise are extremely important in the prevention and treatment of type 2 diabetes as well as in the prevention of complications. Given the association of type 2 diabetes with hypertension and hypercholesterolemia, a diabetic diet warrants attention not only to carbohydrate content and glycemic index but also to cholesterol, salt, and fiber intake. Studies evaluating effects of decreased overall energy, carbohydrate, and fat intake have demonstrated a 1 to 2 percent reduction in HgA1c as early as after three months of dietary intervention.[37–48] Lifestyle and dietary measures such as reduced sodium intake, modest weight loss (as little as 4.5 kg), increased physical activity, a low-fat diet, and moderate alcohol intake have been proven successful in reducing blood pressure.[45] Nutrition consultation is helpful in all forms of diabetes to plan and promote these healthy diet and lifestyle measures.

Dietary regimens drastically differ across continents and cultures. Latino, Chinese, and Indian communities (America and Indian subcontinent), whose diets are often carbohydrate based, are among those hardest hit by the diabetes epidemic. Controlling blood sugar levels may be especially challenging in these populations. Intake of fruits and vegetables, food items linked to decreased risk of cardiovascular disease and obesity, also often varies among cultures.[49] Practitioners must be aware of and sensitive to such cultural differences when making dietary assessments and recommendations.[50–51]

Cost is another huge factor governing diet. In general, healthy diets are expensive, while the most economical foods tend to be fried and are generally the least healthy.[52] Studies suggest that many Americans have shifted their food intake

away from the home toward cheap, fast-food restaurants where they consume foods rich in refined carbohydrates, added sugars, and fats. From 1970 to 1995, the percentage of spending on foods away from home rose from 25 percent to 40 percent, and the proportion of meals and snacks eaten at fast-food restaurants increased by 200 percent.[53] This type of food consumption certainly is not helpful in stemming the diabetes epidemic.

Exercise is essential in controlling type 2 diabetes. Exercise in moderation before and after eating is known to help blunt postprandial hyperglycemia. The long-term benefits of exercise (both aerobic and resistance training) are many, including an increase in lean body mass and insulin sensitivity, lowering of blood pressure, and lowering of cholesterol. The American Heart Association recommends that type 2 diabetic patients should get at least 150 minutes per week of moderate-intensity exercise or 90 minutes per week of vigorous-intensity exercise, or some combination of the two. Patients should exercise on at least three nonconsecutive days each week to maximize benefits with individual sessions at least 10 minutes each or longer. Exercise should include both aerobic and moderate- to high-intensity resistance training whenever possible.[54]

Gaining an upper hand on weight control, reducing obesity in particular, would do much to mitigate the development of this disease. As little as 1 kilogram of weight loss is associated with a 33 percent reduction in type 2 diabetes.[55] Repeated trials have shown that type 2 diabetes can be controlled and maybe even prevented with proper diet and exercise.[7, 10, 56–59]

The Diabetes Prevention Program (DPP) was a multicenter clinical research study of 3,200 overweight prediabetic adults aimed at discovering whether intensive lifestyle intervention that includes modest weight loss through dietary changes and increased physical activity (healthy diet, moderate physical activity of 30 minutes a day five days a week) or treatment with the oral diabetes drug Metformin could prevent or delay the onset of type 2 diabetes compared with control groups who received less intense dietary counseling and no treatment with Metformin.[56] The Finnish Diabetes Prevention Study (DPS) study was smaller, involving roughly 500 participants, and also evaluated the effect of intense lifestyle intervention and modest weight loss on the risk for developing diabetes.[57] Both the DPS and DPP studies showed a remarkable 58 percent risk reduction for development of diabetes in the intense lifestyle group compared with the control group. The DPS also showed a 33 percent reduction in diabetes risk in patients treated with Metformin compared with the control group. These studies support the concept that an increase in physical activity paired with a decrease in weight is an extremely effective modality for preventing and treating type 2 diabetes.

Other medications have been shown to prevent diabetes in high-risk nondiabetic individuals, including acarbose (an inhibitor of intestinal carbohydrate absorption), orlistat (an inhibitor of intestinal fat reabsorption), and thiazolidinediones (insulin sensitizers).[7, 60–62] In the STOP Non-Insulin-Dependent Diabetes Mellitus (STOP-NIDDM) trial, patients with impaired glucose tolerance were assigned to treatment with acarbose versus placebo, with development of diabetes as the primary endpoint. By the study's end, 32 percent of patients randomized to acarbose versus 42 percent randomized to placebo developed diabetes.[60] The study authors concluded that acarbose could be used, either as an alternative or in addition to

changes in lifestyle, to delay development of type 2 diabetes in patients with impaired glucose tolerance.[60]

In the XENical in the Prevention of Diabetes in Obese Subjects (XENDOS) Study, patients classified as obese (BMI greater than 30) or with impaired glucose tolerance were treated with lifestyle intervention plus either orlistat or placebo.[61] Primary endpoints were time to onset of type 2 diabetes and change in body weight. By study close, the cumulative incidence of diabetes was 9 percent with placebo and 6.2 percent with orlistat, corresponding to a risk reduction of 37.3 percent.[61] Exploratory analyses indicated that the preventive effect was explained by the difference in subjects with impaired glucose tolerance. The study authors concluded that compared with lifestyle changes alone, orlistat plus lifestyle changes resulted in a greater reduction in the incidence of type 2 diabetes over four years.[61]

In the Diabetes Reduction Assessment with Ramipril and Rosiglitazone Medication (DREAM) trial, patients with impaired fasting glucose, impaired glucose tolerance, or both, were treated with Rosiglitazone versus placebo with a composite endpoint of either diabetes or death.[62] By study close, 11.6 percent of individuals given rosiglitazone versus 26 percent given placebo developed the composite primary outcome, and 50.5 percent of individuals in the rosiglitazone group versus 30.3 percent in the placebo group became normoglycemic. The authors concluded that treatment with rosiglitazone substantially reduces incident type 2 diabetes and increases the likelihood of normalization of blood glucose levels in individuals at high risk for diabetes.[62]

Early and aggressive treatment of diabetes has been shown to prevent long-term diabetes-related complications. In the United Kingdom Prospective Diabetes Study, patients with newly diagnosed type 2 diabetes were randomized to intensive glucose control with diet and medications (with a goal fasting glucose of less than 108 mg/dl) versus conventional control with diet (with a goal fasting glucose of "the best achievable" with diet alone). Primary endpoints included the development of diabetic complications, diabetes-related death, and all-cause mortality. By study close, the HgA1c of the intensive group was 7 percent versus 7.9 percent in the conventional group. The study showed that intensive glucose therapy in patients with newly diagnosed type 2 diabetes was associated with a reduced risk of clinically evident microvascular complications and a nonsignificant reduction of 16 percent in the relative risk of myocardial infarction.[63] Post-trial monitoring was done to determine whether this improved glucose control persisted and whether such therapy had a long-term effect on macrovascular complications.[64] During the 10 years of post-trial follow-up, the study revealed a "legacy effect": despite an early loss of glycemic differences between groups, a continued reduction in microvascular risk and emergent risk reductions for myocardial infarction and death from any cause was seen.[64] This legacy effect suggests that early and aggressive control of diabetes is vital to prevent complications from developing later on in the disease. The legacy effect was also seen among type 1 diabetics in the Diabetes Control and Complications Trial and the Epidemiology of Diabetes Interventions and Complications studies.[65]

Fingersticks are an excellent way to control all forms of diabetes. While an oral glucose tolerance test is helpful to diagnose diabetes in a relatively controlled fashion, fingersticks have the benefit of showing real-time glucose fluctuations. If

patients with known glucose dysregulation monitor their fasting and postprandial (and possibly preprandial) fingerstick values, they become a more informed and empowered patient and an active partner in controlling their diabetes. Once armed with real-time information, patients can evaluate what foods or activities elevate their glucose values above normal. The goal would be to provide reinforcement of lifestyle behaviors that promote normalization of glucose values and thus help to avoid diabetic complications and progression. A feeling of powerlessness over their disease can lead some patients toward frustration and lassitude, prompting them to simply give up. Empowering patients with the tools that enable them to control of their disease could do much to improve their quality of life.

When lifestyle measures alone fail to control diabetes, medications often are necessary (see table 1.1).[66] Diabetes medications control type 2 diabetes by increasing insulin sensitivity, decreasing glucose output, affecting carbohydrate absorption, and augmenting insulin levels. Various classes of diabetes medications exist, including sulfonylureas, meglintinides, biguanides, thiazoldinediones, incretin mimetic, Dipeptidyl protein IV (DPP-IV) inhibitors, and insulin. As with all medications, side effects such as weight gain and hypoglycemia are associated with the sulfonylureas and meglitinides, and weight gain, hypoglycemia, and edema are associated with the thiazoldinediones (TZDs).[67] Perhaps more alarmingly, two TZDs, rosiglitazone (Avandia) and less so for pioglitazone (Actos), increase the risk for heart failure in some individuals. Specifically, there is recent evidence that shows that rosiglitazone in particular contributes to an increased risk for heart attacks.[67] Of course, insulin (all formulations) is a treatment option for type 2 diabetes and often is used in conjunction with other diabetes medications. Some studies suggest early implementation of insulin may be beneficial in terms of promoting beta-cell preservation.[68] The pharmaceutical industry is actively developing and marketing new medications. Included among these are other versions of incretin mimetics and DPP-IV inhibitors, sodium-glucose transporter (SGLT) inhibitors (which include renal reabsorption and gut absorption of glucose), and inhaled insulins.

Surgical options have attracted tremendous interest among those treating obese diabetics. Bariatric surgery, both for prevention of and treatment of type 2 diabetes, is an exciting and new treatment modality for diabetes. Currently, bariatric surgery is a recommended treatment option for patients with type 2 diabetes who have a BMI greater than 35. Bariatric surgery employs either gastric bypass or gastric reduction measures to decrease food intake, leading to a decrease in weight with a concurrent increase in insulin sensitivity and decrease in type 2 diabetes. Results suggest an average weight loss of 20 to 50 kilograms (44 to 110 pounds).[69–70] Recent studies suggest that greater than 80 percent of cases of diabetes are completely reversed following bariatric surgery.[71] Similar improvements have been seen for patients with hyperlipidemia (83 percent), hypertension (66 percent), and sleep apnea (88 percent).[69–70]

The surgical control of diabetes translates into an improvement in diabetes-related mortality, with studies suggesting a decrease in long-term diabetes-related mortality of 92 percent.[69–70] As bariatric surgical techniques have been improved upon, the mortality associated with this surgery has fallen over the past few years.[71] Regarding complications, the overall reported complication rates for gastric bypass (open and laparoscopic) range from 10 to 20 percent with 95 percent

Table 1.1
Medications to Control Diabetes

Drug Class	Drug Name	Mechanism of Action	Potential Benefit (beyond glycemic control)	Potential Side Effects
Biguanides	Metformin (Glucophage, Glumetza)	Decreased hepatic glucose output, increases insulin sensitivity in muscle	Weight loss, diabetes prevention [65]	Nausea, flatulence
Thiazolidinediones (TZDs)	pioglitazone (Actos), rosiglitazone (Avandia)	Increase insulin sensitivity in fat and muscle	Diabetes prevention [62]	Weight gain, edema, congestive heart failure, myocardial infarction [67]
Sulfonylureas	glyburide (Micronase, Diaeta), glipizide (Glucotrol), glimepiride (Amaryl)	Insulin Secretagogue (long acting)		Weight gain, hypoglycemia
Meglintinides	repaglinide (Prandin), nataglinide (Starlix)	Insulin Secretagogue (short acting)		Weight gain hypoglycemia
Alpha Glucosidase Inhibitors	acarbose (Precose), meglitol (Glyset)	Inhibit intestinal carbohydrate absorption	Diabetes prevention [60]	Flatulence, diarrhea
Incretin Mimetics	Exenatide (byetta)	Decrease hepatic glucose output, decrease gastric emptying, promot glucose-dependant insulin secretion, promote satiety	Weight loss, beta cell preservation [137]	Nausea, pancreatitis [53]

Table 1.1
(Continued)

Drug Class	Drug Name	Mechanism of Action	Potential Benefit (beyond glycemic control)	Potential Side Effects
Dipeptidyl protein IV (DPP-IV) inhibitors	Sitagliptin (Januvia), saxagliptin (Onglyza)	Decrease hepatic glucose output, decrease gastric emptying, promote glucose-dependant insulin secretion, promote satiety	No weight gain	Nasopharyngitis, upper respiratory infection, headache
Rapid-Acting Insulin Analogs	Aspart (Novalog), lispro (Humalog), glulisine (Apidra)		Beta cell preservation [68]	Hypoglycemia, weight gain, cancer [138]
Short-Acting Insulin	Regular		Beta cell preservation [68]	Hypoglycemia, weight gain
Intermediate Acting Insulin	Neutral Protamine Hagedorn (NPH)		Beta cell preservation [68]	Hypoglycemia, weight gain
Long-Acting Insulin Analogues	Glargine (Lantus), detemir (Levemir)		Beta cell preservation [68]	Hypoglycemia, weight gain, cancer [138]

Source: Author's compilation except as noted.

of these complications treated successfully without residual disability.[71] Laparoscopic surgeries generally have fewer complication rates. Complications include an anastomotic leak, wound infection, pulmonary complications, and hemorrhage.[71]

There have been many hypotheses about the mechanism behind the surgical cure of diabetes. Both human and animal studies suggest that weight loss, reduced caloric intake, and malabsorption cannot entirely explain the improvement in glycemic control observed following bariatric surgery.[69–71] Rubino proposes an "anti-incretin theory" as an explanation. Incretin hormones, including both glucose-dependent insulinotropic polypeptide (GIP) and glucagon-like peptide-1 (GLP-1), are produced in response to nutrient ingestion and enhance glucose-dependent insulin secretion, enhance insulin action, and promote beta-cell growth and proliferation. Rubino proposes that anti-incretins normally are produced by the proximal small bowel to balance the action of incretin hormones, resulting in the maintenance of normal glucose homeostasis.[71] Accordingly, a dysfunction in the incretin/anti-incretin system, for example, the overproduction of anti-incretins, would result in decreased insulin secretion, decreased insulin action (insulin resistance), and a depletion in beta-cell mass, contributing to type 2 diabetes. Surgical correction of this dysfunction in the anti-incretin system by duodenal exclusion may explain the resolution of diabetes following bypass surgery.

Type I Diabetes

Type 1 diabetes generally is diagnosed in children and young adults, and arises from the autoimmune-mediated destruction of pancreatic beta cells leading to a state of insulin deficiency.[72] Both genetic and environmental factors have been associated with this autoimmune process leading to type 1 diabetes, although no single factor has been shown consistently to be protective or to confer increased risk. Overall, males and females appear to be equally affected.[73–74] Diagnosis most typically occurs in the pubertal years for both sexes; however, the incidence tends to decrease thereafter for females but remains relatively elevated for males until they reach 35 years of age.

Although less prevalent than type 2 diabetes, type 1 diabetes is equally as serious. In the United States, roughly 3 million people suffer from type 1 diabetes, which accounts for roughly 5 to 10 percent of all cases of diabetes. This translates into 15,000 new cases annually, or roughly 40 new cases daily according to the Juvenile Diabetes Research Foundation (JDRF).[3, 75] In the United States, non-Hispanic white youth had the highest rate of new cases of type 1 diabetes.[76]

The data regarding the global incidence of type 1 diabetes are difficult to obtain primarily because of the lack of a centralized database and limited access to screening tests. Some estimates report 430,000 global cases of type 1 diabetes with 65,000 new cases diagnosed annually.[77] That being said, it appears that (as with type 2 diabetes), the incidence of type 1 diabetes is increasing by 3 to 5 percent annually, particularly in populations historically viewed as having a lower incidence (such as China and Japan).[75–77] Annual incidence varies from 0.61 cases per 100,000 people in China, to 41.4 cases per 100,000 people in Finland.[75]

Globally, extreme variations exist among incidence rates. Significant variations in type 1 diabetes incidence rates are observed between neighboring countries

with differing lifestyles, such as Estonia versus Finland, and between genetically similar populations, such as in Iceland versus Norway. A further example is the difference in incidence between mainland Italy (8.4 cases per 100,000 population) and the island of Sardinia (36.9 cases per 100,000 population). Such variations strongly support the role of environmental factors in the development of type 1 diabetes.[75, 77] Most countries report that incidence rates doubled or more in the last 20 years.[78] The cause for this dramatic rise in incidence is unclear, but may be secondary to environmental triggers, the increased inheritance of genes associated with type 1 diabetes, and increased identification of cases as methods of diagnosis and surveillance improve.

Type 1 diabetics need to be mindful of not only what they eat, but also how heavy they get. Being overweight or obese has an effect on their level of insulin. Type 1 diabetics require daily insulin shots and have to monitor their blood sugar levels carefully. In lieu of simple increases in insulin dose, other treatments may be considered, such as the use of oral hypoglycemic agents like Metformin (off-label) and Symlin.[79] Symlin is an injectable synthetic form of amylin, a molecule normally co-secreted with insulin that decreases glucose levels and blunts postprandial hyperglycemia. Additionally, Symlin may result in a lower insulin dose and weight loss.[80–81] Currently, Symlin is approved for use only as a separate injection with meals.

Evidence supports a genetic component for type 1 diabetes. For example, while the risk of developing type 1 diabetes in the general population is roughly 0.4 percent, the frequency of diabetes developing in children with a diabetic mother is 2 to 3 percent and 5 to 6 percent if the father has type 1 diabetes. The risk rises to almost 30 percent if both parents are diabetic. The inheritance of human leukocyte antigen class II molecules DR3 and DR4 are associated strongly with type 1 diabetes. More than 90 percent of non-Hispanic whites with type 1 diabetes express one or both of these molecules as compared with 50 to 60 percent in the general population.[82–84]

Environmental factors associated with type 1 diabetes include infections, diet, and geography. Risk factors include enteroviral infection in pregnant females, older maternal ages (39 to 42 years), Cesarean section, increased birth weight, preeclampsia, early introduction of cow's milk proteins, and proximity to the equator.[85–89] Protective factors include optimal vitamin D supplementation in early life.[90]

Viral infections have been linked to the development of type 1 diabetes and are thought to initiate or modify an autoimmune process leading to destruction of insulin-producing beta cells located in the pancreas. Possible culprit viruses include the enteroviruses (Coxsackie and echovirus), rubella virus, mumps virus, rotavirus, parvovirus, and cytomegalovirus.[88–89] These viruses have been investigated in experimental and clinical studies aimed at defining their roles in the pathogenesis of type 1 diabetes. Outbreaks of Coxsackie virus, mumps, and congenital rubella have been linked with an increased incidence of type 1 diabetes. One survey suggests enteroviral infection during pregnancy carries an increased risk of type 1 diabetes in offspring.[88–89] Paradoxically, type 1 diabetes incidence is higher in areas where the overall burden of infectious disease is lower, which suggests the interplay of other risk factors.[89]

Various dietary factors have been associated with increased risk for type 1 diabetes. Some studies suggest that breastfed infants have a lower risk for type 1 diabetes and have demonstrated a direct relationship between per capita consumption of cow's milk, which contains proteins that have antigenic similarities to islet cell antigens, and the incidence of diabetes.[87] Additionally, nitrosamines, chemicals found in smoked foods and some water supplies, have been shown to cause type 1 diabetes in animal models, although evidence in humans is lacking.[89]

The association of increasing incidence of type 1 diabetes with distance from the equator may be linked to vitamin D levels. Reduced exposure to ultraviolet light and lower vitamin D levels, both of which are more likely found in the higher latitudes, are associated with an increased risk of type 1 diabetes.[89] Furthermore, taking vitamin D supplements may offer protection against the development of type 1 diabetes. Meta-analysis of data from four studies, which included children from many different European countries, indicated that children taking supplements had a 29 percent reduction in risk of developing type 1 diabetes compared with their peers who were not taking supplements.[90] The exact mechanism by which vitamin D supplementation protects against type 1 diabetes is unclear, but the identification of receptors for the active form of vitamin D in both beta cells and immune cells has led to a suggestion of immune modulation by vitamin D.

Morbidity and mortality data on type 1 diabetes are startling, particularly in the developing world. Studies suggest that, globally, people with type 1 diabetes have a mortality rate three and a half times that of nondiabetics with mortality rates directly correlated to access to health care resources.[91–92] Mortality data varies by geographic location and, globally, type 1 diabetes has a tenfold variation in death: the lowest rates are found in Canada and Western Europe, and the highest rates are found in Japan, Eastern Europe, and Russia. In the developing world, mortality rates are even more striking; survival after diagnosis in many African countries is quite poor.[20, 93] The lack of access to health care providers and facilities as well as the lack of medication contribute significantly to the poor survival among so many in the developing world.

In the United States, overall mortality seems to be improving and evidence indicates that recently diagnosed individuals have improved survival compared with those diagnosed in prior years, independent of age of diagnosis.[20, 91] The authors infer that such improvement in mortality likely stems from better access to health care, better overall glucose control and monitoring, and better management of comorbid conditions.[20, 91] A combination of early and persistent tight control of type 1 diabetes remains the best method for preventing adverse outcomes related to the disease.

Mental health issues have to be addressed in type 1 diabetics. Body image disorders are more common among type 1 diabetics, including eating disorders such as bingeing and "diabulimia."[94–95] Studies suggest that female type 1 diabetics are nearly twice as likely to develop an eating disorder such as bulimia and anorexia nervosa as their nondiabetic peers.[94] Other studies suggest that nearly half of adolescent females with type 1 diabetes skip insulin shots to avoid weight gain. Logically, these individuals have higher rates of diabetes-related complications such as nephropathy and microvascular complications and are at a threefold risk of premature death.[94–95] Depression also is more common in diabetics, with

depression rates twice those compared with individuals without diabetes.[95–96] Depression rates are similar in both type 1 and type 2 diabetes, and they are highest in women. Studies suggest that adults with type 1 diabetes and depression are more likely to be single, have less income, and have poor health care follow-up, and they are twice as likely to seek primary care through emergency room visits.[96]

Much research is being done on finding a cure for type 1 diabetes both in terms of prevention and intervention.[96–101] Studies are under way to look at administration of agents such as oral insulin, vitamin D, glutamic acid decarboxylase (GAD) protein, omega-3 fatty acids, and docosahexaenoic acid to patients at high risk for developing type 1 diabetes, in hopes of preventing disease.[96–97] Research is ongoing regarding immunological manipulation in newly diagnosed type 1 diabetics, using agents such as CTLA-4 Ig (Abatacept) and anti-CD20 (Rituximab), with hopes of prolonging c-peptide production via manipulation of the immune system and arresting the progressive destruction of beta cells.[96–97] Transplantation, including pancreatic, islet cell, and stem cell, holds promise, but issues exist with all modalities, limiting the cohort of patients for whom they may be appropriate. Antirejection medications are often poorly tolerated by patients and also are toxic to transplanted beta cells, precipitating transplant failure. Although islet and stem cell transplants are less invasive and require less toxic immune-suppressive agents, they have not proven to yield long-term benefits as patients invariably experience a progressive decline in c-peptide levels and concurrent increase in insulin requirements.[100] More research is needed before these methods of treatment can be widely recommended.

Gestational Diabetes Mellitus

The American Diabetes Association estimates more than 200,000 cases of GDM are diagnosed annually in the United States.[102–103] By definition, GDM refers to glucose intolerance that is first recognized during pregnancy. In normal, nondiabetic pregnancies, insulin production increases accordingly and keeps blood glucose levels in check. Women with pre-GDM or diagnosed GDM fail to release enough insulin to counter this rise in insulin resistance and hyperglycemia ensues. This hyperglycemia percolates across the placenta and bathes the developing fetus, which is the cause of GDM complications involving the placenta and fetus.

Reaching a diagnosis of DM in pregnancy requires breaching a different and more restrictive set of "normal" glucose values than those used in type 2 diabetes (see table 1.2). Evaluation for gestational diabetes is traditionally done at 24 to 28 weeks gestation or earlier in high-risk patients. Initial testing is usually performed with a one-hour 50-gram oral-glucose challenge test, which if abnormal (more than 140 mg/dl) may be followed by a diagnostic 75-gram or 100-gram oral-glucose challenge test.[7] In making a positive diagnosis, two or more of the venous plasma concentrations must be met or exceeded. The fasting glucose tests should be done in the morning after an overnight fast of between 8 and 14 hours and after at least three days of unrestricted diet and unlimited physical activity.

The incidence of GDM, which has increased over time and probably is related to the increase in overweight or obese women becoming pregnant, ranges between 4 and 14 percent of all pregnancies.[103–105] Studies suggest the relative risk of GDM increases for overweight, obese, and severely obese women by a factor of

Table 1.2
Diagnosis of GDM with a 100g or 75g Glucose Load

	mg/dl	mmol/l
100g glucose load		
Fasting	95	5.3
1-hour	180	10.0
2-hour	155	8.6
3-hour	140	7.8
75g glucose level		
Fasting	95	5.3
1-hour	180	10.0
2-hour	155	8.6

Source: American Diabetes Association. Standards of medical care in diabetes, 2009. *Diabetes Care.* 2009;32(Suppl 1):S13–S61.
Note: GDM = gestational diabetes mellitus.

2.1, 3.6, and 8.6, respectively.[105] Although 40 to 60 percent of women diagnosed with GDM have no identifiable risk factors,[104] factors other than obesity include a previous diagnosis of gestational diabetes or prediabetes (either impaired fasting glucose or impaired glucose tolerance), a family history of type 2 diabetes, increased maternal age (especially over 35 years), a previous pregnancy that resulted in a child with a high birth weight (greater than 90th percentile, or more than 4,000 grams [8 pounds 12.8 ounces]), a previous poor obstetric history, smoking, polycystic ovarian syndrome, and short maternal stature.[102–109]

Specific ethnic groups are at particularly high risk for GDM. Globally, marked variation in GDM prevalence exists among different racial and ethnic groups with a higher prevalence documented among Asian, African American, and Hispanic populations compared to non-Hispanic whites.[110–119] In the United States, Native Americans, Hispanic, Asian, and African American women are at greater risk for GDM than non-Hispanic white women.[111, 117] In Australia, GDM prevalence is higher in women born in China or India than in women whose country of birth was in Europe or Northern Africa,[115] and also is higher in Aboriginal women compared with non-Aboriginal women.[116] In Europe, GDM is more common in women of Asian background than among women of European heritage.[118] Residency location also plays an important role in GDM prevalence: the proportion of pregnancies complicated by GDM in Asian countries has been reported to be lower than the proportion observed in Asian women living in other continents.[117–118] In India, GDM has been found to be more common in women living in urban areas than in women living in rural areas.[119] This correlation between GDM and ethnicity and country of residence strengthens the argument that genetics and environment play a role in the development of this disease.

Following pregnancy, studies show that 5 to 10 percent of women with GDM subsequently develop type 2 diabetes. And data further suggest that women with GDM are at greater risk of a recurrence during their next pregnancy; GDM recurs in 30 to 69 percent of subsequent pregnancies.[120] The risk of development of type 2 diabetes and recurrent GDM is directly correlated to postpartum

weight—that is, woman who fail to lose significant weight gained during pregnancy are at greater risk postpregnancy and during future pregnancies.[120–122]

A linear relationship exists between the magnitude of glucose elevation above both fasting glucose goals (70 to 90 mg/dl) and postprandial glucose goals (less than 120 mg/dl) and GDM complications.[123] Congenital malformations resulting from poorly controlled GDM can affect multiple organ systems—including the brain and spinal cord, the heart and major vessels, the kidneys, the gut, and skeletal structures—and can result in pre- or postnatal complications.[123–125] Such complications include increased Cesarean delivery, neonatal hypoglycemia, birth defects, premature delivery, shoulder dystocia or birth injury, hyperbilirubinemia, preeclampsia, and both micro and macrosomia.[123–125] Even subtle nuances in glycemic control can increase the frequency of these adverse pregnancy outcomes [123–127] with studies indicating that even in planned pregnancies with optimal prepregnancy care, the incidence of malformations in diabetic pregnancies is still at least twice that in nondiabetic pregnancies.[123]

Diet is the cornerstone to management of GDM. All GDM patients should consult with a nutritionist, and counseling should pair dietary intake to activity and blood glucose fingerstick records. Fingerstick monitoring is essential to GDM management, and strict control of both fasting and postprandial glucose values have been shown repeatedly to decrease macrosomia and GDM complications.[126–130] Fingerstick and periodic HgA1c and fructosamine testing may be employed to monitor the level of success in control.

Exercise in moderation is always recommended.[129–130] For those women at high risk for GDM, exercise in moderation during pregnancy might help prevent the disease, and women with GDM who exercise during pregnancy are less likely to require treatment with insulin.[129]

When dietary and lifestyle measures fail to control gestational diabetes, pharmacologic treatment is warranted. Various insulins have been approved or are in the process of approval for use in pregnant women. Insulin requirements will vary from patient to patient and pregnancy to pregnancy, and insulin doses should be individualized accordingly. No oral agents are as of yet approved for use in pregnancy, although numerous studies suggest glyburide and Metformin may be safe for use. Studies are ongoing regarding this issue, and use of any oral hypoglycemic agents during pregnancy is currently off-label.[131–133]

THE INCREDIBLE COST BURDEN OF DIABETES

The costs associated with all forms of diabetes are astronomical. Globally, the cost of diabetes has been estimated to be $232 billion, a number projected to skyrocket to a dizzying $302.5 billion by 2025.[17, 20, 93, 134–136] The burden of this disease is huge both for developed countries as well as developing nations, where the disease has increased substantially in recent years. The United States, with only 8 percent of the world's burden of diabetes, still accounts for more than half of global costs related to the disease. Europe, Japan, and Australia account for one-quarter of the global diabetic costs. But it is in the developing world, where the incidence and prevalence of diabetes have soared, that the economic burden of diabetes is felt acutely. In countries such as India and China, the human

and economic cost of the disease is extreme. For example, in India, the poor allocate more than 25 percent of their annual income for diabetes care (in the United States, this figure approximates 10 percent).

In the United States, more than $174 billion is spent on diabetes care.[18, 30, 66, 134–136] Physician visits alone accounted for $9.9 billion dollars in 2007. Costs of prescription medications too are substantial. Between 2001 and 2007, the cost of diabetes medication nearly doubled from $6.7 billion to $12.5 billion perhaps due to the increased use of newer and more costly medications.[23, 135–136] One-third of all Medicare expenditures currently are attributed to diabetes-related costs.[66] According to Medicare data, $11,744 is spent annually per diabetic compared with $2,935 per nondiabetic. In fact, diabetics incur health care costs at 2.3 times that of nondiabetics.[66] In 2007, 22 percent of hospital inpatient days in the United States were for diabetes-related care, costing the U.S. health care system an astounding $58.3 billion.[18, 30, 66] In this era of health care reform, this situation must be addressed as the nation cannot continue along the current path.

DIABETES PREVENTION AND DETECTION

The future in controlling diabetes lies in its prevention, including reduction of known risk factors and early identification of those at elevated risk for diabetes. Blood glucose screening, for example, is an excellent way to identify individuals with diabetes, and intervention at an early stage often can mitigate the ravages of this disease. Lifestyle changes, frequent monitoring, and even medication can do much to better manage the prediabetic and diabetic. Postprandial hyperglycemia, a known important independent risk factor for microvascular and macrovascular diabetic complications, is often the first abnormality seen in the early stages of diabetes.[139] By the time fasting glucose and HgA1c are elevated, in most cases, the disease has progressed. Subtle elevations in HgA1c should raise the alarm that postprandial hyperglycemia may be present.

FUTURE CHALLENGES

Improvements in diabetes care, both by means of availability of medications as well as diabetes testing supplies, are extending the lives of many diabetics. But, even though diabetics are living longer, if control remains suboptimal, diabetes complications and concurrent costs inevitably will rise dramatically. Much more needs to be done to stem the epidemic domestically as well as internationally. For a chance of success, prevention efforts must include a partnership between the individual and the health care provider. Changes in diet and lifestyle are essential in the management of this disease. Cooperation of the food industry and government toward efforts in reducing the fat, salt, and sugar content of processed foods; decreasing portion sizes; promoting affordable and nutritious food choices; and supporting healthy food and lifestyle marketing could greatly truncate the diabetes and obesity epidemics. Promotion of affordable self-management counseling and education materials in nutrition and physical activity is needed. Recognition of the impact of such measures on both physical and economic global health should sway governments to allocate the resources required for implementation of such measures.

DM, be it type 2, type 1, or gestational, is not a "new" disease by any stretch of the imagination. The disease knows no economic, social, or geographic boundaries; both rich and poor, all racial and ethnic groups, and developed and developing countries are dealing with this crisis. Obesity is the catalyst for the explosion in diabetes and complicates and compounds the disease's significant health and economic consequences. Taking a pill or a shot each day should not be the way to deal with this insidious disease. While pharmaceutical management is necessary and appropriate, what is needed is a strategy to address the issues relating to lifestyle changes. Better education about the disease and its causes, more screening programs, and more public health awareness messages are needed. The key to control and management of diabetes lies in its prevention, an effort that requires the concerted efforts and resources of all individuals and all nations.

REFERENCES

1. Aretaeus. *De Diabete*. Paris; 1554.
2. Willis T. *Practice of the Physick*. London; 1684.
3. Ahmed AM. History of Diabetes Mellitus. *Saudi Med J*. 2002;23(4):373–378.
4. Polonsky KS, Given BD, Van Cauter E. Twenty-four-hour profiles and pulsatile patterns of insulin secretion in normal and obese subjects. *J Clin Invest*. 1988;81(2):442–448.
5. American Diabetes Association. Postprandial blood glucose: consensus statement. *Diabetes Care*. 2001;24(4):775–777.
6. Simon D, Senan C, Garnier P, Saint-Paul M, Papoz L. Epidemiological features of glycated haemoglobin A1c-distribution in a healthy population. *Diabetologia*. 1989; 32(12):864–869.
7. American Diabetes Association. Standards of medical care in diabetes 2009. *Diabetes Care*. 2009;32(Suppl 1):S13–S61.
8. HealthInsite. http://www.healthinsite.gov.au.
9. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes estimate for year 2000 and projections for 2030. *Diabetes Care*. 2004;27(5): 1047–1053.
10. Harris MI, Flegal KM, Cowie CC, et al. Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in US adults: the third National Health and Nutrition Examination Survey (NHANES III) 1988–1994. *Diabetes Care*. 1998;21(4):518–524.
11. Centers for Disease Control and Prevention, National Center for Health Statistics. NHANES III 1988–1994. http://www.cdc.gov/nchs/nhanes.htm.
12. International Diabetes Federation. Diabetes atlas. http://www.eatlas.idf.org.
13. Roglic G, Unwin N, Bennett PH, et al. The burden of mortality attributable to diabetes: realistic estimates for the year 2000. *Diabetes Care*. 2005;28(9): 2130–2135.
14. Cowie CC, Rust KF, Byrd-Holt DD, et al. Prevalence of diabetes and impaired fasting glucose in adults in the US population: NHANES 1999–2002. *Diabetes Care*. 2006;29(6):1263–1268.
15. Centers for Disease Control and Prevention, National Center for Health Statistics. NHANES 2003–2006. http://www.cdc.gov/nchs/nhanes.htm.

16. Centers for Disease Control and Prevention, National Center for Health Statistics. 2004–2006 NHIS. http://www.cdc.gov/nchs/nhis.htm.
17. US Census Bureau. Resident population estimates for September 1, 2007. http://www.census.gov/popest.
18. Centers for Disease Control and Prevention. National diabetes fact sheet: United States, 2005. http://www.ndep.nih.gov/diabetes/.
19. Centers for Disease Control and Prevention. Age-specific percentage of civilian, noninstitutionalized population with diagnosed diabetes, by race and sex: United States, 2006.
20. World Health Organization. http://www.who.int/diabetes.
21. Centers for Disease Control and Prevention, National Center for Health Statistics. NHANES data on the prevalence of overweight among children and adolescents: United States, 2003–2006. Health e-stat.
22. Ogden CL, Carroll MD, Flegal KM. High body mass index for age among us children and adolescents, 2003–2006. *JAMA*. 2008;299:2401–2405.
23. Centers for Disease Control and Prevention. http://www.cdc.gov/obesity.
24. Joffe B, Zimmet P. The thrifty genotype in type 2 diabetes: an unfinished symphony moving to its finale. *Endocrine*. 1998;9(2):139–141.
25. Neel JV. Diabetes mellitus: a thrifty genotype rendered detrimental by progress? *Am J Hum Genet*. 1962;14:353.
26. Coughlin SS, Calle EE, Teras LR, Petrelli J, Thun MJ. Diabetes mellitus as a predictor of cancer mortality in a large cohort of US adults. *Am J Epidemiol*. 2004;159:1160–1167.
27. Morrish NJ, Wang SL, Stevens LK, Fuller JH, Keen H. Mortality and causes of death in the WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia*. 2001;44(Suppl 2):S14–S21.
28. Donnelly CW, et al. Vascular complications of diabetes. *Br Med J*. 2000;320:1062–1066.
29. Haffner SM, Lehto S, Ronnemaa T, Pyorala K, Laakso M. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *N Engl J Med*. 1998;339:229–234.
30. International Diabetes Federation. *Diabetes and Cardiovascular Disease: Time to Act*. Brussels, Belgium: International Diabetes Federation; 2001. http://www.idf.org/webdata/ docs/Diabetes%20and%20CVD.pdf.
31. KRC Research for Juvenile Diabetes Research Foundation. *Type 1 Diabetes: 2004*. Juvenile Diabetes Research Foundation; 2005.
32. Gu K, Cowie CC, Harris MI. Diabetes and decline in heart disease mortality in US adults. *JAMA*. 1999;281:1291–1297.
33. Larsson SC, Orsini N, Wolk A. Diabetes mellitus and risk of colorectal cancer: a meta-analysis. *J Natl Cancer Inst*. 2005;97:1679–1687.
34. Huxley R, Ansary-Moghaddam A, Berrington de Gonzalez A, Barzi F, Woodward M. Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies. *Br J Cancer*. 2005;92:2076–2083.
35. Larsson SC, Mantzoros CS, Wolk A. Diabetes mellitus and risk of breast cancer: a meta-analysis. *Int J Cancer*. 2007;121:856–862.
36. Smith U, Gale EA. Does diabetes therapy influence the risk of cancer? *Diabetologia*. 2009;52(9):1699–1708.
37. Franz MJ, Bantle JP, Beepe CA, et al. Evidence-based nutrition principles and recommendations for the treatment and prevention of diabetes and related complications. *Diabetes Care*. 2002;25:148–198.

38. Canadian Diabetes Association, Clinical Practice Guidelines Expert Committee. Clinical practice guidelines for the prevention and management of diabetes in Canada. *Can J Diabetes*. 2003;27(Suppl 2):S10–S13.
39. Goldhaber-Fiebert JD, Goldhaber-Fiebert SN, Tristan ML, Nathan DM. Randomized controlled community-based nutrition and exercise intervention improves glycemia and cardiovascular risk factors in type 2 diabetic patients in rural Costa Rica. *Diabetes Care*. 2003;26:24–29.
40. Ziemer DC, Berkowitz KJ, Panayioto RM, et al. A simple meal plan emphasizing healthy food choices is as effective as an exchange-based meal plan for urban African Americans with type 2 diabetes. *Diabetes Care*. 2003;26:1719–1724.
41. Lemon CC, Lacey K, Lohse B, Hubacher DO, Klawitter B, Palta M. Outcomes monitoring of health, behavior, and quality of life after nutrition intervention in adults with type 2 diabetes. *J Am Diet Assoc*. 2004;104:1805–1815.
42. Polonsky WH, Earles J, Smith S, et al. Integrating medical management with diabetes self-management training: a randomized control trial of the Diabetes Outpatient Intensive Treatment program. *Diabetes Care*. 2003;26:3048–3053.
43. Banister NA, Jastrow ST, Hodges V, Loop R, Gillham MB. Diabetes self-management training program in a community clinic improves patient outcomes at modest cost. *J Am Diet Assoc*. 2004;104:807–810.
44. Prevention and management of diabetes in Canada. *Can J Diabetes*. 2003; 27(Suppl 2):S24–S26.
45. Whitworth JA, Chalmers J. World Health Organization, International Society of Hypertension—hypertension guidelines. *Clin Exp Hypertens*. 2004;26:747–752.
46. Boule NG, Haddad E, Kenny GP, Wells GA, Sigal RJ. Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: a meta-analysis of controlled clinical trials. *JAMA*. 2001;286:1218–1227.
47. Wei M, Gibbons LW, Kampert JB, Nichaman MZ, Blair SN. Low cardiorespiratory fitness and physical activity as predictors of mortality in men with type 2 diabetes. *Ann Intern Med*. 2000;132:605–611.
48. Church TS, Cheng YJ, Earnest CP, et al. Exercise capacity and body composition as predictors of mortality among men with diabetes. *Diabetes Care*. 2004;27:83–88.
49. Cullen KW, et al. Social-environmental influences on children's fruit, juice, and vegetable consumption reported by parents: reliability and validity of measures. *Public Health Nutr*. 2000;3:345–356.
50. Chyun DA, Amend AM, Newlin K, Langerman S, Melkus GD. Coronary heart disease prevention and lifestyle interventions: cultural influences. *J Cardiovasc Nurs*. 2003;18:302–318.
51. Satia-Abouta J, et al. Dietary acculturation: applications to nutrition research and dietetics. *J Am Diet Assoc*. 2002;102:1105–1118.
52. Drewnoski A, et al. Symposium: Modifying the food environment: energy density, food costs, and portion size. *J Nutr*. 2005;135:900–904.
53. US Food and Drug Administration. http://www.fda.gov.
54. Marwick TH, Hordern MD, Miller T, et al. Exercise training for type 2 diabetes mellitus impact on cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*. 2009;119:3244–3262.
55. Resnick HE, Valsania P, Halter JB, Lin X. Relation of weight gain and weight loss on subsequent diabetes risk in overweight adults. *J Epidemiol Community Health*. 2000; 54(8):596–602.
56. The DPP Research Group. *N Engl J Med*. 2002;346:393–403.

57. Lindström J, Louheranta A, Mannelin M, et al. The Finnish Diabetes Prevention Study: lifestyle intervention and 3-year results on diet and physical activity. *Diabetes Care*. 2003;26(12):3230–3236.
58. LaPorte RE, et al. Prevalence and incidence of insulin-dependent diabetes. In: National Diabetes Data Group, ed. *Diabetes in America*. Bethesda, MD: National Diabetes Data Group, National Institute of Health; 2002:37–46.
59. National Institute of Diabetes and Digestion and Kidney Diseases. *Diabetes Statistics*. Bethesda, MD: National Institutes of Health; 1999. Publication No.99-3892.
60. Chiasson JL, Josse RG, Gomis R, Hanefeld M, Karasik A, Laakso M. Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomized trial. *Lancet*. 2002;359:2072–2077.
61. Torgerson JS, et al. XENical in the Prevention of Diabetes in Obese Subjects study: a randomized study of orlistat as an adjunct of lifestyle changes for the prevention of type 2 diabetes in obese patients. *Diabetes Care*. 2004;27:155–161.
62. Gerstein HC, Yusuf S, Bosch J, et al. Effect of rosiglitazone on the frequency of diabetes in patients with impaired glucose tolerance or impaired fasting glucose: a randomized controlled trial. *Lancet*. 2006;368:1096–1105.
63. UK Prospective Diabetes Study Group. Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes: UKPDS 33. *Lancet*. 1998;352:837–853.
64. Holman RR, Paul SK, Bethel MA, Matthews DR, Neil HA. 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med*. 2008;359(15):1577–1589.
65. Cefalu WT. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes: the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study Research Group. *N Engl J Med*. 2005; 353(25):2643–2653.
66. American Diabetes Association. http://www.diabetes.org.
67. Nissen SE, Wolski K. Rosiglitazone revisted. An updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. *Arch Intern Med*. 2010; 170(14):1191–1201. And Graham, DJ et al. Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with Rosiglitazone or Pioglitazone. *JAMA*. 2010;304(4):411–18.
68. Niswender K, et al. Early and aggressive initiation of insulin therapy for type 2 diabetes: what is the evidence? *Clin Diabetes*. 2009;27(2):60–68.
69. Thaler JP, Cummings DE. Minireview: hormonal and metabolic mechanisms of diabetes remission after gastrointestinal surgery. *Endocrinology*. 2009;150(6):2518–2525.
70. Demaria E. Bariatric surgery for morbid obesity. *N Eng J Med*. 2007;356:2176–2183.
71. Rubino F. Gastrointestinal surgery as treatment for type 2 diabetes. *Curr Opin Endocrinol Diabetes Obes*. 2008;15(2):153–158.
72. Kahn CR, Weir GC. *Joslin's Diabetes Mellitus*. 14th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2005:342–345.
73. Brekke HK, Ludvigsson JF, Van Odijk J, Ludvigsson J. Breastfeeding and introduction of solid foods in Swedish infants: the All Babies in Southeast Sweden study. *Br J Nutr*. 2005;94:377–382.
74. Wahlberg J, Vaarala O, Ludvigsson J. Dietary risk factors for the emergence of type 1 diabetes-related autoantibodies in 21 2-year-old Swedish children. *Br J Nutr*. 2006;95:603–608.

75. Onkamo P, Vaananen S, Karvonen M, Tuomilehto J. Worldwide increase in incidence of type 1 diabetes: the analysis of the data on published incidence trends. *Diabetologia*. 1999;42:1395–1403.
76. Dabelea D, Bell RA, D'Agostino RB, et al. Incidence of diabetes in youth in the United States. *JAMA*. 2007;297(24):2716–2724.
77. Soltesz G, Patterson CC, Dahlquist G. Worldwide childhood type 1 diabetes incidence: what can we learn from epidemiology? *Pediatr Diabetes*. 2007;8(Suppl 6): S14–S36.
78. Amos AF, McCarty DJ, Zimmet P. The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabet Med*. 1997; 14(Suppl 5):S1-S85.
79. Gomez R, et al. Metformin adjunctive therapy with insulin improves glycemic control in patients with type 1 diabetes mellitus: a pilot study. *J Pediatr Endocrinol Metab*. 2002;15(8):1147–1151.
80. Karl D, et al. Pramlintide as an adjunct to insulin in patients with type 2 diabetes in a clinical practice setting reduced A1C, postprandial glucose excursions, and weight. *Diabetes Technol Ther*. 2007;9(2):191–199.
81. Holman RR, et al. Addition of biphasic, prandial, or basal insulin to oral therapy in type 2 diabetes. *N Engl J Med*. 2007;357(17):1716–1730.
82. Felner EI, et al. Genetic interaction among three genomic regions creates distinct contributions to early- and late-onset type 1 diabetes mellitus. *Pediatr Diabetes*. 2005;6(4):213–220.
83. Kyvik KO, et al. Concordance rates of insulin dependent diabetes mellitus: a population based study of young Danish twins. *BMJ*. 1995;311:913–917.
84. Patterson CC, Carson DJ, Hadden DR, Northern Ireland Diabetes Study Group. Epidemiology of childhood IDDM in Northern Ireland 1989–1994: low incidence in areas with highest population density and most household crowding. *Diabetologia*. 1996;39(9):1063–1069.
85. LaPorte RM, et al. Prevalence and incidence of insulin-dependent diabetes. In: National Diabetes Data Group, ed. *Diabetes in America*. 2nd ed. Bethesda, MD: National Diabetes Data Group, National Institute of Health; 1999: 37–46.
86. National Institute of Diabetes and Digestion and Kidney Diseases. *Diabetes Statistics*. Bethesda, MD: National Institutes of Health; 1999. Publication No.99-3892.
87. Scott FW. Cow milk and insulin-dependent diabetes mellitus: is there a relationship? *Am J Clin Nutr*. 1990;51:489–491.
88. Wagenknecht LE, et al. Increased incidence of insulin-dependent diabetes mellitus following an epidemic of Coxsackievirus B5. *Am J Epidemiol*. 1999;133: 1024–1031.
89. Dorman J, et al. Risk factors for insulin-dependent diabetes. In: National Diabetes Data Group, ed. *Diabetes in America*. 2nd ed. Bethesda, MD: National Diabetes Data Group, National Institutes of Health; 1999:165–178.
90. Zipitis CS, Akobeng AK. Vitamin D supplementation in early childhood and risk of type 1 diabetes: a systematic review and meta-analysis. *Arch Dis Child*. 2008;93:512–517.
91. Makinen VP, Forsblom D, Thorn, et al. Metabolic phenotypes, vascular complications, and premature deaths in a population of 4,197 patients with type 1 diabetes. *Diabetes*. 2008;57(9):2480–2487.
92. National Institute of Diabetes and Digestion and Kidney Diseases. *Diabetes Statistics*. Bethesda, MD: National Institutes of Health; 1999. Publication No.99-3892.

93. International Diabetes Federation. *Diabetes Atlas*. 3rd ed. Brussels, Belgium: International Diabetes Federation; 2007.
94. Mannucci E, Rotella F, Ricca V, Moretti S, Placidi GF, Rotella CM. Eating disorders in patients with type 1 diabetes: a meta-analysis. *J Endocrinol Invest*. 2005;28(5):417–419.
95. Rydell AC, Rodin GM, Olmsted MP, Devenyi RG, Daneman D. Disordered eating behavior and microvascular complications in young women with insulin-dependent diabetes mellitus. *N Engl J Med*. 1997;336:1849–1854.
96. Anderson RJ, Freedland KE, Clouse RE, Lustman PJ. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes Care*. 2001; 24(6):1069–1078.
97. National Institute of Diabetes and Digestive and Kidney Diseases. Type 1 diabetes (T1D) research. http://www.t1diabetes.nih.gov.
98. Type 1 Diabetes TrialNet. http://www.diabetestrialnet.org.
99. Couri CE, Oliveira MC, Stracieri AB. C-peptide levels and insulin independence following autologous nonmyeloablative hematopoietic stem cell transplantation in newly diagnosed type 1 diabetes mellitus. *JAMA*. 2009;301(15):1573–1579.
100. The DPP Research Group. *N Engl J Med*. 2002;346:393–403.
101. Serreze DV, DiLorenzo TP. The good turned ugly: immunopathogenic basis for diabetogenic CD8+ T cells in NOD mice. *Immunol Rev*. 2005;204:250–263.
102. American Diabetes Association. Gestational diabetes mellitus: position statement. *Diabetes Care*. 2004;27(Suppl 1):S88–S90.
103. MacNeill S, Dodds L, Hamilton DC, Armson BA, Vandenhof M. Rates and risk factors for recurrence of gestational diabetes. *Diabetes Care*. 2001;24(4):659–662.
104. Dabelea D, et al. Increasing prevalence of gestational diabetes mellitus over time and by birth cohort. *Diabetes Care*. 2005;28:579–584.
105. Berkowitz GS, Lapinski RH, Wein R, Lee D. Race/ethnicity and other risk factors for gestational diabetes. *Am J Epidemiol*. 1992;135:965–973.
106. Chu SY, et al. Maternal obesity and risk of gestational diabetes mellitus. *Diabetes Care*. 2007;30(8):2070–2076.
107. England LJ, et al. Glucose tolerance and risk of gestational diabetes mellitus in nulliparous women who smoke during pregnancy. *Am J Epidemiol*. 2004;160(12): 1205–1213.
108. Toulis KA, et al. Risk of gestational diabetes mellitus in women with polycystic ovary syndrome. *Fertil Steril*. 2008;90(2):367–372.
109. Ma RM, et al. Relationship between leg length and gestational diabetes mellitus in Chinese pregnant women. *Diabetes Care*. 2007;30(11):2960–2961.
110. Green JR, et al. Glucose tolerance in pregnancy: ethnic variation and influence of body habitus. *Am J Obstet Gynecol*. 1990;163:86–92.
111. Dooley SL, Metzger BE, Cho NH: Gestational diabetes mellitus: influence of race on disease prevalence and perinatal outcome in a US population. *Diabetes*. 1991;40(Suppl 2):S25–S29.
112. King H. Epidemiology of glucose intolerance and gestational diabetes in women of childbearing age. *Diabetes Care*. 1998;21(Suppl 2):B9–13.
113. Anna V, Van der Ploeg HP, Cheung NW, Huxley RR, Bauman AE. Sociodemographic correlates of the increasing trend in prevalence of gestational diabetes mellitus in a large population of women between 1995 and 2005. *Diabetes Care*. 2008;31(12):2288–2293.
114. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public health perspective. *Diabetes Care*. 2007;30(Suppl 2):S141–S146.

115. Beischer NA, et al. Incidence and severity of gestational diabetes mellitus according to country of birth in women living in Australia. *Diabetes*. 1991;40(Suppl 2): 35–38.
116. Ishak M, Petocz P. Gestational diabetes among Aboriginal Australians: prevalence, time trend, and comparisons with non-Aboriginal Australians. *Ethn Dis*. 2003;13:55–60.
117. Dornhorst A, et al. High prevalence of gestational diabetes in women from ethnic minority groups. *Diabet Med*. 1992;9:820–825.
118. Yang X, et al. Gestational diabetes mellitus in women of single gravidity in Tianjin City, China. *Diabetes Care*. 2002;25:847–851.
119. Zargar AH, et al. Prevalence of gestational diabetes mellitus in Kashmiri women from the Indian subcontinent. *Diabetes Res Clin Pract*. 2004;66:139–145.
120. MacNeill S, Dodds L, Hamilton DC, Armson BA, Vandenhof M. Rates and risk factors for recurrence of gestational diabetes. *Diabetes Care*. 2001;24(4):659–662.
121. Resnick HE, Valsania P, Halter JB, Lin X. Relation of weight gain and weight loss on subsequent diabetes risk in overweight adults. *J Epidemiol Community Health*. 2000;54(8):596–602.
122. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. *Diabetes Care*. 2002;25:1862–1868.
123. HAPO Study Cooperative Research Group, Metzger BE, et al. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med*. 2008;358:1991–2002.
124. Loeken MR. Challenges in understanding diabetic embryopathy. *Diabetes*. 2008; 57:3187–3188.
125. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. *Diabetes Care*. 2002;25:1862–1868.
126. Jovanovic L. The role of continuous glucose monitoring in gestational diabetes mellitus diabetes technology and therapeutics. 2000;2(Suppl 1):67–71.
127. Gabbe SG, Graves C. Management of diabetes mellitus complicating pregnancy. *Obstet Gynecol*. 2003;102:857–868.
128. American Diabetes Association. Gestational diabetes mellitus: position statement. *Diabetes Care*. 2004;27(Suppl 1):S88–S90.
129. Snapp C. Gestational diabetes mellitus: physical exercise and health outcome. *Biol Res Nurs*. 2008;10(2):145–155.
130. Dye TD, et al. Physical activity, obesity, and diabetes in pregnancy. *Am J Epidemiol*. 1997;146(11):961–965.
131. Langer O. Oral antidiabetic drugs in pregnancy: the other alternative. *Diabetes Spectr*. 2007;20:101–105.
132. Langer O, et al. A comparison of glyburide and insulin in women with gestational diabetes mellitus. *N Engl J Med*. 2000;343:1134–1138.
133. Elliott BD, et al. Comparative placental transport of oral hypoglycemic agents in humans: a model of human placental drug transfer. *Am J Obstet Gynecol*. 1994;171:653–666.
134. Amos AF, McCarty DJ, Zimmet P. The rising global burden of diabetes and its complications: estimates and projections to the year 2010. *Diabet Med*. 1997;14(suppl 5):S1–S85.
135. Fenton JJ, Van Korff M, Lin EH, Ciechanowski P, Young BA. Quality of preventive care for diabetes: effects of visit frequency and competing demands. *Ann Fam Med*. 2006;4:32–39.
136. Finkelstein EA, Trogdon JG, Cohen JW, Dietz W. Annual medical spending attributable to obesity: payer- and service-specific estimates. *Health Aff*. 2009; 28(5):w822–w831.

137. Wajchenberg B. Beta-cell failure in diabetes and preservation by clinical treatment. *Endocrine Rev*. 2007;28:187–218.
138. Hemkens LG, Grouven U, Bender R, et al. Risk of malignancies in patients with diabetes treated with human insulin or insulin analogues: a cohort study. *Diabetologia*. 2009;26:3080–3086.
139. Monnier L, Colette C, Monnier L, Colette C. Contributions of fasting and post-prandial glucose to hemoglobin A1c. *Endocr Pract*. 2006;12(Suppl 1):42–46.

This page intentionally left blank

CHAPTER 2

Global Cardiovascular Disease: An Expanding Burden with Expanding Solutions

Helina Kassahun, MD, and William B. Borden, MD, MBA

Historically, in a country such as India, the vast majority of the population suffered from malnutrition and acute diseases, especially infectious diseases. Chronic diseases, or "diseases of excess," were not prevalent. In the 21st century, fast-food restaurants abound, with obesity and diabetes becoming symbols of prosperity in Indian society.[1] However, it is not only in rapidly developing countries, such as India, where cardiovascular disease and its predisposing conditions are problematic. Globally, chronic conditions have surged, especially cardiovascular diseases. As used herein, cardiovascular disease refers to any abnormal condition characterized by dysfunction of the heart and blood vessels. Cardiovascular disease includes atherosclerosis (especially coronary heart disease, which can lead to heart attacks), cerebrovascular disease (stroke), and hypertension (high blood pressure).

Cardiovascular disease is among the leading causes of morbidity and mortality globally. In 2004, according to the World Health Organization (WHO), 17.1 million people died from cardiovascular disease, which represented 29 percent of global deaths. By 2010, cardiovascular disease is predicted to be the leading cause of death in developing countries.[2] Hypertension, hyperlipidemia, obesity, and diabetes are among the known risk factors for cardiovascular disease and, together, these conditions have been referred to as the "diseases of affluence" or "Western diseases." As life expectancy increases worldwide, more people in the developing world are living long enough to develop these chronic diseases, a process accelerated by the spread of inexpensive, high-fat, high-caloric foods. Much more information is known about cardiovascular disease in the developed world and significant improvements have been achieved, with cardiovascular deaths decreasing by 24.7 percent from 1994 to 2004 in the United States.[3] Because of the lack of resources and research, however, data related to cardiovascular disease in lower- to middle-income countries are scarce. What information is available is mostly extrapolated from data based on countries in the Western Hemisphere. It is known that a disturbing trend is developing with low-income countries and low-middle-income countries experiencing large increases in the

rates of obesity, hypertension, diabetes, hyperlipidemia, and subsequent cardiovascular disease.

Cardiovascular disease stems not only from individual biological characteristics, but also from environmental conditions and behavioral attributes. For example, people living in economically deprived areas have a 65 percent increase in death from cardiovascular disease compared with those in the most affluent areas, suggesting that lower socioeconomic class is one possible predictor of cardiovascular disease.[4] A retrospective cohort study of mortality was carried out on 3,360 patients who attended Glasgow Blood Pressure Unit between 1991 and 2000, followed for a mean period of 8.1 years. The study found that residents of the most deprived areas had a hazard ratio for all-cause mortality of 1.46 (95 percent confidence interval 1.04, 2.04). Adjusted hazard ratio for death from cardiovascular disease in the most deprived areas was 1.65 (95 percent confidence interval 1.04, 2.60) compared with those from the most affluent areas.

This chapter explores the surge in cardiovascular disease and its risk factors worldwide, the characteristics and implications of this growth, and public health initiatives that can stem or even reverse this trend.

EPIDEMIOLOGY

Historically, the health status and disease profile of a society has been linked to its level of economic and social development. Industrialization has led to improvements in the treatment of diseases that previously afflicted people, such as infectious diseases and conditions arising from nutritional deficiencies. Subsequently, noninfectious chronic diseases such as cancer and cardiovascular disease have taken preeminence as the leading causes of death in contemporary society.[5] This phenomenon, termed "the epidemiologic transition," was described first in 1971 by Abdel Omran, an epidemiologist from Johns Hopkins.[6] The decline in infectious disease mortality and the rise in cardiovascular disease may occur at different times chronologically, but the steps of transition are the same globally. Omran's stages of epidemiologic transition can be applied easily to the contemporary international development of cardiovascular disease (see table 2.1).[5]

At any given time, different countries are at varying stages of the epidemiologic transition. Currently, the developing world and middle-income countries are continuing to deal with the burden of infectious and nutritional disease while simultaneously experiencing a rise in cardiovascular disease. With low-cost, high-caloric foods now more readily available in most countries around the world, and with globalization increasing the purchasing power of many individuals, many people are shifting toward a Western diet that is high in saturated fats, trans fats, and processed carbohydrates. Moreover, urbanization transforms the workforce from labor-intensive agricultural work to more sedentary office-based and service-oriented occupations, leading to a decrease in daily physical activity. This shift began in England in the 1920s and later in the 1950s in countries as diverse as Japan and Sri Lanka. Most of the developing world, such as urban China, India, and Latin America are in the third stage, known as the "age of degenerative and manmade disease," wherein cardiovascular disease accounts for 50 percent of all deaths.[7] While in Western Hemisphere countries, this transition in the epidemiology of

Table 2.1
Modified Model of the Stages of Epidemiologic Transition as It Pertains to Cardiovascular Diseases

Stages of Development	Deaths from Cardiovascular Disease, Percent of Total Deaths	Predominant Cardiovascular Diseases and Risk Factors	Regional Examples
I. Age of pestilence and famine	5–10	Rheumatic heart disease, infections, and nutritional cardiomyopathies	Sub-Saharan Africa, rural India, South America
II. Age of receding pandemics	10–35	As above plus hypertensive heart disease and hemorrhagic strokes	China
III. Age of degenerative and manmade diseases	35–65	All forms of strokes, ischemic heart disease at young ages, increasing obesity, and diabetes	Urban India, former socialist economies, aboriginal communities
IV. Age of delayed degenerative diseases	<50	Stroke and ischemic heart disease at old age	Western Europe, North America, Australia, New Zealand
V. Age of health regression and social upheaval	35–55	Reemergence of deaths from rheumatic heart disease, infections, increased alcoholism, and violence; increase in ischemic and hypertensive diseases in the young	Russian Federation

During stages I to IV, life expectancy increases, whereas in stage V life expectancy decreases compared with stages IV and even III.

Source: Adapted from Yusuf S, Reddy S, Ounpuu S, and Anand S. Global burden of cardiovascular diseases. *Circulation*. 2001;104:2746–2753.

cardiovascular disease occurred over centuries, alarmingly, the developing world is now seeing this shift occurring over the course of decades.[8] Whereas many developing countries have made significant improvements in reducing or eliminating infectious diseases, none have totally eradicated their infectious disease burdens. Thus, this accelerated change in disease profile adds to existing public health and economic strain, creating a double challenge for countries dealing with infectious and, now, cardiovascular diseases.

The WHO and the World Bank published the Global Burden of Disease study to chronicle the current state of health worldwide. Not surprisingly, the study demonstrated that ischemic heart disease and stroke were the two leading causes of death worldwide with a decline in infant mortality and an improvement in life expectancy in many of the developing countries. Moreover, while the birth rate decreased worldwide, life expectancy has increased to such a degree that the average age, worldwide, has become older. With more people living to old age, it is not surprising to see a trend toward diseases of the elderly—that is, degenerative diseases such as cardiovascular disease.[9] Yet, despite the growing prevalence and importance of chronic illness and improvements in treating most infectious diseases, infectious diseases still afflict far too many people in the developing world. Malaria, HIV/AIDS, and tuberculosis remain epidemic, consuming most of the research, resources, and public health focus in those areas. Now, as the world recognizes the new epidemics of obesity, hypertension, dyslipidemia, diabetes, and cardiovascular disease, additional resources and a communal commitment are needed to decrease the growth of these "modern" conditions.

ECONOMIC EFFECTS

The rise of cardiovascular disease has had impacts that range beyond the health burden to also create a large economic challenge for developing countries. Deaths from cardiovascular disease in these countries tend to occur one to two decades earlier than in Western Hemisphere countries with half of cardiovascular deaths in the developing world occurring before age 70.[10] In these low- and middle-income countries, the fact that cardiovascular disease occurs in younger people of working age affects the number of people in the workforce and reduces economic growth. Mortality from cardiovascular disease among working-age people in Brazil, India, and South Africa are 1.5 to 2 times higher than in the United States.[11] The regions most affected by these early deaths are those regions that are most in need of a young, vibrant workforce to energize their economies. Countries that already are plagued by poverty will not be able to cope with the economic impact of losing a large fraction of their workforce to cardiovascular disease. Several reports suggest direct associations between the rate of rise in chronic diseases and the rate of decline in economic growth (see table 2.2).[12]

Added to the lost productivity, developing countries must provide long-term and expensive medical care in treating these chronic diseases, requiring resources, both human and financial, that many do not have. Even in the United States, health care spending and lost productivity from cardiovascular disease exceeded $400 billion in 2006, while in a country such as South Africa, 2 to 3 percent of the country's gross national income was devoted to the treatment of cardiovascular disease.[13, 14]

Table 2.2
Effect of Expected Changes in Chronic Disease Mortality Rates on Economic Growth, 2002–2030

Region	Projected Chronic Disease Increase (percent change from 2002 to 2030)	Estimated Effect on Economic Growth Rates in 2030 (percent per year)
World	21.90	−1.02
High income	12.06	−0.56
Eastern Europe and Central Asia	5.73	−0.27
East Asia and Pacific	52.23	−2.43
South Asia	27.89	−1.30
Latin America	48.01	−2.24
Middle East and North Africa	28.28	−1.32
Sub-Saharan Africa	12.13	−0.57

Source: Adapted from Stuckler D, Population causes and consequences of leading chronic diseases: a comparative analysis of prevailing explanations. *The Milbank Quarterly*. 2008;86:273–326.

Thus, the rise in cardiovascular disease will generate a growing financial burden on not only the developed nations, but also developing countries, thus straining already limited resources.

PUBLIC HEALTH INTERVENTIONS

As the cardiovascular disease epidemic adds to existing burdens in developing countries, the challenges are great, but a multitude of opportunities also are available to stem this tide, starting with population level interventions. Fortunately, current knowledge about the causes of cardiovascular disease can direct public health interventions. Although genetic variability exists among different ethnic groups, common environmental risk factors such as obesity, high saturated fats diets, and smoking play major roles in the development of cardiovascular disease. Public health efforts aimed at minor changes in behaviors such as smoking cessation, lifestyle and diet modifications, and exercise can lead to significant reductions in cardiovascular morbidity and mortality. In primary prevention, the strategy is often on a population level, seeking to reduce the burden of disease in the community as a whole, frequently resulting in only smaller benefits to individuals. In secondary prevention, the strategy is to treat high-risk patients, providing larger benefits, but to fewer individuals.[15] In both cases, limited resources unfortunately hinder treatments and accessibility to primary and specialty care providers.

In seeking solutions to this problem, one must first identify the public health intervention challenges, and then, the opportunities. For many years, population control, infant survival, and infectious diseases have been the focus of global health resources. Many countries must deal with a multitude of barriers to the provision of high-quality health care, such as lack of education, health care provider shortages, and limited access to medications. Currently, few investments and fewer resources are directed toward preventing chronic disease compared with the resources directed at infectious diseases. In addition, political strife, migration (forced or otherwise),

rebellions and war, and the inability to achieve cooperation among leaders have been major obstacles to creating sustainable long-term solutions. Public policy efforts in cardiovascular health need to become a government priority to create the appropriate environment for reducing this disease burden.

The development of a health care infrastructure, such as having a trained health care workforce, pharmaceutical distribution networks, and community clinics and hospitals, must be established to allow community leaders and health care workers to administer basic services and education programs. Fortunately, in many countries, an infrastructure exists for the treatment of infectious diseases such as HIV/AIDS and tuberculosis. Developing countries can build on the infrastructure for these infectious diseases by providing additional cardiovascular training for health care workers, utilizing similar drug procurement programs to obtain antihypertensive and lipid-lowering agents, and creating policies that facilitate the screening and treatment of cardiovascular disease at the same clinics and outreach programs that currently focus on infectious diseases.

In 2001, the WHO called for the creation of national commissions on macroeconomics and on health with the health effort focusing mostly on poverty, HIV, and infectious disease in the developing countries.[16] The purpose of the commissions was to help the world's poorest countries allocate appropriate resources, train personnel, and make public health a vital national issue. These same programs could be applied to cardiovascular disease. For example, the Global Fund to Fight AIDS, Tuberculosis, and Malaria is an international financing institution that invests billions of dollars in 140 countries to support large-scale prevention, treatment, and care programs against the three diseases.[17] This type of program can be applied toward setting up funds to provide resources to combat the rapidly emerging cardiovascular epidemic.

Effective and inexpensive interventions to combat cardiovascular disease are needed. First, worldwide, it is important to focus on the promotion and education of healthy lifestyles, such as have been widely applied in Western Hemisphere countries. In much of the developed world, health fairs in schools, workplaces, and the community are common. In contrast, in the developing world, resources are minimally allocated toward health education. Moreover, there is a dearth of resources directed at health care research. The limited research publications consist mostly of case reports and case series, rather than randomized controlled trials, and most of the research continues to focus on infectious diseases rather than on cardiovascular health.[1]

CIGARETTE SMOKING

Cigarette smoking is a major cause of death and an important risk factor for cardiovascular disease. According to the WHO, about one-third of the adult male population smokes and, in the developing world, tobacco consumption is rising by 3.4 percent per year. The WHO estimates that by 2020, tobacco is expected to be the single greatest cause of death and disability worldwide, accounting for about 10 million deaths per year.[18] Globally, China, with its population of more than 1 billion, is the world's largest producer and consumer of tobacco. A prospective cohort study

of 169,871 Chinese patients showed a dose-response association between pack-years smoked and all-cause mortality in both men and women.[19] This study and others have led to further recognition of the need for government-regulated programs for tobacco control and cessation. In 2003, the WHO Framework Convention on Tobacco Control called for countries to adopt such initiatives as tobacco price and tax increases, advertising bans, and warning labels for packages.[20]

In the United States, cigarette manufacturers are mandated to place warning labels on cigarette packets and tobacco advertisements geared toward children are regulated. Increasing numbers of public service announcements focus on the deleterious consequences of tobacco as well as bans on smoking inside most buildings, restaurants, and public areas. More recently, in 2009, Congress passed a bill empowering the Food and Drug Administration to regulate, though not outlaw, tobacco. A New York State smoking ban resulted in an 8 percent decrease in admissions for acute myocardial infarction with a savings in direct health care costs of $56 million in 2004.[21] Internationally, in 2004, Ireland became the first country to enact a nationwide comprehensive smoking ban in all workplaces, restaurants, and bars. These smoking bans are effective. Other countries across Europe and North America have followed suit and have mandated smoking bans in public areas. If these initiatives are applied worldwide, they could do much to reduce the global burden of smoking-related morbidity and mortality. Individually and collectively, these efforts represent both effective and potentially cost-saving ways to promote cardiovascular health.

OBESITY

Obesity is a well-known risk factor for cardiovascular disease, which affects both adults and children. According to the WHO, in 2005, 1.6 billion adults worldwide were overweight and 40 million adults were obese. Based on these numbers, the WHO predicts that by 2020, there will be 5 million deaths attributable to obesity each year.[22] In developing countries, being overweight historically has been viewed as a sign of wealth. With the global marketing of high-calorie, processed meals that are cheap and readily available, evidence shows that it is the underprivileged who tend to be overweight. Although this leads to populations having high rates of obesity and the attendant diseases that obesity creates, this cohort still has significant malnutrition because of the lack of access to nutritionally complete foods. Another factor contributing to the rise in obesity in developing countries is the shift to reduced physical activity in all aspects of daily life. With urbanization, more people have desk jobs, use cars and buses for transportation, and participate in sedentary leisure activities such as watching television and movies. Therefore, the combination of less active lifestyles and the transition to high-fat, high-calorie diets has made obesity a prevailing public health problem internationally. Individual governments and health care leaders have the opportunity to educate the public. For example, Mexico has initiated major efforts to use the media and community-based programs to educate children and parents on healthy lifestyles.[11]

Several tangible public health measures implemented in the United States can be applied to the developing world. Companies are required by law to label their commercial foods with the ingredients as well as the caloric, fat, carbohydrate, and

sodium contents. Certain municipalities have similar requirements for restaurants to label their foods as well and to encourage the offering of healthy menu options. With the rise in childhood obesity, schools have made significant efforts to provide healthier lunches to students. Although not all of the measures are applicable in the developing world where poverty and limited resources are common, the same strategies can be used to implement mass media to educate the public on nutrition, exercise, and risk factors for cardiovascular disease. For example, public service messages can be used to emphasize the rise in cardiovascular disease and associated risk factors. Another option would be to organize health fairs at which the public can be provided with education materials and pamphlets, as well as screening for chronic diseases such as hypertension and diabetes.

PHARMACOLOGIC THERAPY

Nearly half of the improvement in cardiovascular mortality in the United States is attributable to primary prevention of risk factors. The other half is due to secondary prevention after cardiovascular events.[23] For both targets of prevention, medications such as aspirin, beta-blockers, angiotensin-converting enzyme inhibitors, and lipid-lowering medications are necessary; yet, especially worldwide, these medications are not readily available to everyone.[24] Significant numbers of future cardiovascular events could be prevented with programs that provide affordable and accessible medications to those who need them. One proposed solution is the development of a polypill where different classes of medications are combined in one pill. A randomized controlled trial evaluating the efficacy of a polypill randomized 2,053 individuals in 50 centers across India to the five-in-one Polycap pill (thiazide, atenolol, ramipril, simvastatin, and aspirin) or to each of the drugs individually. The results showed that the Polycap was equivalent to the individual components in reducing blood pressure and heart rate, although the degree of low-density lipoprotein cholesterol reduction was less than that with simvastatin alone.[25] The concept of the polypill shows great promise as a convenient and potentially affordable way to bring primary cardiovascular prevention to large populations worldwide.

Governments can work with pharmaceutical companies to manufacture and provide cheap and generic drugs for the developing world. Currently, many medications to prevent and treat cardiovascular disease are not available in the developing world or are available at much higher costs. Some reasons for the greater expense of drugs in these countries include the lack of generic drug availability, and the expensive distribution networks through which intermediaries sequentially mark up prices.[26] Government legislation is needed to ensure drug quality, facilitate the availability and distribution of appropriate generic medications, and create financial systems, such as insurance plans, that will allow individuals to pay for medications.

FUTURE SOLUTIONS

Several next steps toward implementing cardiovascular disease–reduction programs include utilizing the established resources for delivering health care and developing risk assessment tools to target high-risk population groups. An efficient

and central health care infrastructure is an important and sustainable solution for managing the rise in the cardiovascular epidemic. Primary care providers are critical for population-based risk factor modification, cardiovascular disease prevention, and treating chronic diseases. Economic support and resources, such as medical equipment and education pamphlets, should be made available to primary care providers who can reach large numbers of patients. Efforts should be made to pursue health programs that are affordable and accessible.

More resources and funding need to be allocated toward understanding the barriers to health care, such as the specific beliefs and practices of different cultures. Civic and religious leaders are important allies in the public heath efforts to gain the attention and trust of their communities. Because many areas around the world have physician shortages, training of nonmedical individuals can expand substantially the workforce needed for education and screening. One example is directly observed therapy for the treatment of tuberculosis, a program in which trained nonmedical personnel monitor patients as they take every dose of medication. This program has been used worldwide with success in reducing the transmission of tuberculosis.[27] The same concept can be applied toward establishing programs that ensure that patients have the support and reinforcement they need for the treatment of chronic disease that predispose to cardiovascular diseases.

Current recommendations for the treatment and prevention of cardiovascular disease are based on the long-term risk of myocardial infarction or death. Traditional risk factor assessment tools, such as the Framingham Risk Score, are based on a specific population and can overestimate cardiovascular risk in other ethnic groups. Efforts have been made to develop screening tools that more closely fit individual populations. For example, the Systematic Coronary Risk Evaluation Project developed a long-term risk estimation system for clinical practice in European populations.[28] Continued funding and research should focus on calibrating the risk assessment tools to fit different ethnic groups. Moreover, development of simple and inexpensive methods for screening laboratory tests as well as efficient and accurate methods for recording mortality and morbidity are needed.

CONCLUSION

Cardiovascular disease is already the major cause of death both in the developed world and is rapidly becoming a serious chronic problem in the developing world. Social, environmental, and cultural determinants of cardiovascular health, such as obesity, tobacco use, and access to health care, need to be addressed globally to reduce the incidence of cardiovascular disease. The rise in chronic diseases, in addition to the preexisting infectious disease burden, will be difficult to manage in countries that have limited economic resources. The economic problem is made worse by the fact that cardiovascular disease affects the working-age population, leading to a reduction in the available workforce, further stymieing economic growth in the developing countries. Unless public health interventions are implemented, the combination of scarce resources and the rise in cardiovascular diseases will be catastrophic. Improving cardiovascular health care delivery infrastructures, increasing public awareness, and furthering preventive health research will be the keys to controlling the emerging worldwide epidemic in cardiovascular disease. As

evidenced in the developed world, this large public health problem does not have simple solutions, and the fight against cardiovascular disease internationally will require the efforts of both individual countries and the global health community.

REFERENCES

1. Kleinfield N. Modern ways opens India's door to diabetes. *New York Times*. September 12, 2006.
2. World Health Organization. International cardiovascular disease statistics: statistical fact sheet for populations, 2008 update. http://www.who.int/ncd/cvd.
3. Rosamond W, Flegal K, Furie K, et al. Heart disease and stroke statistics, 2008 update: a report from the American Heart Association Statistics Committee and Stroke Statistics. *Circulation*. 2008;117:e25–146.
4. Stewart L, McInnes G, Murray L, et al. Risks of socioeconomic deprivation on mortality in hypertensive patients. *J Hypertens*. 2009;27:730–735.
5. Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases. *Circulation*. 2001;104:2746–2753.
6. Omran A. The epidemiologic transition: a theory of the epidemiology of population change. *Milbank Memorial Fund Q*. 1971;49:509–538.
7. Yusuf S, Hawken S, Ounpuu S. Effect of potentially modifiable risk factors associated with myocardial infarction in fifty-two countries: the INTERHEART Study case-control study. *Lancet*. 2004;364:937–952.
8. Greenberg H, Raymond S, Leeder S. Global health assistance for chronic illness: a look at the practical. *Prog in Cardiovasc Dis*. 2008;51:89–96.
9. Murray C, Lopez A. Regional Patterns of disability-free life expectancy and disability-adjusted life expectancy: Global Burden of Disease Study. *Lancet*. 1997;349:1498–1504.
10. Lopez A and Murray C. The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases. Injuries and Risk Factors in 1990 and Projected to 2020. Boston, MA: Harvard School of Public Health.
11. Fuster V, Voute J, Hunn M, Smith S. Low priority of cardiovascular and chronic diseases on the global health agenda. *Circulation*. 2007;116:1966–1970.
12. Stuckler D. Population causes and consequences of leading chronic diseases: a comparative analysis of prevailing explanations. *The Milbank Q*. 2008;86:273–326.
13. Gaziano T. Cardiovascular disease in the developing world and its cost-effective management. *Circulation*. 2005;112:3547–3553.
14. Mensah G, Brown D. An overview of cardiovascular disease burden in the United States. *Health Aff*. 2007;26:38–48.
15. Rose G. *The Strategy of Preventive Medicine*. New York: Oxford University Press; 1992.
16. World Health Organization. *Report of the Commission on Macroeconomics and Health: Investing in Health for Economic Development*. Geneva, Switzerland: World Health Organization; 2001.
17. The Global Fund to Fight AIDS, Tuberculosis, and Malaria. http://www.theglobalfund.org.
18. World Health Organization. Tobacco free initiative. http://www.who.int/tobacco/.
19. Gu D, Kelly T, Wu X, et al. Mortality attributable to smoking in China. *N Engl J Med*. 2009;360:150–159.
20. World Health Organization. http://www.who.int/mediacentre/factsheets.

21. Juster H. Declines in hospital admissions for acute myocardial infarction in New York state after implementation of a comprehensive smoking ban. *Am J Public Health*. 2007;11:2035–2039.
22. www.who.int/mediacentre/factsheets.
23. Ford E, Ajani U, Croft J, et al. Explaining the decrease in US deaths from coronary disease, 1980–2000. *N Engl J Med*. 2007;356:2388–2398.
24. Mendis S, Fukino K, Cameron A, et al. The availability and affordability of selected essential medicines for chronic diseases in six low and middle-income countries. *Bull World Health Organ*. 2007;85:279–288.
25. Yusuf S, Pais P, Afzal R, et al. Effects of a polypill (Polycap) on risk factors in middle-aged individuals without cardiovascular disease: a phase II, double-blind, randomized trial. *Lancet*. 2009;373(9672):1341–1351.
26. Joshi R, Jan S, Wu Y, MacMahon S. Global inequalities in access to cardiovascular health care. *J Am Coll Cardiol*. 2008;52:1817–1825.
27. Wright J. Direct observation of treatment of tuberculosis: a randomized controlled trial of community health workers versus family members. *Trop Med Int Health*. 2004;9:559–565.
28. Conroy R, Pyörälä K, Fitzgerald A, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J*. 2003;24:987–1003.

This page intentionally left blank

CHAPTER 3

The Obesity Epidemic: Implications for the Future

Abraham Bornstein, MD, and Rubin S. Cooper, MD

INTRODUCTION

The scientific, medical, and lay communities currently are confronted with a serious medical and public health problem related to the unremitting worldwide pandemic of obesity.[1, 2, 3, 4] The ever-increasing prevalence of obesity has been accompanied by a host of inherently associated comorbidities. As a result, obesity is fast becoming the major cause of premature death in the developed as well as the developing world.[2, 4, 5–7] More than 1.1 billion people in the world are estimated to be overweight, of which approximately 320 million are obese. As many as 1.7 billion people may be exposed to weight-related health risk, with greater than 2.5 million deaths per year attributable to obesity, a figure that is expected to double by 2030.[8]

Over the past two decades, the number of overweight and obese adults, adolescents, and children has increased dramatically. Of greater concern are the children and adolescents ***who already have early obesity-related degenerative diseases***, such as hypertension, dyslipidemia, the metabolic syndrome, and type 2 diabetes mellitus, as well as manifestations of early preclinical atherosclerotic cardiovascular disease that has not previously been observed in this age-group.[7, 8] This chapter will present a review of the obesity epidemic in the United States, including the scope of the problem and the implications for the future. While adult obesity has received considerable attention, we also highlight the alarming epidemic of obesity among children and adolescents. Recent estimates indicate that more than 1 million adolescents and young adults in the United States would be considered obese.[9] Particularly alarming is the fact that greater than 80 percent of obese teenagers are likely to remain obese into adulthood.[10] Because of the associated comorbidities and resulting health care expenditures, which are much higher in severely obese individuals, this situation does not bode well for the future health of the population or for the economic burden associated with caring for this population.[11]

Table 3.1
BMI-Based Obesity Definitions for Adults and Children

Risk class	Adults BMI	Children BMI for age
Normal	18.5 to 24.9 kg/m^2	5th to 84.9th percentile
At risk of overweight	–	85th to 95th percentile
Overweight	25 to 29.9 kg/m^2	>95th percentile
Obese, class I (Moderate obesity)	30 to 34.9 kg/m^2	–
Obese, class II (Severe obesity)	35 to 39.9 kg/m^2	–
Obese, class III (Extreme obesity)	40 to 49.5 kg/m^2	–
Super obese	50 to 59.9 kg/m^2	–
Super, super obese	>60 kg/m^2	–

Definitions of normal through class II obesity based on WHO and NIH recommendations have been reviewed.

Source: Defining Overweight and Obesity. Centers for Disease Control and Prevention. www.cdc.gov/obesity/defining.html.

Note: BMI = body mass index; NIH = National Institutes of Health; WHO = World Health Organization.

DEFINING OBESITY

In adults, overweight and obesity are generally defined using body mass index (BMI), a noninvasive screening measure of weight relative to height that has been shown to be closely correlated with total body fat content in both adults and children.[9, 10] According to the National Heart, Lung, and Blood Institute, overweight in adults is defined as a BMI of 25 to 29.9 kg/m^2; obesity as more than 29 kg/m^2; and extreme obesity as more than 39 kg/m^2 (see table 3.1).[11] BMI trends in the adult population over the past 15 years have been extremely disconcerting in that the prevalence of BMI $\geq$30 kg/m^2 has doubled, that of BMI $\geq$40 kg/m^2 has quadrupled, and that of BMI $\geq$50 kg/m^2 has increased fivefold.[2] Unfortunately, parallel trends have been noted in children and adolescents as well.[1, 3, 4, 9, 10]

HEALTH CONSEQUENCES OF OBESITY

Obesity has been linked to several relatively common degenerative diseases such as type 2 diabetes mellitus, hyperlipidemia, hypertension, cardiovascular disease (CVD), early atherosclerosis, osteoarthritis, sleep apnea, and some forms of cancer.[7] All of the major cardiovascular disease risk factors increase progressively as the BMI increases.[12] Extensive data indicate that weight loss could reverse or arrest many of the harmful effects of obesity.[13] Lifestyle intervention studies have shown the effectiveness of weight loss in improving cardiovascular risk factors, including elevated blood pressure, insulin resistance, type 2 diabetes mellitus, lipid disorders, and the metabolic syndrome.[13, 14]

Several longitudinal studies demonstrated the serious consequences related to obesity in adolescence.[15, 16] A study following 200,000 children over a 30-year period showed that those with a BMI above the 95th percentile (mean BMI = 31 kg/m^2) had twice the all-cause mortality risk during the follow-up period. A definitive dose-response relationship between BMI during adolescence and risk of death in adulthood was demonstrated.[13] Other evidence shows an association between

extreme obesity (BMI $\geq$45 kg/m^2) and years of life lost in young adults (ages 20 to 30 years) with the loss in white men and women being 13 years and 8 years, respectively, while the loss in black men and women was greater than 20 years and 5 years, respectively.[16–18] The rising prevalence of obesity and obesity-related comorbidities also has dramatically raised health care costs and reduced quality of life for many.[18–21]

In addition to obesity-related physical complications, obesity has psychosocial sequelae, especially among children and adolescents.[22, 23] Obese adolescents are more often stigmatized, victimized by peers, and likely to establish fewer friendships than their lean counterparts.[24] Compared with normal-weight adolescents, obese adolescents have more psychological symptoms that are internalized, a poorer self-concept, lower expectations of finishing high school or college, and a health-related quality of life essentially indistinguishable from that of children with cancer.[24–27]

The association of obesity with asthma is noteworthy because asthma is the most common chronic disease of childhood. In the late 20th century, the increase in asthma incidence paralleled that of obesity.[28] In addition to the observation that asthmatic children could become overweight because asthma limits their physical activity, prospective studies also supported the hypothesis that overweight children were more likely than their peers to develop asthma. Among 3,792 children and adolescents 7 to 18 years of age who were assessed annually between 1993 and 1998, those who were overweight or obese were nearly twice as likely as their leaner peers to develop asthma.[29, 30]

EPIDEMIOLOGIC TRENDS IN OBESITY

In the United States, obesity prevalence increased from 13 percent to 32 percent between the 1960s and 2004. Some minority and low-socioeconomic status (SES) groups—including non-Hispanic black women and children; Hispanic and Mexican American women and children; black men, women, and children; Native Americans and Pacific Islanders—are disproportionately affected.[31–33] The key findings from an extensive review of the literature found disturbing facts. First, in 2003–2004, 66 percent of U.S. adults were overweight or obese and 16 percent of children and adolescents were overweight with 34 percent at risk of becoming overweight. Second, white children and adolescents had the lowest prevalence of becoming overweight or being overweight compared with their black and Hispanic counterparts; Asians had a lower obesity prevalence when compared with other ethnic groups, but Asians born in the United States were four times more likely to be obese than their foreign-born counterparts. Third, population-based surveys show a higher prevalence of obesity in individuals with lower SES. Fourth, women 20 to 34 years old had the fastest rise in the rate of obesity and overweight. Fifth, 80 percent of black women age 40 years or over are overweight and 50 percent are obese. Sixth, except in the case of black women, people who are less educated had a higher prevalence of obesity than their counterparts. Seventh, geographic differences in the prevalence of obesity has been observed by state as well as by degree of urbanization. For example, interview data (using self-reported height and weight data) from the 2005 Behavioral Risk Factor Surveillance System (BRFSS) survey revealed that

the highest prevalence of obesity was seen in Louisiana, Mississippi, and West Virginia, whereas the lowest prevalence rates were seen in Colorado and Hawaii. Population-based data also show a higher prevalence of obesity reported for rural populations compared with urban and suburban populations. Rural-urban-suburban differences in obesity and health may reflect socioeconomic differences at least in part, with rural areas more often characterized by local poverty and lack of resources.[33–35]

Because childhood obesity is so predictive of adult obesity, the continued increased incidence of overweight and obesity among children and adolescents in both the United States and worldwide puts society at risk for a dramatic increased incidence of clinical cardiovascular disease as these overweight youth reach adulthood. Even during the pediatric years, obese youngsters demonstrated a higher incidence of hypertension and vascular dysfunction as well as autopsy evidence of atherosclerosis compared with their nonobese peers.[36, 37, 38] In the United States, childhood obesity rates have increased threefold over the past 30 years, with 33 percent of children currently overweight or obese as defined by a BMI greater than the 85th percentile.[31–34] The obesity epidemic is not limited only to the United States. Obesity rates are also rising worldwide, with some European and Asian countries also reaching levels seen in North America.[2, 4, 6] The reasons for the current epidemic include numerous causal influences, such as increased television/screen time, decreased physical activity, changes in the living environment, increased variety and consumption of calorie-dense foods and drinks, increased portion sizes, eating away from home, and increased food marketing and advertising that target children.[39]

Healthy People 2010 identified overweight and obesity as 1 of 10 major health indicators and called for a reduction in the proportion of children and adolescents who are overweight or obese, but the United States has not made any significant progress toward meeting the target goal, despite the fact that the number of overweight and obese children has continued to increase unabated.[18] Recent National Health and Nutrition Examination Survey (NHANES) survey data also confirmed no statistically significant changes in obesity prevalence among children and adolescents in the United States between 2003–2004 and 2005–2006.[31–35] Among adolescents 12 to 17 years of age, the percentage of those considered overweight increased from 4.6 to 17.4 percent (a relative increase of 278 percent) between the periods of 1966–1970 and 2003–2004, more than three times the target prevalence rate set at 5 percent for children in Healthy People 2010.[31–35] These rates of obesity raise major concerns. Johns Hopkins researchers predict that if the rate of obesity and overweight were to continue at the current pace, by 2015, 75 percent of adults would be overweight and 41 percent would be obese. Also, by 2015, nearly 24 percent of U.S. children and adolescents would be overweight or obese.[39]

Information on the prevalence and extent of atherosclerotic changes in teens and young adults in the United States comes from several large prospective and cross-sectional databases.[40–42] Researchers are now able to relate risk factors measured during childhood to preclinical vascular changes in young adults. In the Muscatine Study, for example, more than 14,000 school children (ages 8–18 years, predominantly white) from Muscatine, Iowa, underwent biennial examinations between 1971 and 1981 that assessed their height, weight, blood pressure, triceps

skinfold thickness, and total cholesterol and triglyceride levels.[42] Every 10 years, a subset of the original sample underwent repeat testing, which included more advanced noninvasive assessments that were not available at the original recruitment time. In the late 1990s, a subset of 750 of the original participants (with equal proportions of males and females) underwent repeat testing, including measurement of carotid intimal-medial thickness. This assessment revealed that total cholesterol level, measured during childhood, was a significant independent risk factor for carotid thickening during adulthood for both men and women; elevated BMI during childhood was also a significant independent risk factor for women only.[41]

The Bogalusa Heart Study, one of the longest and most detailed prospective studies of children, focused primarily on the early natural history of coronary artery disease and essential hypertension. The study began in 1973 in Bogalusa, Louisiana, and is ongoing. The original population consisted of both white (65 percent) and black (35 percent) school-age children and young adults up to 35 years of age. Observations have confirmed that adult cardiovascular disease, including atherosclerosis, coronary artery disease, and hypertension, all had their origins in childhood, with anatomic vascular changes documented as early as 5 to 8 years of age.[40, 41] A recent examination of 500 young adults from the Bogalusa Heart Study revealed that elevated LDL-cholesterol level and elevated BMI during childhood were found to be independent risk factors for increased carotid thickening in young adulthood.[40, 41] These results emphasize the serious adverse, cumulative effects of childhood-onset obesity that persist into adulthood.

Individuals previously studied as part of the Bogalusa Heart Study, who had subsequently died as a result of an accident or homicide, underwent autopsy. The Bogalusa investigators found that the prevalence and extent of the arterial vascular surface covered with fatty streaks in the coronary arteries increased with age as did the prevalence of fibrous plaques or plaque burden. Fatty streaks were present in 50 percent of individuals during childhood and in 85 percent of young adults. The prevalence of fibrous plaques increased from 8 percent in childhood to 69 percent in young adulthood. Atherosclerotic plaque burden correlated with elevations in serum total and LDL cholesterol, triglycerides, blood pressure, and BMI during childhood, and rose exponentially as the number of risk factors increased.[40, 41]

NUTRITION AND DIETARY TRENDS

Poor diet characterized by dense caloric content and high glycemic indices often has been cited as a major cause of the obesity epidemic. Consumption of these "empty calorie" foods displaces healthier nutrient-dense options, such as fruits, vegetables, whole grains, and calcium-rich foods. Trend data suggest that changes in eating patterns and food consumption are correlated with increases in obesity. In the 21st-century environment, children and adolescents are eating more food away from home, drinking more sugar-sweetened drinks, and snacking more frequently. Convenience has become one of the main criteria for food choices made by Americans, leading more people to consume quick service meals away from home and to buy ready-to-eat, low-cost, quickly accessible meals to prepare at home. Notable trends derived from studies that used U.S. Department of Agriculture's (USDA) Nationwide Food Consumption Survey data and Continuing Survey of Food Intakes

by Individuals data have demonstrated changes in eating patterns among American youth that illustrate the complexity of relating food intake to the increased prevalence of obesity.[43–50] Nevertheless, these trends support the common perception that children are getting more of their food away from home (that is, fast-food places) and that daily total energy intake that children derived from energy-dense high-caloric snacks increased by approximately 121 kilocalories between 1977 and 1996.[43–50] In addition, portion sizes have increased. Other studies have indicated that children are not eating the recommended servings of foods featured in the USDA food pyramid and there have been significant changes in the types of beverages that children are consuming. For instance, only 21 percent of young people eat the recommended five or more servings of fruits and vegetables each day, and when consumed, nearly half of all vegetable servings are fried potatoes.[32, 34, 43–49] Soda consumption increased dramatically in the early to mid-1990s with 32 percent of adolescent girls and 52 percent of adolescent boys consuming three or more eight ounce servings of soda per day.[47, 48] Children as young as seven months old are consuming soda; milk consumption has declined during the same period.[32, 34, 43–50]

PHYSICAL INACTIVITY AND SEDENTARY BEHAVIORS

Physical activity and physical fitness are strong determinants of health outcomes and mortality risk. Unfortunately, the majority of American children and adults are not particularly physically active. A recent study showed that only 42 percent of 6- to 11-year-olds and 8 percent of 12- to 19-year-olds met the surgeon general's recommendation of 30 to 60 minutes of physical activity three to five days per week.[51–53] Research indicates that a decrease in daily energy expenditure without a concomitant decrease in total energy consumption may be the underlying factor for the increase in childhood obesity.[50–52] Physical activity trends data for children are limited, but cross-sectional data indicate that 33 percent of adolescents are not getting the recommended levels of moderate or vigorous activity; 10 percent are completely inactive; and physical activity levels fall as adolescents age. The amount of activity objectively measured by physical activity monitors tends to be significantly lower than what is reported on surveys.[32, 33] To further complicate matters, overweight youth are involved in even less physical activity and are even less physically fit than their leaner counterparts.[54, 55] Physical activity increases physical fitness, decreases adiposity, and enhances skeletal muscle health. Skeletal muscle is the primary site for glucose disposal and fatty acid oxidation and therefore plays a significant role in insulin resistance, lipid profile, and the risk of developing type 2 diabetes and cardiovascular disease.[56]

Over the last several decades, watching television, using the computer, and playing video games (that is, screen time) has occupied a large percentage of children's leisure time and is associated with a negative impact on physical activity levels. Children in the United States are currently spending 25 percent of their waking hours watching television, and those who watch the most hours of television have the highest incidence of obesity. This observation may be true, not only because little energy is expended while viewing television, but also because of the concurrent consumption of high-calorie snacks.[57–60]

THE PHYSICAL ENVIRONMENT

Experts increasingly have looked to the physical environment as a factor in the rapid increase of obesity in the United States.[61] Few studies have examined the direct effects of the physical environment on physical activity in adults, with even less research available in children.[62, 63] The percentage of trips where children walked to school declined from 20 percent in 1977 to 12 percent in 2001.[64, 65] Because children spend a substantial amount of time traveling to and from school, this may be an area in which to incorporate increased physical activity into daily habits of children. Additionally, in-school environments have had a negative impact on health.[66] In urban areas, outdoor recreational space is often scarce, preventing kids from having a protected place to play. Neighborhood crime, unattended dogs, or lack of street lighting may inhibit children from being able to safely walk outdoors. Busy traffic can impede commuters from walking or biking to work as a means of daily exercise. Recent signs of improvement in these trends are evidenced by Toronto's 23 percent increase in bicycle use after the addition of bicycle lanes, and London's increase in footpath use, in the range of 34–101 percent (depending on location), as a result of improved lighting.[67, 68]

In suburban areas, the progression of "sprawl" has prevented residents from walking or biking, thereby contributing to the ever-increasing dependence on vehicle use. Suburban residents often lack adequate resources for physical recreation or even sidewalks in some areas. In the first national study to establish a direct association between the form of the community and the health of the people who live there, analysts from Smart Growth America and the Centers for Disease Control and Prevention (CDC) found that "sprawl" appeared to have direct relationship to increased BMI and obesity.[69–71] Residents of the highest sprawling areas are likely to weigh 6 pounds more, on average, than residents of the most compact areas. Researchers reported that people in high-sprawl counties were likely to weigh more, walk less, and have a higher prevalence of hypertension.[69–71]

GENETICS AND OBESITY

To date, data regarding obesity seem to indicate that body-weight regulation is controlled by a complex interplay of environmental, biological, and genomic factors, the combined impact of which has produced the increased prevalence of obesity around the world. A better understanding of the genetic contribution to both weight gain and the intra-abdominal distribution of fat (central obesity) has resulted in the identification of certain ethnic groups and susceptible families who are especially at high risk.[72–74] The inheritance pattern of obesity is complex as the identification of susceptibility genes and genetic variants for obesity requires various types of methodological approaches. Identification of the underlying genetic causes of obesity likely will provide a basis both for the development of new therapeutic agents and for the personalized prevention of this condition. Thus far, however, advances in identifying the multiple genes associated with the most common form of obesity have been limited. Since October 2005, when the latest update of the Human Obesity Gene Map became available, single mutations in 11 genes were strongly implicated in 176 cases of obesity globally.[75–78] Also, 50 chromosomal locations pertinent to obesity have been mapped, with potential causal genes

identified in most of those regions. This expanding body of investigation may be particularly useful in the development of targets for future antiobesity pharmacological agents as well as in new genetic testing to predict the risk of obesity.

THE "THRIFTY GENOTYPE" HYPOTHESIS

From a teleological point of view, in the distant past, humans lived under different environmental conditions in which periods of food scarcity alternated with overt famine. To thrive under these conditions, humans, as "hunter-gatherers," developed "energy-thrifty genes" critical to their survival as a species.[79] According to the thrifty genotype hypothesis, the same genes that were key to human survival when food sources were scarce, are now being influenced by an environment in which food, for the most part, is plentiful.[79–81] The thrifty genotype may be just part of a wider spectrum of ways in which genes could enhance fat accumulation in a given environment. The different ways include (1) a tendency to overeat (diminished regulation of appetite and satiety), (2) tendency to physical inactivity, (3) decreased ability to use dietary fats as fuel, and (4) enlarged capacity for storing body fat. Not all people who live in industrial nations with an abundance of food become obese and not all obese people are at risk for the same resultant health outcomes. The variability of how different people interact with the same environmental conditions suggests a genetic role in the development of obesity.[77, 78, 82] Major efforts are now focused on assessing the interactions of genes and environment precipitating the obesity epidemic. From a genetic standpoint, translating this research into clinically useful public health recommendations and practice will take continued effort and time.

OBESITY, INSULIN RESISTANCE, AND THE METABOLIC SYNDROME

The concept of the metabolic syndrome initially emerged from the work of pioneers such as Jean Vague, a French physician, who, in 1947, established a correlation between body fat distribution and the risk of diabetes and cardiovascular disease.[83–85] In 1988, Gerald Reaven, an American endocrinologist and professor emeritus in medicine at the Stanford University School of Medicine, coined the term Syndrome X to refer to the clustering of (abdominal) obesity, hypertriglyceridemia, reduced levels of HDL cholesterol, hyperinsulinemia, glucose intolerance, and hypertension. He later explained how insulin resistance and the compensatory mechanism of hyperinsulinemia could lead to diverse metabolic disturbances, such as type 2 diabetes and cardiovascular disease. Reaven estimated that the prevalence of insulin resistance within the sedentary adult population of North America was approximately 25 percent and was closely linked to visceral obesity.[83–85] The syndrome is currently referred to as the metabolic syndrome, with insulin resistance as the critical biochemical abnormality. Other clinical findings related to the metabolic syndrome have included component traits such as visceral obesity (increased waist circumference), microalbuminuria, and a prothrombotic state.[85–88] Although the exact prevalence may vary depending on the definition used, in the United States about 25 percent of adults over age 20 and 40 percent over age 60 exhibit the metabolic syndrome.[85–88]

Although the exact definition of the metabolic syndrome itself remains controversial, childhood overweight is also related to its individual components. *For example, in the Bogalusa Heart Study, overweight children were 12 times more likely to have high levels of fasting insulin than their leaner peers.*[49] Paradoxically, the relative risk seemed greater for whites than for blacks, but the race difference may reflect the fact that blacks appear to have lower insulin sensitivity than whites independent of body fatness.[53] Higher BMI was associated with higher blood pressure and abnormal lipid profiles (including higher triglycerides) in children and adolescents.[49] Girls who were overweight at age nine, were 10 times more likely to have elevated systolic blood pressure, six times more likely to have low HDL cholesterol levels, and two to three times more likely to have elevated diastolic blood pressure, triglyceride levels, and total, as well as LDL cholesterol than girls of normal weight.[52]

Recently, a joint scientific statement, issued by the American Heart Association Atherosclerosis, Hypertension, and Obesity in the Young Committee, the Council on Cardiovascular Disease in the Young, the Council on Cardiovascular Nursing, and Council on Nutrition, Physical Activity, and Metabolism, addressed the progress made as well as the continued challenges in defining the metabolic syndrome in children and adolescents.[53] The statement addressed primarily the progress that has been made in recognizing the components of the metabolic syndrome in children, the interaction between these risk factors, and their importance as predictors of longitudinal risk for atherosclerotic cardiovascular disease and type 2 diabetes mellitus. The need for early detection and preventive measures regarding cardiometabolic risk factors in children and adolescents was addressed, with a strong focus on obesity, inflammation, insulin resistance, dyslipidemia, and hypertension (core elements of morbidity).

Obesity, prediabetes, and type 2 diabetes all have been associated with elevated triglyceride levels, small dense LDL cholesterol (which is highly atherogenic), and low HDL cholesterol, a dyslipidemic lipoprotein profile characteristic of the metabolic syndrome culminating in an accelerated atherosclerotic process.[62] Fructose and glucose, both commonly used caloric sweeteners, are consumed in large and ever-increasing quantities in the United States and other countries where the incidence of obesity occurs in epidemic proportions.[63–65] Research carried out in lean and overweight adult volunteers at The Rockefeller University Hospital Clinical Translational Science Center, for example, demonstrated that chronic ingestion of a weight-maintaining, very low-fat, high-sugar diet caused a large increase in hepatic de novo lipogenesis and blood levels of saturated fat that correlated with an increase in serum triglycerides and decrease in HDL cholesterol.[66–68] In contrast, adipose tissue de novo lipogenesis was minimally responsive to dietary carbohydrates with no evidence for weight gain. More recent preliminary data show that fructose, unlike glucose, acutely stimulated the production of saturated fat and triglycerides because of its unique hepatic metabolism.[66–68] The evidence indicates that dietary sugars, particularly fructose, not only caused or exacerbated dyslipidemia, but also promoted the deposition of fat in the viscera, especially in the liver, causing or exacerbating insulin resistance.

There is a general need to better understand the heterogeneity of increased triglyceride levels in response to dietary carbohydrates as a clue to the precursors of the metabolic syndrome and type 2 diabetes mellitus, as well as new strategies to address the

diagnosis, treatment, and prevention of these diseases.[68, 69] Elevated triglycerides and other components of the metabolic syndrome frequently are seen in overweight children, and they may precede excessive weight gain, but the lipogenic response to dietary carbohydrate has not been fully explored in this age-group.[69, 70, 71]

LEPTIN, ADIPONECTIN, AND GHRELIN

Leptin plays a key role in regulating the balance between energy intake and energy expenditure, including appetite and metabolism. In obesity, especially with excess visceral adiposity, leptin is thought to contribute to insulin resistance and is considered to be an important link between obesity, insulin resistance, and the atherosclerotic process. Recent studies looking at type 2 diabetics demonstrated that hyperleptinemia, independent of insulin resistance, was directly associated with atherosclerosis. The absence of a functional hormone (or its receptor) leads to uncontrolled food intake and obesity.[89, 90]

Adiponectin is an adipocyte secretory protein, the circulating levels of which are decreased in obese and diabetic states. This protein has been demonstrated to play a role in insulin sensitivity in the liver as well as in the body's metabolic processes. As opposed to leptin, adiponectin has been shown to be inversely correlated with obesity, insulin resistance, and atherosclerosis. Adiponectin seems to maintain cardiovascular health, and low levels are a highly sensitive serum marker for the prediction of future cardiovascular events.[91–97]

Ghrelin is the only known orexigenic hormone. It is released from the stomach and upper intestine shortly before individual meals and is rapidly suppressed by food intake.[98, 99] Moreover, ghrelin appears to play a role in long-term body-weight regulation. As with leptin therapy, the most useful application of ghrelin-receptor blockade might be to prevent weight regain that already has been achieved by other means, instead of using ghrelin-receptor blockade to initiate weight loss.

PREVENTION AND TREATMENT OF OVERWEIGHT AND OBESITY

What can be done to modify the sobering outlook related to obesity and its associated comorbidities? The basic therapeutic principles have been well established: diet and exercise. Woo et al. reported the results of dietary modification alone or in combination with exercise training in a cohort of 82 overweight children 9 to 12 years of age.[100] Findings show conclusive evidence that aggressive lifestyle modification by diet and physical exercise training are needed in overweight children to correct endothelial dysfunction, which is a marker of cardiovascular risk. The results are consistent with numerous other studies addressing the impact of hypercholesterolemia and sedentary lifestyle on endothelial function and the value of diet and exercise for improving vascular relaxation.[101]

Rosenbaum et al. point to the difficulty of maintaining a reduced weight and stress the need to prevent obesity.[102] Societal interventions designed to alter behavior in a manner that decreases the likelihood that an individual will become obese (prevention) are more likely to succeed than interventions designed to help sustain weight loss (treatment). On the other hand, pharmacological interventions designed to assist in sustaining a reduced weight are more likely to be successful

than those designed to prevent weight gain. Overall, the difficulties in preventing weight gain or sustaining weight loss must be viewed as physiological and not only as the product of inactivity and overeating. Significant health benefits can be achieved from even small amounts of sustained weight loss, and many of the methods used to lose weight, including exercise and a healthful diet, have independent and complementary health benefits in terms of metabolic profile and cardiovascular risk, even if weight loss is neither achieved nor sustained.

PHARMACOLOGIC THERAPY FOR OBESITY

The first-line therapy for obesity should be food restriction and exercise. Most people, however, find it difficult to lose weight despite the availability of a wide choice of diets and exercise programs, and find it even more difficult to sustain that weight loss. In terms of managing obesity and obesity-related comorbidities, current recommendations include pharmacologic therapy in conjunction with lifestyle modification for obese individuals (BMI > 30 kg/m^2).[103–106] This therapy is recommend particularly for individuals with comorbid conditions such as type 2 diabetes, hypertension, or dyslipidemia when conservative measures such as behavioral therapy, diet, and exercise have not resulted in the weight loss desired.

Drug treatment for obesity has been plagued by various problems. Since the introduction of thyroid hormone to treat obesity in 1893, many drugs that have been developed for the treatment of overweight and obesity unfortunately have led to undesirable side effects or clinical outcomes that resulted in their termination.[103–106] Therefore, in the short term, caution must be used in embracing new drugs for treatment of obesity, until the long-term safety profile has been proven.

That being said, several pharmacological options currently are available for the treatment of overweight and obesity in the United States (see Appendix 3. A). For the most part, these potential antiobesity drugs can be broadly classified into four categories: (1) pharmacologic agents that primarily decrease appetite through central action; (2) pharmacologic agents that primarily increase metabolic rate or affect metabolism through peripheral action; (3) pharmacologic agents that act on the gastrointestinal tract; and (4) pharmacologic agents that not only affect obesity but also the metabolic syndrome.[103–110] A comprehensive review of the efficacy, pharmacology, and side effects of all the currently available antiobesity medications as well as those under clinical investigation is beyond the scope of this chapter. Suffice it to say that all of these medications have a similar maximum weight loss potential, and have been demonstrated to be equally effective for weight loss.

Findings have shown that antiobesity agents, such as orlistat, sibutramine, and rimonabant, typically result in no more than a weight loss of 5 to 10 percent.[103–110] Even this limited amount of weight loss can result in a significant change in cardiometabolic status associated with improvement in cardiovascular risk profile and a reduced incidence of type 2 diabetes. Yet, most studies have demonstrated that maximum benefits achievable with any of these medications are only manifest when taken in addition to a hypocaloric diet.

Because bariatric specialist physicians may have prescribing habits that are different from nonspecialist physicians, a recent survey was conducted to see how physician members of the American Society of Bariatric Physicians treat obesity.[111]

Almost all prescribed medications. Phendimetrazine, metformin, and phentermine plus L-5-hydroxytryptophan (5-HTP) with carbidopa were all used more frequently than either orlistat or sibutramine. The combination of sibutramine and orlistat as well as 5-HTP/carbidopa (a combination not previously reported for the treatment of obesity) was prescribed by 14 percent and 20 percent, respectively. Twenty percent of the obesity specialists used phentermine. A controlled, randomized, clinical trial to evaluate the safety and efficacy of this combination in treating obesity is warranted.

NONPHARMACOLOGICAL APPROACHES TO WEIGHT MANAGEMENT

An extensive discussion of weight loss surgical procedures is beyond the scope of this chapter, but a limited overview is warranted. All contemporary bariatric surgical procedures significantly restrict dietary intake, resulting in a period of negative energy balance, achieved primarily by either hypocaloric intake or malabsorption. This results in a loss of 25 percent to 35 percent of body weight, which is lost primarily as fat mass in both adults and adolescents.[112–114]

Gastric Bypass

For clinically severe obesity, gastric bypass (open laparotomy or laparoscopic surgery) has become a commonly performed operation. The operation not only causes a loss of appetite, but also is an inherent deterrent to the ingestion of carbohydrates. In adult gastric bypass patients, the surgery produces an average loss of 33 percent of initial body weight.[112–114] Adjustable gastric banding has been used increasingly worldwide, but, in the United States, the procedure has been approved for use in adults only since 2001. The adjustable gastric band restricts food entry into the stomach with the food filling a small proximal gastric pouch instead. This procedure, which is reversible, is the least invasive of all the common bariatric operations. Weight loss tends to occur more slowly with adjustable gastric banding than with other bariatric procedures.[112–114] Postoperatively, maximal weight loss occurs in two to three years, as opposed to only 12 to 18 months for gastric bypass. Short-term results suggest that laparoscopic adjustable gastric banding is just as effective as gastric bypass, but the long-term efficacy of adjustable gastric banding remains unproven.

Surgical Intervention in Adolescents

A relevant question that has been raised often is whether extreme obesity during the teenage years justifies considering the use of bariatric surgery in obese adolescents, rather than delaying surgery until adulthood. Although many of the health risks associated with extreme obesity in adolescence will not manifest as clinical disease for years, many adolescents weighing 100 percent or more above their ideal weight manifest obesity-related diseases as teenagers that will predictably worsen over time. Bariatric treatment guidelines for adolescents that take these considerations into account. Furthermore, once an adolescent or teenager has become extremely obese and has failed traditional weight loss options currently available, chances are slim that a healthy weight will be achieved and sustained in the absence of a more aggressive intervention. Finally, although adolescents, still considered

minors, cannot legally consent to any medical or surgical treatment plan on their own, the general consensus among pediatricians is that those adolescents who are counseled appropriately and capable of assent should have a voice in the decision-making process involving their health care, especially with regard to elective surgical procedures.

Commercial Diets and Weight Loss Programs

More than 64 different commercial diet plans are available with each reporting variable degrees of weight loss success.[115, 116] But, with so many different diet plans available, finding an effective weight loss program that not only is right for an individual's short-term goals but also effective over the long term can be a daunting task. Studies have shown that fad diets usually result in the regaining of all the weight lost and, in many cases, more than was originally lost. Studies also have shown the importance of exercise as a critical part of any weight loss program in keeping weight off. One recent study, however, concluded that the diet program that an individual chooses is irrelevant as long as the diet reduces caloric intake, and is low in saturated fat and cholesterol or is heart healthy.[117] The big problem is that most diets are effective in the short term, but they are not as effective for long-term maintenance.[118, 119] They are either too restrictive, making long-term adherence impossible, or do not address the important biobehavioral or psychological issues critical to successful and sustained weight loss. Furthermore, complicating the already murky picture, with the exception of Weight Watchers, the evidence to support the effectiveness of major commercial weight loss programs is limited (see Appendix 3. B).[120–130] USDA researchers and researchers from other institutions have conducted extensive reviews of the scientific literature to evaluate the efficacy of popular diets. Based on their analysis of existing data collected regarding weight-reducing diets, calorie restriction in itself is the key measure for successful weight loss. Unfortunately, evidence for long-term health safety and maintenance of weight loss associated with fad diets remains elusive.

Despite the apparently poor long-term success rate of fad diets, these weight-reducing plans continue to remain popular among Americans. Millions of all Americans have lost or are currently trying to lose weight, spending more than $33 billion on weight loss products and services. For example, more than 50 percent of the 50 best-selling diet books have been published since 1999.

A PUBLIC HEALTH APPROACH TO DECREASING OBESITY

With nearly 119 million American adults, or 65 percent of the population, currently overweight or obese, the burden of the obesity epidemic superimposed on an already fragmented and ailing health care system will result in escalating disease rates and health care costs that already are immense. The overall direct and indirect costs of obesity in America are more than $117 billion per year.[131] Still, obesity rates have continued to increase, essentially doubling since 1980. In 2001, the Surgeon General's Call to Action to Prevent and Decrease Overweight and Obesity was issued to provide the necessary framework for coordinated efforts between public and private organizations working in collaboration to implement the broad range

of changes needed to reduce obesity in the United States.[129] Public health interventions to decrease obesity prevalence must employ a multifaceted and coordinated approach similar to that which reduced tobacco use not only to change individual behavior patterns, but also to effectively address the environmental barriers to physical activity and healthful food choices. Part of the problem is that new farming practices, subsidies, and innovations in processing, packaging, preservation, and refrigeration have resulted in abundant food supplies, easily stored and transported across states, nations, and continents.[130, 131] The food industry spends approximately $26 billion on advertising annually, and the proliferation of restaurants and fast-food chains has made their fare inexpensively and widely available.[130] To raise public consciousness regarding the content of healthful foods, public health interventions have focused on making the caloric and nutritional content of foods more available by labeling them. Other interventions have included encouraging the sale of more healthful foods in fast-food restaurants, providing tax incentives for the production and marketing of healthy foods, and limiting the sale of high-caloric snacks in schools.[130]

The food industry spends more than $11 billion annually on advertising to children and adolescents, often using innovative methods, such as Internet advertising, Internet games, and product placements on popular television shows.[131–137] Studies have suggested that advertising can significantly shape the eating habits of young people and the purchasing patterns of their parents.[133, 134] America's youth are annually exposed to approximately 40,000 food advertisements, the vast majority of which are for candy, sugar-coated cereal, and fast foods.[133, 134] This practice is quite deliberate. Young children are unable to understand the persuasive intent of advertising or to view it critically. Empirical studies, including recent reviews by the American Psychological Association and the Institute of Medicine, have demonstrated that ads actually do influence product preferences and eating behaviors in children, thereby achieving their intended effects.[138] Potential regulatory strategies, therefore, could include restricting food advertising during children's programs, counteradvertising to promote good nutrition and physical activity, limiting the use of cartoon characters, and restricting Web-based games and promotions. Regulating food advertising has constitutional implications and, therefore, has been somewhat contentious. The U.S. Constitution permits and the public supports the regulation of potentially misleading claims and messages directed at children and adolescents, but no overriding consensus on what determines or what constitutes a misleading claim or message.

Trust for America's Health (TFAH), a nonprofit, nonpartisan organization dedicated to protecting the health of each community and to working to make disease prevention a national priority, recently released a report entitled, "F as in Fat: How Obesity Policies Are Failing in America," which, after a comprehensive review of government preparedness and government efforts, found that national and state policies were inadequate to gain control of the obesity epidemic, falling far short of established obesity reduction goals. The organization concluded that the United States did not yet have a sufficiently aggressive or coordinated national and state strategies in place to adequately address the crisis that threatens to contribute to the obesity epidemic. To date, 41 states and Washington, D.C., had obesity levels greater than 20 percent. Alabama was ranked as the heaviest state with a 28.4 percent obesity

prevalence rate, and Colorado ranked as the least heavy at 16 percent.[131] All states were projected to fall short of the national goal of reducing the proportion of adults who are obese to 15 percent or lower by the year 2010. In 40 states and Washington, D.C., more than 6 percent of adults have diabetes mellitus, far above the projected national goal of 2.5 percent by the year 2010, with Mississippi having the highest rate at 11 percent and Colorado the lowest at 4.7 percent.[131] In terms of a state-by-state evaluation of program successes and failures, Alabama had the worst results, whereas Colorado had the best, but overall no state had made the grade.

The study found that the federal government was facing significant organizational problems, including a lack of designated leadership, a bureaucratic jumble of involved agencies, and the competing interests of industry and public health. The study found that school food and physical activity programs and policies were in disarray and urgently required attention and support. Only four states, California, Hawaii, Texas, and West Virginia, have set appropriate nutritional standards for foods sold in schools (referred to as competitive foods) that were not part of the federally sponsored school lunch program. Eighteen states have limited the availability of competitive foods beyond federal requirements, including Colorado.[131] Only two states, Oklahoma and South Dakota, lacked any laws or rules regarding some form of physical education in elementary and secondary schools, but those established in almost all states often were not enforced, and many of the programs were woefully inadequate. The study also found that state policies and actions aimed at lowering obesity rates were fragmented and inadequate. For instance, although the effectiveness of "snack and soda taxes" was unknown, and may have had even negative consequences, 17 states and the Washington, D.C., had enacted various forms of these taxes to discourage consumption of calorie-dense foods that were low in nutrients.[131] Despite the fact that low-income groups and communities had the highest levels of overweight and obesity, few states and communities tried to improve access to low-cost, nutritious food in those areas or had initiatives to foster increased physical activity, such as increasing sidewalks, cycling paths, and park development.

To combat the obesity crisis, TFAH recommended some crucial government actions, including that startup funds to combat obesity need to be increased to save lives as well as taxpayer dollars. To establish order, the CDC should be designated as the command center to manage the obesity epidemic. Initiatives that the CDC should support include forming an interagency task force, with external experts serving in an advisory capacity. The CDC should investigate the root causes for unhealthy eating, physical inactivity, and obesity, as well as the impact of marketing and advertising on children's diet and health. The CDC should establish nutritional guidelines while centralizing and monitoring the effectiveness of obesity-related public education campaigns. The CDC should become the central repository for all research on obesity so that successful community programs and efficacious treatment strategies can be immediately implemented or at least, fast-tracked.[131]

PUBLIC OPINION AND POLICY REGARDING OBESITY

The role of U.S. agricultural, economic, and nutritional public policies, with respect to the development and distribution of the American food supply and the impact that these policies may be making, should be considered.[139–142] These

policies have resulted in the creation of a commercial environment affecting social, economic, technological, and political factors that favor the development of obesity in America.[139]

As an example, in response to recent lawsuits against fast-food companies, Congress and many states are considering legislation that *protects* food manufacturers, distributors, and sellers from "frivolous" lawsuits. In fact, 11 states have already passed legislation to limit obesity-related lawsuits.[139]

As with many other types of policies, the political and meaningful success of those regarding obesity, depend largely on *public opinion.* To examine the strength of public opinion regarding support for active government involvement in the obesity problem, a nationwide telephone survey was conducted in 2001.[139] The survey, entitled American Attitudes Towards Obesity (AATO), measured public knowledge and attitudes about obesity, negative stereotypes of obesity, discrimination against the obese, and support for public policies targeting obesity.[140] Among the health problems mentioned, only depression was identified by a smaller percentage of the sample as being a serious problem than was obesity. Even though obesity was a source of far more deaths than AIDS, most Americans considered it a less serious health problem.[139] Furthermore, few participants in the survey saw their own weight as a serious health matter. For the majority of respondents, obesity was not viewed as a significant health concern in either a larger or a personal context.[139]

Also in 2001, a survey conducted by Robert Blendon et al. reported that although obesity was listed by 9 percent of Americans as an important health concern, it did not appear among any of the top issues that merited government action.[139, 140] A seeming paradox exists between the lack of concern about the personal or national importance of obesity and attentiveness to information about personal diet, which needs to be reconciled. For example, although most Americans did not see obesity as major health concern for either the nation or themselves, they actively sought information about their health in food labels and news sources.[140]

An important clue comes from beliefs held by people about the cause of individuals being overweight and obese.[140] Respondents in the AATO survey, for example, were far more likely to attribute the growth in obesity to the lack of individual responsibility.[140] The most popular explanation for obesity was individuals' lack of willpower with regard to diet or exercise, a belief held by 65 percent of respondents.[140] Respondents were willing to blame environmental factors: 62 percent agreed that too many unhealthy food options in restaurants and supermarkets were to blame and 57 percent agreed that currently available diets were ineffective. Only 40 percent of respondents thought obesity was inherited from parents, while 18 percent agreed that people were obese because they were "simply born that way."[140] Similarly, 45 percent agreed that people were obese simply from not caring about what they ate and accepting their weight.[140] Given the relative lack of concern about obesity, it is not surprising that support for obesity-targeted policies has fallen behind other major public health initiatives.

ECONOMIC COSTS OF OBESITY

Obesity could shorten the average life span of an entire generation by two to five years, which, if true, would result in the first reversal in life expectancy since

data collection began in 1900. The essence of the problem can be framed in this way: because personal choice regarding diet, exercise, and lifestyle is often a root cause of obesity, public health agencies face difficult challenges in trying to prevent overweight and obesity in the general population. Because personal choice is an individual's right, obesity is more often considered a matter of individual and not governmental responsibility. Although obesity affects the individual primarily, it also has significant global socioeconomic costs. The aggregate result of individual choice is countless preventable disabilities and deaths, affecting families and community.[143] Obesity-attributable medical expenditures reached more than $78 billion in the United States in 2003, with substantial additional indirect costs in lost productivity.[144, 145]

Although critics of state and federal regulation have argued that individuals should absorb the cost of their own illness, taxpayers still finance approximately 50 percent of all medical costs through Medicare and Medicaid, with employers covering most of the rest. These figures suggest that the government has a legitimate stake in controlling medical and social expenditures related to individuals' unhealthy behaviors because they are borne by society at large. Moreover, nonwhite and poor individuals are disproportionately affected by the health burdens related to obesity with poor diet and sedentary lifestyles contributing to these socioeconomic disparities.[142, 146]

IMPLICATIONS OF FOOD PROHIBITIONS TO BATTLE OBESITY

Outright bans on foods or ingredients considered particularly harmful have been considered by many government organizations, and occasionally implemented. For example, a growing body of scientific evidence demonstrates the strong link between the consumption of trans-fatty acids and the subsequent development of coronary heart disease. The Institute of Medicine concluded that trans fat provided *no* demonstrable benefit to human health and *no* level of consumption was deemed safe.[147] New York City, for example, recently restricted the service of products containing trans fat in all food service establishments.[148] Supporters of an outright ban argued that eliminating a known health risk from the food supply would decrease morbidity and premature mortality in the population. Because a trans-fat ban could drive the market back toward saturated fats, a key question was whether evidence was sufficiently robust that trans fat was more dangerous than saturated fat to justify new public policies.[135–137] Conversely, the food industry has asserted that government prohibitions undermine competitive markets and free trade and also has argued that removing trans fat from foods was a significant added expense and would negatively affect the food's taste and desirability.

In the recent past, lawsuits have been filed against food companies as obese individuals have sought compensation for their obesity-related health problems. In particular, a highly publicized lawsuit filed against the fast-food chain McDonald's was brought on behalf of obese children in New York City.[149] This suit alleged that McDonald's negligently failed to warn consumers of the risks of eating Big Macs and french fries, and that the company's marketing constituted deceptive business practices under the state's consumer-protection laws. Similar claims have been threatened against soft-drink companies. Whatever one may think of

such legal action, obesity-related lawsuits are difficult to mount. Plaintiffs must prove that the food or corporate practice actually caused injury and that the dangers were not obvious to the average consumer. Perhaps in reaction, 21 states enacted personal responsibility laws that hold fast-food companies harmless from obesity-related tort claims. Yet, by attracting attention, the lawsuits may have contributed to the prudent decision made by some food companies to provide more healthful products.

Public health always has maintained the assertion that legislation could be used to create conditions that allow people to lead healthier lives and that the government has an obligation to regulate private behavior to promote public health. In the past, the government has intervened in other areas of private behavior, such as the use of alcohol and tobacco, with the development of a scientific body of evidence accompanied by social disapproval, both of which are now in evolution with regard to obesity. By defining obesity as a medical problem (Medicare classified obesity as a disease) and by emphasizing that each employer and taxpayer is a stakeholder, emerging research regarding the overwhelming economic and human cost of obesity has garnered support for the government to play a greater role.

State legislative and local regulatory activity has focused mostly on schools. States and school boards have crafted new policies to reduce students' access to foods that are poor in nutrients; measures include restriction of foods that compete with school lunch programs, closed-campus policies that keep students at school for lunch, and a mandate for increased physical education. Another, more controversial approach is to tax junk foods, typically by excluding them from the general exemption of foods from state sales taxes.[150] As of 2000, 19 states taxed foods that are not nutritious (such as soft drinks and candy). Several other states previously had such taxes, but repealed them in the 1990s because of pressure from the food and beverage industry and because of difficulties enforcing them (for example, some states had difficulty determining which foods met the definition of a taxable item).[151]

ACTIVITY AT THE FEDERAL LEVEL

In 1973, the Food and Drug Administration (FDA) began expanding its regulation of food labels under a voluntary program. In 1994, under the Nutrition Labeling and Education Act (NLEA), the FDA began to require a nutrition facts label on most food products.[152–155] The label is required to provide accurate information about fat, cholesterol, sodium, carbohydrate, sugar, and other contents. Restaurant food had been excluded from this mandate except for restaurants that make health claims about an item. After reviewing the findings of empirical studies, the FDA concluded that only a small percentage of consumers use nutrition labels for controlling weight, nevertheless, label use has been associated with more healthful food choices.[152] Recently, the agency has revisited its labeling regulations to address obesity. In July 2003, the FDA required all food manufacturers to list the trans-fat content on product nutrition labels by 2006. In addition, it convened an Obesity Working Group in 2004, which recommended that the FDA evaluate how the nutrition facts panel might better emphasize caloric information and its importance. The group advised the FDA also to do the following: consider the authorization of health claims on reduced-calorie or low-calorie foods, encourage manufacturers to

make dietary guidance statements and comparative labeling statements to encourage healthful food substitutions, seek the cooperation of restaurants in a voluntary program of standardized nutritional information for restaurant food, and enforce the requirement that serving sizes listed on nutrition labels be accurate. In April 2005, the FDA issued notices of proposed rules regarding information about caloric content and serving size on nutrition labels. Although often unobtrusive, many fast-food restaurants now are required to provide nutritional information on in-store posters and Web sites.[152]

CONCLUSION

Addressing the obesity epidemic sooner rather than later is an important public health and medical issue. This chapter has illustrated the huge scope of the problem and the serious deleterious effect that obesity has on the individual and society. The problem is complex and multifactorial, but individual personal efforts no longer are sufficient; more concerted efforts must be initiated by government agencies to combat the ''fattening'' of Americans. What is most alarming is the huge increase in the number of obese children and adolescents. If this issue is not addressed in the immediate future, the legacy of caring for these individuals in adulthood will haunt society for decades. The huge economic burden (present and future) of caring for obesity-related diseases should be a wake-up call, especially as health care reform is being formulated.

Appendix 3. A
Drug Treatment of Obesity

Drug class	Trial Results and Drug information
Phentermine (Adipex P; Ionamin; Fastin; Mirapront; Obephen; Obermine)—Adrenergic agent	a) Appetite suppressant of the amphetamine and phenethylamine class. b) Approved as an appetite suppressant to help reduce weight in obese patients when used short-term and combined with exercise, diet, and behavioral modification. c) Most typically prescribed for individuals who are at increased medical risk because of weight. d) Works by assisting in the release of certain chemicals in the brain that control appetite.
Orlistat (Xenical)—Lipase inhibitor	a) No long-term obesity-related morbidity and mortality data on orlistat treatment. b) In a four-year double-blind, RCT of 3,305 Swedish obese patients, orlistat reduced weight by 2.7 kgms on average and decreased the incidence of type 2 diabetes from 9.0 percent to 6.2 percent. Among 43 percent of patients who completed the study the beneficial effects were observed in all patients with impaired glucose tolerance at baseline.

(Continued)

Appendix 3. A
(Continued)

Drug class	Trial Results and Drug information
Sibutramine (Meridia in United States; Ectiva in South Africa; Reductil in Europe)—Adrenergic/serotonergic agent: serotonin-norepinephrine reuptake inhibitor related to amphetamines	a) Long-term clinical benefits in major obesity-related morbidity and mortality are not available; ongoing Sibutramine Cardiovascular Outcomes trial is assessing efficacy of sibutramine in reducing MI, stroke, and cardiovascular mortality in 9,000 obese and overweight patients.
Rimonabant (Acomplia, Bethin, Monaslim, Riobant)—Selective CB1 endocannabinoid receptor antagonist	a) Selective antagonist of the cannabinoid type 1 receptor; through actions in the hypothalamus, hindbrain, mesolimbic reward centers and vagus nerve, rimonabant enhances anorexia, potentiates satiation signals and lessens the motivation to consume palatable, rewarding foods; effects lead to reduction in food intake and weight loss. By enhancing glucose uptake in muscle and impeding de novo lipogenesis in liver, it increases insulin sensitivity, reduces steatosis and ameliorates dyslipidemia. b) Four randomized, double-blind trials, comprising the Rimonabant In Obesity (RIO) Program, compared rimonabant 5 or 20 mg daily in conjunction with a low-calorie diet versus placebo in more than 6,600 individuals; RIO-Europe, RIO-Lipids, RIO-North America and RIO-Diabetes have published one-year results; overall data from clinical trials are remarkably consistent with weight reduction of 5 to 10 percent on average. c) Remains to be seen whether safety and efficacy will persist with widespread use; several rimonabant studies examining clinical end-points and surrogate measurements of atherosclerotic burden are currently under way.
Diethylpropion (Tenuate)—Sympathomimetic agent that is also an anorectic agent	a) In a meta-analysis of studies, in combination with lifestyle intervention, the drug was associated with mild weight loss of borderline statistical significance. b) Pharmacologic effect similar to that of amphetamines; some common side effects reported included CNS stimulation, dizziness, headache, insomnia, restlessness, mild increases in BP, palpitations. c) Although no serious adverse events reported in RCTs, rate of serious adverse events could be as high as 15 per 1,000.
Fluoxetine (Prozac)—Antidepressant: selective serotonin reuptake inhibitor (SSRI)	a) Ordinarily prescribed for treatment of depression, OCD (obsessive-compulsive disorder), bulimia. b) Nine studies of fluoxetine treatment reported weight loss outcomes; statistical analysis of study results revealed too much heterogeneity, so pooled analyses could not be rationally presented.

Appendix 3. A
(Continued)

Drug class	Trial Results and Drug information
	c) Seven studies of fluoxetine however, reported weight loss outcomes at six months; 86 percent of studies reported statistically significant weight loss in fluoxetine-treated patients; in contrast to studies reporting six-month outcomes, only 50 percent of these studies reported statistically significant weight loss in fluoxetine-treated patients.
	d) Weight loss relative to placebo ranged from 14.5 kgms lost to 0.4 kgms gained; in adverse event analysis, fluoxetine-treated patients experienced increased nervousness, sweating, tremors, nausea and vomiting, fatigue, somnolence, and insomnia compared with placebo recipients.
Sertraline (Zoloft)—Antidepressant: selective serotonin reuptake inhibitor (SSRI)	a) One study assessed effect on maintaining weight loss in women who had completed a 26-week weight reduction program that combined a very low-calorie diet and behavioral therapy; at the end of 54-week evaluation, sertraline-treated patients had regained an average of 17.7 kgms, while placebo recipients had regained an average of 11.8 kgms, a difference that was not statistically significant.
Bupropion (Wellbutrin)—Antidepressant inhibiting reuptake of dopamine, serotonin, and norepinephrine	a) Five studies assessed the efficacy of bupropion for weight loss. In three of these studies, total weight loss in the bupropion-treated patients was 4.44 kgms; significant heterogeneity among studies; sensitivity analysis assuming no weight loss among patients lost to follow-up yielded a mean weight loss of 2.66 kgms (confidence interval [CI], 1.0 to 4.33 kgms) favoring bupropion; the adverse event analysis showed an increase in dry mouth and nonsignificant increases in diarrhea and constipation.
	b) Research literature on bupropion for depression and smoking cessation is abundant; dry mouth and insomnia are commonly reported side effects in these studies.
Topiramate (Topomax)—Anticonvulsant	a) Nine studies assessed the efficacy for weight loss; all of these studies reported their data only as percentage of weight loss; higher dosage produced significantly more weight loss than the lower dosage over the duration of the study, so only the data on the higher dose was analyzed; individual percentage weight loss values for the six studies using the higher dose reported six-month weight loss outcomes with total percentage of weight lost in the topiramate-treated patients at 8 percent; significant heterogeneity among studies; all six individual studies and pooled results report statistically significant weight loss, but magnitude varied substantially; sensitivity analysis assuming no weight loss among patients lost

(Continued)

Appendix 3. A
(Continued)

Drug class	Trial Results and Drug information
	to follow-up yielded a mean weight loss favoring topiramate of 3.6 percent (CI, 2.6 percent to 4.8 percent).
Zonisamide (Zonegran)—Anticonvulsant	a) Novel antiepileptic drug that has been associated with modest weight loss in patients with epilepsy. When studied in patients with obesity, zonisamide (600 mg daily) was associated with a 5 kgms weight loss compared with placebo. In patients treated for epilepsy, rare adverse effects include Stevens-Johnson syndrome and increases in serum creatinine. b) Associated with dizziness, confusion, and difficulty concentrating.

Source: Dickerson LM, Carek PJ. Pharmacotherapy for the obese patient. *Primary Care: Clinics in Office Practice.* 2009;36(2):407–15.

Appendix 3. B
Type of Diet in Terms of Total Calories and Fat Grams

Type of Diet	Total Calories	Fat grams (percent of calories)	Carbohydrate grams (percent of calories)	Protein grams (percent of calories)	Nutrition Adequacy
Typical American Diet	2200	85 (35 percent)	275 (50 percent)	82.5 (15 percent)	
High-Fat, Low-Carbohydrate Diet • Dr. Atkins Diet • Zone Diet • Sugar Busters • Protein Power	1414	96 (60 percent) fat level: 35–65 percent	35 (10 percent)	105 (30 percent)	Low in several nutrients: vitamins A, B6, D, E, thiamin, folate, calcium, magnesium, iron, zinc, potassium and dietary fiber; this type of diet also contains excess amounts of total fat, saturated fat and dietary cholesterol. Nutritional supplementation is highly recommended.
Moderate Fat Diet	1450	40 (25 percent)	218 (60 percent)	54 (15 percent)	Usually nutritionally balanced eating plan

Appendix 3. B
(Continued)

Type of Diet	Total Calories	Fat grams (percent of calories)	Carbohydrate grams (percent of calories)	Protein grams (percent of calories)	Nutrition Adequacy
• USDA Food Guide Pyramid • DASH Diet • American Diabetic Association • Weight Watchers • Jenny Craig		fat level: 21–34 percent			assuming the dieter eats a variety of foods from all food categories, but limiting certain food categories can lead to deficiencies in certain nutrients especially calcium, iron, and zinc.
Low- and Very Low-Fat Diet • Volumetrics • Dean Ornish's Eat More, Weigh Less • New Pritikin Program	1450	20 (13 percent) fat level: 10–20 percent	235–271 (70 percent)	54–72 (17 percent)	Deficient in zinc and vitamin B12 due to infrequent meat consumption; this type of diet can be inadequate in vitamin E, a nutrient found in oils, nuts, and other foods rich in fat.

Source: Freedman M, King J, and Kennedy E, Popular diets: a scientific review. *J of Obesity Research.* 2001(Suppl 1).
Note: USDA = U.S. Department of Agriculture.

REFERENCES

1. Larson NI, Story M. The pandemic of obesity among children and adolescents: what actions are needed to reverse current trends? *J Adolesc Health*. 2007; 41(6):521–522.
2. Hossain P, Kawar B, Nahas M. Obesity and diabetes in the developing world: a growing challenge. *N Engl J Med.* 2007;356:213–215.
3. Institute of Medicine, Committee on Prevention of Obesity in Children and Youth, Food and Nutrition Board, Board on Health Promotion and Disease Prevention. *Preventing Childhood Obesity: Health in the Balance*. Washington, DC: National Academies Press; 2005.
4. Popkin B, Conde W, Hou N, et al. Is there a lag globally in overweight trends for children compared with adults? *Obesity.* 2006;14:1846–1853.
5. Kumanyika SK, et al. Population-based prevention of obesity. *Circulation.* 2008; 118:428–464.
6. Yach D, Stuckler D, Brownell KD. Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. *Nat Med.* 2006;12(1):62–66.
7. Schelbert KB. Comorbidities of obesity. *Prim Care Clin Office Pract.* 2009; 36:271–285.

8. Deitel M. Overweight and obesity worldwide now estimated to involve 1.7 billion people. *Obes Surg.* 2003;13:329–330.
9. Hedley A, Ogden C, Johnson C, Carroll M, Curtin L, Flegal K. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. *JAMA.* 2004;16(23):2847–2850.
10. Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. Predicting obesity in young adulthood from childhood and parental obesity. *N Engl J Med.* 1997;337(13):869–873.
11. Sullivan PW, Ghushchyan V, Ben-Joseph RH. The effect of obesity and cardiometabolic risk factors on expenditures and productivity in the United States. *Obesity.* 2008;16(9):2155–2162.
12. McLaughlin T, Allison G, Abbasi F, Lamendola C, Reaven G. Prevalence of insulin resistance and associated cardiovascular disease risk factors among normal weight, overweight, and obese individuals. *Metabolism.* 2004;53(4):495–499.
13. Avenell A, Broom J, Brown TJ, et al. Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement. *Health Technol Assess*. 2004;21:iii–iv, 1–182.
14. Vidal J. Updated review on the benefits of weight loss. *Int J Obes Relat Metab Disord.* 2002;26(Suppl 4):S25–S28.
15. Aronne L. Classification of obesity and assessment of obesity-related health risks. *Obes Res.* 2002;10:S105–S115.
16. Must A, Strauss RS. Risks and consequences of childhood and adolescent obesity. *Int J Obes Relat Metab Disord.* 1999;23(Suppl 2):S2–S11.
17. Daniels SR, Arnett DK, Eckel RH, et al. Overweight in children and adolescents: pathophysiology, consequences, prevention, and treatment. *Circulation.* 2005; 111(15):1999–2012.
18. Field AE. Impact of overweight on the risk of developing common chronic diseases during a 10-year period. *Arch Inter Med.* 2001;161:1581–1586.
19. Manson JE, Bassuk SS. Obesity in the United States: a fresh look at its high toll. *JAMA.* 2003; 289:229–230.
20. Popkin BM, Kim S, Rusev ER, Du S, Zizza C. Measuring the full economic costs of diet, physical activity and obesity-related chronic diseases. *Obes Rev.* 2006;7:271–293.
21. Colditz GA. Economic costs of obesity and inactivity. *Med Sci Sports Exerc.* 1999;31(Suppl 11):S663–S667.
22. Erermis S, Cetin N, Tamar M, Bukusoglu N, Akdeniz F, Goksen D. Is obesity a risk factor for psychopathology among adolescents? *Pediatr Int.* 2004;46(3):296–301.
23. Vila G, Zipper E, Dabbas M, et al. Mental disorders in obese children and adolescents. *Psychosom Med.* 2004;66(3):387–394.
24. Latner JD, Stunkard AJ. Getting worse: the stigmatization of obese children. *Obes Res.* 2003;11:452–456.
25. Jackson AS, Stanforth PR, Gagnon J, et al. The effect of sex, age and race on estimating percentage body fat from body mass index: the Heritage Family Study. *Int J Obes Relat Metab Disord.* 2002;26(6):789–796.
26. Janssen I, Katzmarzyk PT, Ross R. Body mass index, waist circumference, and health risk: evidence in support of current National Institutes of Health guidelines. *Arch Intern Med.* 2002;162(18):2074–2079.
27. Janssen I, Heymsfield SB, Allison DB, Kotler DP, Ross R. Body mass index and waist circumference independently contribute to prediction of nonabdominal, abdominal subcutaneous, and visceral fat. *Am J Clin Nutr.* 2002;75(4):683–688.

28. Gold DR, Damokosh AI, Dockery DW, Berkey CS. Body-mass index as a predictor of incident asthma in a prospective cohort of children. *Pediatr Pulmonol.* 2003;36:514–521.
29. Gilliland FD, Berhane K, Islam T, et al. Obesity and the risk of newly diagnosed asthma in school-age children. *Am J Epidemiol.* 2003;158:406–415.
30. Tantisira KG, Weiss ST. Complex interactions in complex traits: obesity and asthma. *Thorax.* 2001;56(Suppl 2):S64–S73.
31. Lutfiyya MN, Nika B, et al. Primary prevention of overweight and obesity: an analysis of national survey data. *Gen Intern Med.* 2008;23(6):821–823.
32. Centers for Disease Control and Prevention, National Center for Health Statistics. Prevalence of overweight among children and adolescents: United States, 2003–2004. http://www.cdc.gov/nchs/products/pubs/pubd/hestats/overweight/overwght_child_03.htm.
33. US Department of Health and Human Services, Institute of Medicine. Preventing childhood obesity: life in the balance, 2004. http://aspe.hhs.gov/health/reports/child_obesity/.
34. Centers for Disease Control and Prevention, National Center for Health Statistics. Prevalence of overweight and obesity among adults: United States, 2003–2004. http://www.cdc.gov/nchs/products/pubs/pubd/hestats/overweight/overwght_adult_03.htm.
35. Yun S, Zhu BP, Black W, Brownson RC. A comparison of national estimates of obesity prevalence from the behavioral risk factor surveillance system and the National Health and Nutrition Examination Survey. *Int J Obes.* 2006;30(1):164–170.
36. Strong JP, Malcom GT, McMahan CA, et al. Prevalence and extent of atherosclerosis in adolescents and young adults: implications for prevention from the Pathobiological Determinants of Atherosclerosis in Youth Study. *JAMA.* 1999;281:727–735.
37. McGill HC, McMahan CA, Malcom GT, Oalmann MC, Strong JP, Pathobiological Determinants of Atherosclerosis in Youth Research Group. Relation of glycohemoglobin and adiposity to atherosclerosis in youth. *Arterioscler Thromb Vasc Biol.* 1995;15:431–440.
38. Freedman DS, Mei Z, Srinivasan SR, Berenson GS, Dietz WH. Cardiovascular risk factors and excess adiposity among overweight children and adolescents: the Bogalusa Heart Study. *J Pediatr.* 2007;150(1):12–17, e2.
39. Johns Hopkins Bloomberg School of Public Health Center for Human Nutrition. Obesity rates continue to climb in the United States. *Public Health Center News Letter*. July 10, 2007.
40. Berenson GS, Srinivasan SR, Bao W, Newman WP III, Tracy RE, Wattigney WA. Association between multiple cardiovascular risk factors and atherosclerosis in children and young adults: the Bogalusa heart study. *N Engl J Med.* 1998;338:1650–1656.
41. Bao W, Srinivasan SR, Berenson GS. Persistent elevation of plasma insulin levels is associated with increased cardiovascular risk in children and young adults: the Bogalusa Heart Study. *Circulation.* 1996;93:54–59.
42. Davis PH, Dawson JD, Riley WA, Lauer RM. Carotid intimal-medial thickness is related to cardiovascular risk factors measured from childhood through middle age: the Muscatine Study. *Circulation.* 2001;104:2815–2819.
43. Lin BH, Guthrie J, Frazao E. Quality of children's diets at and away from home: 1994–1996. *Food Rev.* 1999;22:2–10.

44. Enns CW, Mickle SJ, Goldman JD. Trends in food and nutrient intakes by children in the United States. *Family Econ Nutr Rev*. 2002;14(2):56–58.
45. Jahns L, Siega-Riz AM, Popkin BM. The increasing prevalence of snacking among US children from 1977 to 1996. *J Pediatr.* 2001;138(4):493–498.
46. Nielsen SJ, Popkin BM. Patterns and trends in food portion sizes, 1977–1998. *JAMA.* 2003;289(4):450–453.
47. Cavadini C, Siega-Riz AM, Popkin BM. US adolescent food intake trends from 1965 to 1996. *Arch Dis Child.* 2000;83(1):18–24.
48. French SA, Lin BH, Guthrie JF. National trends in soft drink consumption among children and adolescents age 6 to 17 years: prevalence, amounts, and sources, 1977/1978 to 1994/1998. *J Am Diet Assoc.* 2003;103(10):1326–1331.
49. Centers for Disease Control and Prevention. Physical activity and good nutrition essential elements to preventing chronic disease and obesity. http://www.cdc.gov/nccdphp/ aag/pdf/aag_dnpa2004.pdf. Accessed June 2009.
50. Slyper AH. The pediatric obesity epidemic: causes and controversies. *J Clin Endocrinol Metab.* 2004;89(6):2540–2547.
51. Patrick K, Norman GJ, Calfas KJ, et al. Diet, physical activity, and sedentary behaviors as risk factors for overweight in adolescence. *Arch Pediatr Adolesc Med.* 2004;158(4):385–390.
52. Centers for Disease Control and Prevention. Physical activity levels among children aged 9–13 years: United States, 2002. *MMWR.* 2003;52(33):785.
53. Hayman LL, Meininger JC, Daniels SR, et al. Primary prevention of cardiovascular disease in nursing practice: focus on children and youth: a scientific statement from the American Heart Association Committee on Atherosclerosis, Hypertension, and Obesity in Youth of the Council on Cardiovascular Disease in the Young, Council on Cardiovascular Nursing, Council on Epidemiology and Prevention, and Council on Nutrition, Physical Activity, and Metabolism. *Circulation.* 2007;116:344–357.
54. Berkey CS, Rocket HRH, Gillman MW, Colditz GA. One-year changes in activity and in inactivity among 10- to 15-year-old boys and girls: relationship to change in body mass index. *Pediatrics.* 2003;111(4):836–843.
55. Eisenmann JC, Bartee RT, Wang MQ. Physical activity, TV viewing, and weight in US youth, 1999: Youth Risk Behavior Survey. *Obes Res.* 2002;10(5):379–385.
56. Jensen MD. Fatty acid oxidation in human skeletal muscle. *J Clin Invest.* 2002;110:1607–1609.
57. Pate RR, Reedson PS, Sallis JF, et al. Compliance with physical activity guidelines: prevalence in a population of children and youth. *Ann Epidemiol.* 2002;12(5):303–308.
58. Caspersen CJ, Pereira MA, Curran KM. Changes in physical activity patterns in the United States, by sex and cross-sectional age. *Med Sci Sports Exerc.* 2000;32(9):1601–1609.
59. Epstein LH, Roemmich JN, Robinson JL, et al. A randomized trial of the effects of reducing television viewing and computer use on body mass index in young children. *Arch Pediatr Adolesc Med.* 2008;162:239–245.
60. Schonfeld-Warden N. Pediatric obesity: an overview of etiology and treatment. *Pediatr Clin North Am.* 1997;44(2):339–361.
61. Hill J, Peters J. Environmental contributions to the obesity epidemic. *Science.* 1998;280: 1371–1374.
62. Li F, Harmer PA, Cardinal BJ, et al. Built environment, adiposity, and physical activity in adults aged 50–75. *Am J Prev Med.* 2008;35(1):38–46.

63. Lovasi GS, Moudon AV, Pearson AL, et al. Using built environment characteristics to predict walking for exercise. *Int J Health Geogr.* 2008;7:10.
64. Sallis J, Conway T, Prochaska J, et al. The association of school environments with youth physical activity. *Am J Public Health.* 2001;91(4):618–620.
65. Nutbeam, D. The health promoting school: closing the gap between theory and practice. *Health Promot Int.* 1992;7(3):151–153.
66. Li F, Harmer PA, Cardinal BJ, et al. Built environment, adiposity, and physical activity in adults aged 50–75. *Am J Prev Med.* 2008;35(1):38–46.
67. Macbeth AG. Bicycle lanes in Toronto. *ITE J.* 1999;69:38–40, 42, 44, 46.
68. Painter K. The influence of street lighting improvements on crime, fear, and pedestrian street use after dark. *Landsc Urban Plan.* 1996;35:193–201.
69. Ewing R, Schmid T, Killingsworth R, Zlot A, Raudenbush S. Relationship between urban sprawl and physical activity, obesity, and morbidity. *Science Health Promot.* 2003;18(1):47–57.
70. Wakefield J. Deconstructing obesity. *Environ Health Perspect*. 2004;112(11). http://ehp.niehs.nih.gov/docs/allpubs.html. Accessed July 10, 2009.
71. McCann B, Ewing R. Measuring the health effects of sprawl: Smart Growth America Surface Transportation Policy Project. http://www.smartgrowthamerica.org. Published 2003. Accessed February 2005.
72. Barsh G, Farooqi IS, O'Rahilly S. Genetics of body-weight regulation. *Nature.* 2000;404(6): 644–651.
73. Bouchard C. The biological predisposition to obesity: beyond the thrifty genotype scenario. *Int J Obes.* 2007;31:1337–1339.
74. Newell A, Zlot A, Silvey K, Ariail K. Addressing the obesity epidemic: a genomics perspective. *Prev Chronic Dis.* 2007;4(2):1–6.
75. Rankinen T, Zuberi A, Chagnon YC, et al. The human obesity gene map: the 2005 update. *Obesity.* 2006;14(4):529–644.
76. Herbert A, Gerry NP, McQueen MB, et al. A common genetic variant is associated with adult and childhood obesity. *Science.* 2006;312(5771):279–283.
77. Loos R, Bouchard C. Obesity: is it a genetic disorder? *J Int Med.* 2003;254(5): 401–425.
78. Keller K, Pietrobelli A, Faith M. Genetics of food intake and body composition: lessons from twin studies. *Acta Diabetol.* 2003;40:S95–S100.
79. Prentice AM, Hennig BJ, Fulford AJ. Evolutionary origins of the obesity epidemic: natural selection of thrifty genes or genetic drift following predation release? *Int J Obes.* 2008;32(11):1607–1610.
80. Ravussin E, Bogardus C. Energy expenditure in the obese: is there a thrifty gene? *Infusionstherapie.* 1990;17:108–112.
81. Sharma A. The thrifty-genotype hypothesis and its implications for the study of complex genetic disorders in man. *J Mol Med.* 1998;76:568–571.
82. Körner A, Kiess W, Stumvoll M, Kovacs P. Polygenic contribution to obesity: genome-wide strategies reveal new targets. *Front Horm Res*. 2008;36:12–36.
83. The Metabolic Syndrome Institute. Homepage. http://www.metabolicsyndrome institute.com/informations/history/index.php#lien_a.
84. Sarafidis PA, Nilsson PM. The metabolic syndrome: a glance at its history. *J Hypertens.* 2006;24(4):621–626.
85. Phillips LK, Prins JB. The link between abdominal obesity and the metabolic syndrome. *Curr Hypertens Rep.* 2008;10(2):156–164.
86. Bergman RN, Kim SP, Hsu IR, et al. Abdominal obesity: role in the pathophysiology of metabolic disease and cardiovascular risk. *Am J Med.* 2007;120(2 Suppl 1): S3–S8, S29–32.

87. Lawlor DA, Smith GD, Ebrahim S. Does the new International Diabetes Federation definition of the metabolic syndrome predict CHD any more strongly than older definitions? findings from the British Women's Heart and Health Study. *Diabetologia.* 2006;49(1):41–48.
88. Tong PC, Kong AP, So WY, et al. The usefulness of the International Diabetes Federation and the National Cholesterol Education Program's Adult Treatment Panel III definitions of the metabolic syndrome in predicting coronary heart disease in subjects with type 2 diabetes. *Diabetes Care.* 2007;30(5):1206–1211.
89. Patel SB, Reams GP, Spear RM, Freeman RH, Villarreal D. Leptin: linking obesity, the metabolic syndrome, and cardiovascular disease. *Curr Hypertens Rep.* 2008;10(2):131–137.
90. Wannamethee SG, Tchernova J, Whincup P, et al. Plasma leptin: associations with metabolic, inflammatory and haemostatic risk factors for cardiovascular disease. *Atheroscler.* 2007;191(2):418–426.
91. Wang ZV, Scherer PE. Adiponectin, cardiovascular function, and hypertension. *Hypertension.* 2008;51(1):8–14.
92. Störk S, Bots ML, Angerer P, et al. Low levels of adiponectin predict worsening of arterial morphology and function. *Atherosclerosis.* 2007;194:e147.
93. Pischon T, Girman CJ, Hotamisligil GS, Rifai N, Hu FB, Rimm EB. Plasma adiponectin levels and risk of myocardial infarction in men. *JAMA.* 2004;291:1730–1737.
94. Wannamethee G, Sarwar N, Tchernova J, et al. Adiponectin and coronary heart disease: a prospective study and metaanalysis. *Circulation.* 2006;114:623–629.
95. Shibata R, Ouchi N, Ito M, et al. Adiponectin-mediated modulation of hypertrophic signals in the heart. *Nat Med.* 2004;10:1384–1389.
96. Gualillo O, González-Juanatey JR, Lago F. The emerging role of adipokines as mediators of cardiovascular function: physiologic and clinical perspectives. *Trends Cardiovasc Med.* 2007;17(8):275–283.
97. Vázquez-Vela M, Torres N, Tovar A. White adipose tissue as endocrine organ and its role in obesity. *Arch Med Res.* 2008;39(8):715–728.
98. Sun Y, Asnicar M, Smith RG. Central and peripheral roles of ghrelin on glucose homeostasis. *Neuroendocrinology.* 2007;86(3):215–228.
99. Gil-Campos M, Aguilera CM, Cañete R, Gil A. Ghrelin: a hormone regulating food intake and energy homeostasis. *Br J Nutr.* 2006;96(2):201–226.
100. Woo KS, Chook P, Yu CW, et al. Effects of diet and exercise on obesity-related vascular dysfunction in children. *Circulation.* 2004;109(16):1981–1986.
101. Leibel RL, Rosenbaum M, Hirsch J. Changes in energy expenditure resulting from altered body weight. *N Engl J Med.* 1995;332:621–628.
102. Rosenbaum M, Nicolson M, Hirsch J, Murphy E, Chu F, Leibel RL. Effects of weight change on plasma leptin concentrations and energy expenditure. *J Clin Endocrinol Metab.* 1997;82(11):3647–3654.
103. Korner J, Aronne LJ. Pharmacological approaches to weight reduction: therapeutic targets. *J Clin Endocrinol Metab.* 2004;89(6):2616–2621.
104. Padwal RS, Majumdar SR. Drug treatments for obesity: orlistat, sibutramine, and rimonabant. *Lancet.* 2007;369(9555):71–77.
105. Li Z, Maglione M, Tu W, et al. Meta-analysis: pharmacologic treatment of obesity. *Ann Intern Med.* 2005;142(7):532–546.
106. Halford JC. Pharmacotherapy for obesity. *Appetite.* 2006;46(1):6–10.
107. Wilding J. Treatment strategies for obesity. *Obesity Rev.* 20078(Suppl 1):S137–S144.

108. Bays H. Current and investigational antiobesity agents and obesity therapeutic treatment targets. *Obesity Res.* 2004;12(8):1197–1211.
109. Pagotto U, Vanuzzo D, Vicennati V, Pasquali R. Pharmacological therapy of obesity. *G Ital Cardiol.* 2008;9(4 Suppl 1):S83–S93.
110. Foster-Schubert KE, Cummings DE. Emerging therapeutic strategies for obesity. *Endocr Rev.* 2006;27(7):779–793.
111. Hendricks EJ, Rothman RB, Greenway FL. How physician obesity specialists use drugs to treat obesity. *Obesity.* 2009;17:1730–1735.
112. Tsai WS, Inge TH, Burd RS. Bariatric surgery in adolescents: recent national trends in use and in-hospital outcome. *Arch Pediatr Adolesc Med.* 2007;161(3):217–221.
113. Inge TH, Zeller MH, Lawson ML, Daniels SR. A critical appraisal of evidence supporting a bariatric surgical approach to weight management for adolescents. *J Pediatr.* 2005;147(1): 10–19.
114. Schilling PL, Davis MM, Albanese CT, Dutta S, Morton J. National trends in adolescent bariatric surgical procedures and implications for surgical centers of excellence. *J Am Coll Surg.* 2008;206(1):1–12.
115. Tsai AG, Wadden TA. Systematic review: an evaluation of major commercial weight loss programs in the United States. *Ann Intern Med.* 2005;142(1):56–66.
116. Witherspoon B, Rosenzweig M. Industry-sponsored weight loss programs: description, cost, and effectiveness. *J Am Acad Nurse Pract.* 2004;16(5):198–205.
117. Dansinger ML, Gleason JA, Griffith JL, Selker HP, Schaefer EJ. Comparison of the Atkins, Ornish, Weight Watchers, and Zone Diets for weight loss and heart disease risk reduction: a randomized trial. *JAMA.* 2005;293:43–53.
118. Tsai AG, Wadden TA. Systematic review: an evaluation of major commercial weight loss programs in the United States. *Ann Intern Med.* 2005;142(1):56–66.
119. Truby H, Baic S, DeLooy A, et al. Randomised controlled trial of four commercial weight loss programmes in the UK: initial findings from the BBC "diet trials." *BMJ.* 2006;332 (7553):1309–1314.
120. Rating the diets from Atkins to Dr. Sears Zone. *Consum Rep.* 2005;70(6):18–22.
121. Schwartz H. *Never Satisfied: A Cultural History of Diets, Fantasies, and Fat.* New York: Free Press; 1986.
122. Berland T. *Consumer's Guide to Rating the Diets: Everything You Should Know about the Diets Making News.* Skokie, IL: Publications International; 1980.
123. Gardner CD, Kiazand A, Alhassan S, et al. Comparison of the Atkins, Zone, Ornish, and LEARN diets for change in weight and related risk factors among overweight premenopausal women: the A to Z Weight Loss Study: a randomized trial. *JAMA.* 2007;297(9):969–977.
124. Freedman M, King J, Kennedy E. Popular diets: a scientific review. *J Obes Res.* 2001;9(Suppl 1):S1–S40.
125. BestDietforMe.com. Diet Book Reviews. http://www.bestdietforme.com/diet books/ DietBookReviews.aspx. Accessed July 2009.
126. Sacks FM, Bray GA, Carey VJ, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. *N Engl J Med.* 2009;360(9): 859–873.
127. Eckel R. Clinical practice. Nonsurgical management of obesity in adults. *N Engl J Med.* 2008;358(18):1941–1950.
128. Bureau of Consumer Protection of the Federal Trade Commission, American Society for Clinical Nutrition, National Institute of Diabetes and Digestive and Kidney Diseases, Centers for Disease Control and Prevention. Report of the

presiding panel: commercial weight loss products and programs: what consumers stand to gain and lose: A public conference on the information consumers need to evaluate weight loss products and programs; October 16–17, 1997; Washington, DC.
129. US Department of Health and Human Services, Office of the Surgeon General. The surgeon general's call to action to prevent and decrease overweight and obesity. http://www.surgeongeneral.gov/topics/obesity/calltoaction/cover.htm. Published 2001. Accessed July 28, 2009.
130. Blumenthal S, Hendi J, Marsillo L. A public health approach to decreasing obesity. *JAMA.* 2002;288(17):2178.
131. Levi J, Trust for America's Health, Robert Wood Johnson Foundation. F as in fat: how obesity policies are failing in America. http://www.rwjf.org/pr/product.jsp?id=45050. Accessed July 22, 2009.
132. Federal Trade Commission. Children's advertising: proposed trade regulation rulemaking and public hearing. *Fed Regist.* 1978;43(82):17967–17972.
133. Story M, French S. Food advertising and marketing directed at children and adolescents in the US. *Int J Behav Nutr Phys Act.* 2004;1:3–13.
134. Powell LM, Szczypka G, Chaloupka FJ. Exposure to food advertising on television among US children. *Arch Pediatr Adolesc Med.* 2007;161(6):553–560.
135. Gostin LO. Law as a tool to facilitate healthier lifestyles and prevent obesity. *JAMA.* 2007;297(1):87–90.
136. Mello MM, Studdert DM, Brennan TA. Obesity—the new frontier of public health law. *N Engl J Med.* 2006;354(24):2601–2610.
137. Wald NJ. Obesity and public health law. *N Engl J Med.* 2006;355(12):1283
138. Petty RD, Seiders K. Taming the obesity beast: children, marketing, and public policy considerations. *J Public Policy Mark.* 2007;26(2):236.
139. Tillotson J. America's obesity: conflicting public policies, industrial economic development, and unintended human consequences. *Annu Rev Nutr.* 2004;24:617–643.
140. Lee T, Oliver JE. Public opinion and the politics of America's obesity epidemic. May 2002. KSG Faculty Research Working Paper Series.
141. Oliver JE, Lee T. Public opinion and the politics of obesity in America. *J Health Polit Policy Law.* 2005;30(5):923–954.
142. Blendon R, Hunt K, Benson J, Fleischfresser C, Buhr T. Understanding the American public's health priorities: a 2006 perspective. *Health Aff.* 2006;25(6):w508–w515.
143. Olshansky SJ, Passaro DJ, Hershow RC, et al. A potential decline in life expectancy in the United States in the 21st century. *N Eng J Med.* 2005;352:1139–1145.
144. Hogan P, Dall T, Nikolov P. Economic costs of diabetes in the US in 2002: American Diabetes Association. *Diabetes Care.* 2003;26(3):917–932.
145. Wolf AM, Colditz GA. Current estimates of the economic cost of obesity in the United States. *Obes Res.* 1998;6:97–106.
146. Caprio S, Daniels SR, Drewnowski A, et al. Influence of race, ethnicity, and culture on childhood obesity: implications for prevention and treatment: a consensus statement of Shaping America's Health and the Obesity Society. *Diabetes Care.* 2008;31(11):2211–2221.
147. Institute of Medicine. *Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids*. Washington, DC: National Academies Press; 2005.

148. New York City Department of Health, Mental Hygiene Board of Health. Notice of adoption of an amendment (81.08) to article 81 of the New York City Health Code. http://www.nyc.gov/html/doh/downloads/pdf/public/notice-adoption-hc-art81-08.pdf. Accessed June 6, 2009.
149. Robinson M, Bloom P, Lurie N. Combating obesity in the courts: will lawsuits against McDonald's work? *J Public Policy Mark.* 2005;24:299–306.
150. Health Promotion Program State Legislation and Statute Database. Obesity, obesity-childhood, physical activity, and nutrition. Washington, DC: National Conference of State Legislatures.
151. Jacobson MF, Brownell KD. Small taxes on soft drinks and snack foods to promote health. *Am J Pub Health.* 2000;90:854–857.
152. Federal Trade Commission, US Food and Drug Administration. Working agreement between FTC and Food and Drug Administration. 4 Trade Reg Rep. (CCH) 9,850.01; 1971.
153. US Food and Drug Administration. Food labeling, health messages and label statements: reproposed rule. *Fed Regist*. 1990;55:5176–5192.
154. Bryd-Bredbenner C. Designing a consumer friendly nutrition label. *J Nutr Educ.* 1994;26:180–190.
155. Levy A, Fein S. Consumers' ability to perform tasks using nutrition labels. *J Nutr Educ.* 1998;30(4):210–217.

This page intentionally left blank

CHAPTER 4

The Global Burden of Asthma

Philippa Shirtcliffe, MD, MBChB, FRACP, and Richard Beasley, MD, MBChB, FRACP, DM, FAAAAI, FRCP, DSc

INTRODUCTION

The rising global burden of chronic, noncommunicable diseases over recent decades has been labeled "the neglected epidemic."[1] One of the most important challenges in global health is how to address this epidemic, despite the unfinished agenda of the Millennium Development Goals that were adopted by 189 nations during the United Nations Millennium Summit in September 2000. Although attempting to encapsulate the development aspirations of the world as a whole, these goals focus particularly on perinatal and maternal health and communicable diseases.[2] In 2001, noncommunicable diseases accounted for 59 percent of global mortality; nearly 54 percent of deaths in low- and middle-income countries, and 87 percent of deaths in high-income countries.[3] A 2005 World Health Organization (WHO) report similarly identified that in 23 developing countries, chronic diseases were responsible for 50 percent of the total disease burden.[4] Modeling has shown that population growth, aging, and the relative success of efforts to reduce communicable disease will drive a substantial increase in the numbers of deaths from chronic disease globally (estimated at nearly 70 percent by 2030),[3] particularly in low- and middle-income countries that at least can sustain the economic effects.

Over recent decades, asthma has become one of the most common chronic diseases in the world and is now the most common chronic disease of childhood in many countries.[5] It is estimated that as many as 300 million people of all ages and ethnic backgrounds suffer from asthma.[6] Considerably higher estimates can be obtained with less conservative criteria for the diagnosis of clinical asthma. The increase in the prevalence of asthma has been associated with an increase in atopic sensitization and is paralleled by similar increases in other allergic disorders such as eczema and rhinitis. The international patterns of asthma prevalence are not explained by the current knowledge of the causation of asthma. Asthma accounts for about 1 in every 250 deaths worldwide.[6] Many of the deaths are preventable

and are the result of suboptimal long-term medical care and delay in obtaining medical help during the final attack.

The economic cost of asthma is considerable both in terms of direct medical costs (such as the cost of pharmaceuticals and hospital admissions) and indirect medical costs (such as time lost from work and premature death). The global burden of this disease to governments, health care systems, families, and patients is substantial. Indeed, the burden of asthma in many countries is of sufficient magnitude to warrant its recognition as a priority disorder in government health strategies.

ASTHMA DEFINITION

It is increasingly apparent that, rather than being a single disease, asthma is a heterogeneous disorder with respect to immunopathology, clinical phenotypes, responses to therapy, and natural history.[7–9] Despite this increased complexity, however, chronic airway inflammation, reversible airflow obstruction, and enhanced bronchial reactivity remain core features, and the diagnosis of asthma is still based on the recognition of a characteristic pattern of symptoms and signs, and in the absence of an alternative explanation. For the purposes of this discussion, the operational description based on the functional consequences of airway inflammation provided by the Global Initiative for Asthma (GINA) is a useful starting point:

> Asthma is a chronic inflammatory disorder of the airways in which many cells and cellular elements play a role. The chronic inflammation is associated with airway hyperresponsiveness that leads to recurrent episodes of wheezing, breathlessness, chest tightness, and coughing, particularly at night or early in the morning. These episodes are usually associated with widespread, but variable, airflow obstruction within the lung that is often reversible either spontaneously or with treatment.[10]

ASTHMA MANAGEMENT

The clinical spectrum of asthma is highly variable, yet a wealth of evidence indicates that the manifestations can be controlled with appropriate treatment. When asthma is controlled, there should be no more than the occasional recurrence of symptoms and severe exacerbations should be rare. Yet, in many areas of the world, persons with asthma do not have access to basic asthma medications or medical care. The current approach to the long-term management of asthma is based on the use of inhaled short-acting beta-agonist "bronchodilator" drugs for symptomatic relief and "controller" or "prophylactic" drugs to control airway inflammation. Inhaled glucocorticosteroids are the most effective controller medications currently available, but they must be taken daily on a long-term basis to keep asthma under clinical control. More recently, long-acting inhaled beta-agonists have been shown to be effective when combined with inhaled corticosteroids, allowing a reduced dose of the latter.

It has been increasingly recognized by the medical community that the key person in the long-term management of asthma is the informed patient. Greater emphasis therefore should be given to self-management of asthma and asthma education, including the use of asthma action plans. This trend follows that of other major chronic diseases, such as diabetes. The increasing emphasis on informed

self-management raises issues of access to asthma health care and asthma education as well as issues of communication between patients and physicians, and social and psychological factors that may affect a patient's asthma self-management skills.

THE GLOBAL BURDEN OF ASTHMA

In the past, a significant contribution of epidemiology to the study of cancer, cardiovascular disease, and other chronic disorders has included the analyses of patterns of disease prevalence and incidence across demographic groups, geographic areas, and time. In particular, many of the epidemiologic hypotheses concerning the causes of chronic disease have stemmed, at least in part, from international comparisons. By the 1990s, it had long been suspected that asthma prevalence had been increasing in the second half of the 20th century, not only in developed countries but also in the developing world. This, however, was a particularly difficult issue to resolve because of the lack of systematic standardized studies measuring changes in asthma prevalence over time. Some reviewers argued that the increases in reported prevalence were largely due to increased awareness, labeling, and diagnosis of asthma symptoms. Nevertheless, most studies that determined the prevalence of asthma symptoms using the same methodology in the same community at different times reported that asthma prevalence has increased in recent decades and that the magnitude of the increase in some cases has been substantial.[11]

To more clearly define the scale of the problem worldwide and identify possible factors affecting prevalence, the International Study of Asthma and Allergies in Childhood (ISAAC) was established in 1991. This effort required the gathering of comparable information on large numbers of children in random population samples around the world. Using a simple comparison of symptom prevalence in a questionnaire survey (written and video), ISAAC collected data on the prevalence of the symptoms of childhood asthma, rhino-conjunctivitis, and eczema. Phase I of this global study was conducted over the early to mid-1990s and involved more than 700,000 children ages 6 to 7 and 13 to 14 years old from 155 centers in 56 countries.[12]

Between 2002 and 2003, phase III of ISAAC, using the same methodology, was conducted on 498,083 children from centers that previously completed phase I, and was undertaken to provide more comprehensive information on how the prevalence of these symptoms varied around the world. The objective was to examine time trends in the prevalence of symptoms of asthma and related disorders. Data were also collected from 128 new centers (involving another 689,413 children), the majority of which were from countries in Latin America, Eastern Europe, and Africa that had little, if any, epidemiologic data on asthma. With the diverse socioeconomic backgrounds of the participating centers, these data are useful not only as a basis for studying the burden of childhood asthma worldwide, but also for exploring the impact of economic development on the prevalence and severity of asthma symptoms.[13, 14]

A similar approach to assess geographic variations in asthma and allergy using the same instruments and definitions in the early to mid-1990s was used by the European Community Respiratory Health Survey (ECRHS). This involved 137,619 adults ages 20 to 44 years of age. The phase I results include data from 48 centers in

22 countries, predominantly in Western Europe, with 9 centers from 6 countries outside Europe.[15] From the ISAAC and ECRHS data, a number of key patterns were identified:

1. Marked differences in the prevalence of asthma exist worldwide.
2. Asthma prevalence is generally higher in developed countries than in developing countries.
3. Asthma symptom prevalence rates are highest in English-speaking countries. It was subsequently observed that English-speaking populations also have the highest prevalence of atopy.
4. Among the non-English-speaking European countries, asthma prevalence is higher in Western Europe, with lower prevalence rates in Eastern and Southern Europe.
5. Asthma prevalence is increasing in developing countries as they become more industrialized and developed. A similar trend is observed when communities move from a rural to an urban environment.
6. The prevalence of other allergic disorders such as allergic rhinitis, atopic eczema, and urticaria are increasing worldwide.

ISAAC phase I revealed marked geographic variations in the prevalence and severity of asthma symptoms, even within genetically similar groups. Specifically, 12-month prevalence rates of asthma from the written questionnaires for each center by country showed that the highest prevalence was a twentyfold difference among the highest center compared with the center with the lowest prevalence (ranging from 1.6 to 36.8 percent), with an eightfold variation seen between the 10th and 90th percentiles (3.9 to 30.6 percent). For asthma symptoms, the highest prevalence rates were from the centers in the United Kingdom, New Zealand, Australia, and the Republic of Ireland (29.1 to 32.2 percent), followed by some Latin American non-English-speaking countries. The lowest prevalence rates were reported from Albania, Romania, Indonesia, Greece, China, and the Russian Federation (2.1 to 4.4 percent).[16] Although asthma prevalence rates generally were similar in centers within the same country, in some countries, such as India, Ethiopia, Italy, and Spain, large variations were seen between centers. Wide variations also existed within regions, especially within Europe and Asia.

The phase I results from the ECRS also found a wide variation in the prevalence of asthma symptoms in adults. The lowest prevalence rates were from India and Algeria, followed by centers in Italy, France, Belgium, and Germany. The highest prevalence rates were predominantly from centers in the United Kingdom, New Zealand, Australia, and the United States—that is, English-speaking countries.[15] When the prevalence rates from ISAAC were compared with the data from ECRHS, the estimates in the latter were consistently lower but generally the correlation was good.

The ISAAC phase III study showed striking variations in the prevalence of asthma symptoms between different geographic areas and populations. The global average for current wheeze based on the written questionnaire was 14.1 percent, ranging from 5.1 percent in Northern and Eastern Europe to 22 percent in Oceania.[14] Fifteen percent of centers had a prevalence of current wheeze $\geq$20 percent, and these were mostly from the English-speaking countries and Latin America. Nine percent of centers had a prevalence of more than 5 percent, and they were mostly in

the Indian subcontinent, Asia-Pacific, Eastern Mediterranean, and Northern and Eastern Europe.

Globally, 6.9 percent of adolescents had symptoms of severe asthma, ranging from 3.8 percent in Asia-Pacific and Northern and Eastern Europe to 11.3 percent in North America. Centers with the highest prevalence of severe asthma symptoms ($\geq$7.5 percent) were found mostly in the English-speaking countries and Latin America, although these also included many centers in Africa. Centers with the lowest prevalence of severe asthma were located mainly in the Indian subcontinent, Northern and Eastern Europe, Asia-Pacific, and the Eastern Mediterranean. A different pattern was observed in the proportion of current wheezers with symptoms of severe asthma. Africa (51 percent), the Indian subcontinent (48.2 percent), and the Eastern Mediterranean (47.2 percent) were ranked higher than the English-speaking countries (46 percent) and Latin America (38.3 percent). This suggests that the disease may be more severe in these less-affluent countries.

The main source of data regarding temporal comparisons of the prevalence of asthma comes from ISAAC phase III.[13] The three main findings are as follows. First, in most high-prevalence countries, particularly the English-speaking countries, the rise in the prevalence of asthma symptoms has peaked and may have even begun to decline. This finding is consistent with those of other recent studies examining time trends in children and adults. Germany and Finland are the exception, showing a relatively high prevalence in both phase I and in phase III. Second, a number of countries that had high or intermediate levels of symptom prevalence in phase I have shown significant increases in prevalence in phase III; these include Latin American countries such as Costa Rica, Panama, Mexico, Argentina, and Chile; Eastern European countries such as the Ukraine and Romania; and North African countries such as Tunisia, Morocco, and Algeria. Third, with the exception of India, all of the countries with very low symptom prevalence rates in phase I reported increases in prevalence in phase III, with increases particularly marked in Indonesia and China.

The ISAAC phase III trends for English-speaking countries and Western Europe are intriguing and, to some extent, reassuring; however, they should not be taken to indicate that the global pandemic of asthma is easing or that the worst is over. The phase III findings show striking increases for Latin American countries to the extent that asthma may soon be described as a Spanish- and Portuguese-speaking disease rather than an English-speaking disease. The increases for China are of potentially major significance given the size of China's population and its rapid economic growth. Furthermore, there are some intriguing contrasting patterns with, for example, decreases in prevalence in India but increases in China, Indonesia, Taiwan (China), and the Republic of Korea, and stronger increases in Morocco, Algeria, and Tunisia. As with the phase I findings, this new evidence on time trends hopefully will lead to further questioning and testing of current theories, and the development of new theories of asthma aetiology.

RISK FACTORS

International patterns of asthma prevalence raise major issues regarding current knowledge of the risk factors and causes of asthma. In particular, it is evident that the current recognized risk factors for the development of asthma such as atopy, air

pollution, smoking, exposure to house dust mites, and other allergens probably cannot fully account for either the worldwide increase in prevalence or the international variations in asthma prevalence that have been observed. For example, the global patterns of asthma prevalence are consistent with the considerable body of evidence that air pollution is not a major risk factor for the development of asthma. Regions such as China and Eastern Europe, where some of the highest levels of traditional air pollution such as particulate matter and sulphur dioxide are found, generally have lower asthma prevalence than the countries of Western Europe, North America, Australia, and New Zealand that have lower levels of pollution.

It appears unlikely that the international prevalence patterns can be explained by differences in smoking rates. This is illustrated by the study of Chinese communities living in the Southeast Asian region, in which the lowest prevalence of asthma was observed in the community living in mainland China, despite the highest level of cigarette smoking.[17] Furthermore, although sensitization to the house dust mite allergen is an important risk factor for the development of asthma, and some evidence indicates that "Westernization" may increase levels of exposure to the house dust mite, variations in house dust mite exposure are unlikely to explain the international prevalence patterns. For example, levels of house dust mite allergens were similar in houses in Hong Kong and Guangzhou despite the considerable difference in asthma prevalence.[18]

Studies in China, Europe, and Africa have shown that marked differences in asthma prevalence may occur in populations despite similar rates of atopic sensitization. These findings support the view that in addition to atopic sensitization, other risk factors may be important in the development of asthma in a susceptible population.[19] This knowledge has resulted in research focusing not only on "established" risk factors, such as allergen exposure, but also on other factors that may "program" the initial susceptibility to sensitization or contribute to the development of asthma independent of atopic sensitization.

Another hypothesis that recently has been proposed is that acetaminophen (paracetamol) use may contribute to the development of asthma. An association has been observed between asthma and exposure to acetaminophen *in utero,* in childhood, and in adult life.[20] Acetaminophen use is particularly common in English-speaking countries, which have among the highest rates of asthma. It has been proposed that this association, if causal, may be due to acetaminophen increasing oxidant-induced airways inflammation and enhancing T_H2 immune responses.

Airway inflammation in asthma is a multicellular process involving mainly eosinophils, neutrophils, CD4+ T (helper) lymphocytes (T_H2 cells), and mast cells.[7] A fundamental feature of asthma associated with allergic sensitization is the ability of the airway to recognize common environmental allergens and to generate a T_H2 cytokine response to them. Based on this understanding, the "hygiene hypothesis" has been proposed as one explanation for an increase in the development of allergic diseases, including asthma. This basically attempts to explain the influence of exposures to microbial sources—both in terms of bacterial and viral infections and also noninvasive exposures in the environment—on our innate and adaptive immune response. At a simplistic level, the mechanism has been described as a skewing of the T helper cell–mediated immune response (T_H1/T_H2 balance) away from T_H1 cells toward allergy-promoting T_H2 cells. This response has been attributed to a

reduced microbial burden during childhood, as a consequence of a Westernized lifestyle.[21, 22] Although mechanisms related to T_H1/T_H2 balance are of undoubted importance for the development of allergies, it is clear that the jigsaw is much more complex, with many dimensions, and it is now clear that asthma is not a single disease but a wide range of different disorders.[7]

MORTALITY

Death from asthma is a complex phenomenon and many factors relevant to the risk of asthma mortality have changed in varying degrees in different countries over time, making international time trends difficult to interpret. These factors potentially could include the characteristics of the disease and its treatment, the characteristics of the population, and changes in the underlying level of asthma prevalence. Ongoing surveillance of international mortality trends is important, however, because it provides a crude signal of major changes in asthma death rates in different populations. In accordance with standard practice, asthma mortality rates are confined to the 5- to 34-year-old age-group because the correct assignment of asthma mortality is more firmly established in this group.

Between 1900 and 1940, asthma mortality was uniformly low and relatively stable. The death rate began to increase gradually in the 1940s in a number of countries, including New Zealand and Australia, in which a threefold increase over a 15-year period was observed. Mortality declined in the late 1950s in New Zealand, England, and Wales, but not in Australia. In contrast, little change in asthma mortality rates was observed in the United States during this period. Examination of the time trends in relation both to revisions in International Classification of Diseases (ICD) coding and to trends from other obstructive respiratory disorders does not suggest that the patterns observed were due to changes in coding or diagnostic fashion. Interestingly, isoprenaline was introduced in a nebulizer during the 1940s when mortality began to increase, but whether this had a role in the increase in mortality was not examined in detail at the time.

Statistics show that a dramatic increase occurred in asthma mortality in the mid-1960s with a mean increase of 53 percent from 0.55 per 100,000 in 1960 and 1961 to a peak of 0.84 in 1966 and 1967. This trend was followed by a progressive decline to a nadir of 0.45 per 100,000 in 1974 and 1975. A gradual increase was then found in asthma mortality rates to a peak of 0.62 per 100,000 in 1985 and 1986, with a mean increase of 38 percent during this period. Since the late 1980s, there has been a widespread and progressive reduction in mortality rates to a level of 0.23 per 100,000 in 2004 and 2005. The magnitude of the reduction has been substantial, with a mean reduction of 63 percent in the countries included. Although the reduction in mortality rates generally was consistent among countries, the timing of its onset varied. For example, in England and Wales, mortality rates peaked at 0.99 in 1987, with a progressive decline to 0.33 in 2005, whereas in the United States, mortality did not peak until 1995, with a subsequent decline from 0.56 to 0.35 in 2005.

A unifying hypothesis has been proposed to explain these international time trends, namely, that they relate to changes in drug treatment.[23] The epidemics of asthma mortality occurred in at least six countries in the 1960s as a result of the

widespread use of isoprenaline forte, and, in New Zealand in the 1970s and 1980s, as a result of the use of fenoterol.[24] In fact, a recent systematic review has identified that the second epidemic was not restricted to New Zealand, with substantive increases in mortality observed during the 1980s in many other countries.[23] In some of these countries, it is likely that the increase in mortality was at least partly due to the introduction and increasing use of fenoterol, which has been shown to increase the risk of mortality when compared with other beta-agonists such as salbutamol. The potential contribution of other beta-agonists to the trend of increasing mortality during this period is uncertain.

The most likely reason for the subsequent marked reduction in asthma mortality since the late 1980s is the widespread and progressive increase in the use of inhaled corticosteroid (ICS) therapy. The increased use of ICS therapy over the past 20 to 30 years has been substantial and has occurred in countries in different regions of the world. Substantive evidence exists that ICS reduces the risk of asthma death, an effect that is achieved with low doses and has not been demonstrated with other asthma therapies.[25, 26] Other improvements in management, such as guided self-management systems of care, also are likely to have contributed to the reduction in mortality. Reassuringly, the reductions have occurred during periods in which the use of long-acting beta-agonist therapy has been increasing, the use of which as a sole therapy has been implicated in asthma mortality. Although these international time trend data do not rule out a contribution to asthma mortality, they do suggest that the widespread use of long-acting beta-agonist therapy with ICS therapy is not associated with major risk.[23]

DALYs

Disability-adjusted life years (DALYs) are another method of quantifying burden of disease. A DALY is a summary measure that combines years of life lost due to premature death and years of life lived with disability. One DALY can be thought of as one lost year of healthy life, and the burden of disease as a measurement of the gap between the present health of a population and an ideal situation in which everyone in the population lives into old age in full health.[4] Comparing DALYs is a useful way to compare the relative importance of chronic diseases.

Asthma was the 25th leading cause of DALYs lost worldwide in 2001. In comparison, perinatal conditions, lower respiratory tract infections, and HIV/AIDS were the top-three diseases that caused lost DALYs. Diabetes mellitus was ranked number 23. In 2005, the WHO estimated that chronic diseases were responsible for 49 percent of the total worldwide burden of disease measured in DALYs. The number of DALYs lost due to asthma worldwide has been estimated to be about 15 million per year. Worldwide, asthma accounts for around 1 percent of all DALYs lost, which reflects the high prevalence and severity of asthma. This figure is similar to those for diabetes, cirrhosis of the liver, and schizophrenia.[6]

BARRIERS TO REDUCING THE GLOBAL BURDEN OF ASTHMA

Asthma represents a huge burden to health care systems in many regions around the world. As detailed in earlier sections, asthma is more prevalent in industrial

Table 4.1
Disability-Adjusted Life Years Lost Due to Asthma Worldwide, as Ranked with Other Common Disorders

Rank	Disorder	Number of DALYs ($\times 10^8$)
1	Perinatal conditions	98.4
2	Lower respiratory tract infections	90.7
3	HIV/AIDS	88.4
4	Unipolar depressive disorders	65.9
5	Diarrheal disease	62.5
6	Ischaemic heart disease	58.7
7	Cerebrovascular disease	45.9
8	Malaria	42.3
9	Road traffic accidents	37.7
10	Tuberculosis	36.0
11	Maternal conditions	30.9
12	Chronic obstructive pulmonary disease	29.9
13	Congenital abnormalities	28.1
14	Measles	26.5
15	Hearing loss—adult onset	25.9
16	Violence	20.2
17	Self-inflicted injuries	19.9
18	Alcohol use disorders	19.8
19	Protein-energy malnutrition	16.7
20	Osteoarthritis	16.4
21	Schizophrenia	15.9
22	Falls	15.7
23	Diabetes mellitus	15.4
24	Cirrhosis of the liver	15.1
25	ASTHMA	15.0
26	Bipolar affective disorder	13.8
27	Pertussis	12.5
28	Alzheimer's and other dementias	12.4
29	Sexually transmitted diseases excluding HIV	12.4
30	Iron-deficiency anemia	12.0

Source: Masoli M., et al. The global burden of asthma: executive summary of the GINA Dissemination Committee Report. Allergy. 2004; 59:469–78.

Note: AIDS = acquired immune deficiency syndrome; DALYs = Disability-adjusted life years; HIV = human immunodeficiency virus.

countries than in low-income countries, although the total number of people with asthma is higher in the developing world. The financial burden can be very high not just for those living with asthma but also for their families, the health care systems, and governments.[27] These costs increase significantly when there is poor case management, and both inadequate treatment and the high costs of medication can contribute to disability, absenteeism, and poverty. Until there is a greater understanding of the factors that cause asthma and novel public health and pharmacological measures become available to reduce the prevalence of asthma, the priority is to

ensure that cost-effective management approaches that have been proven to reduce morbidity and mortality are available to as many people with asthma as possible.

At a conceptual level, a number of myths regarding the nature of chronic diseases exist that have significantly affected their recognition and prioritization by the international health community. These include the perception of chronic diseases as problems of affluent, aging communities that have acquired them by indulging in risk factors for disease.[1] With regards to asthma, this could include the effects of tobacco use, diets richer in fats, sugar, and salt, and physical inactivity. The causes of human behavior are complex, however, and include environmental and economic pressures especially with increasing urbanization in low-income and middle-income countries that results in a lack of resources to pursue healthy choices.[1, 3]

Following this increasing awareness of the international burden of chronic diseases is the realization of the need for a new global-health partnership to address the challenge.[1, 3, 4, 28, 29] The management of global health issues is no longer the sole preserve of nation states, or even international agencies controlled by nation states, but also the preserve of the private sector, nongovernment organizations, and civil society.[3] The details of how this understanding fits into current priorities for the resources of governments and stakeholders in global health and current structures and alliances are vague.

With regards to asthma per se, the 2003 GINA report specifically identified a number of barriers.[6] These include the following:

1. A low public health priority due to the importance of other respiratory illnesses, such as tuberculosis and pneumonia, and the paucity of data on morbidity and mortality from asthma.
2. The organization of health care services. Inherent barriers may exist in terms of a country's geography, the division between public and private care, the tendency of care to be "acute" rather than "routine," shortages of trained health workers and education and training systems. In most developing countries, decades of neglect and insufficient investment have weakened health systems.
3. The lack of symptom-based rather than disease-based approaches to the management of respiratory diseases including asthma. Unsustainable generalizations across cultures and health care systems may make management guidelines developed in high-income countries difficult to implement in low-income countries.
4. The limited availability and use of medications, including omission of basic medications from WHO or national essential drug lists, poor supply and distribution infrastructure, cost, cultural attitudes toward drug delivery systems (for example, inhalers), and overuse or ineffective use.
5. Environmental barriers, including outdoor and indoor air pollution, tobacco smoking, occupational exposures, and global warming.
6. Patient barriers, including cultural factors, lack of information, underuse of self-management, overreliance on acute care, and use of alternative unproven therapies.

URBANIZATION AS A FACTOR IN ASTHMA PREVALENCE

There is no agreed definition on what constitutes an urban environment, with some countries regarding any town with a population of more than 200 as urban,

whereas others set a minimum population of 20,000.[30, 31] Other important factors to consider are population density, infrastructure, availability of health care, and the widely different patterns of social groupings and movements of people.[31] Urbanization reflects a process of social and economic development and the adoption of a modern lifestyle.[32] It has a profound effect on people's living conditions and health status, and rapid urbanization in many low- and middle-income countries has severely strained the capacity of local and national governments to provide basic services.

Between 2007 and 2050, the world population is expected to increase from 6.7 billion to 9.2 billion. Most of this expected population growth will continue to be concentrated in the cities and towns of the less developed regions of the world. As a result of rapid urbanization in the last few decades, in 2008, the world population reached a landmark: for the first time in history, the urban population equalled the rural population and, since then, the world population has been urban in its majority.[33] More developed regions are expected to see their level of urbanization rise from 74 percent in 2007 to 86 percent in 2050. In less developed regions, the proportion of urban dwellers likely will increase from 44 percent to 67 percent over the same period.[33] This compares to a mere 3 percent living in urban areas in 1800.[31] With the projected increase in the proportion of the world's population that is urban, a marked increase in the number of asthmatics worldwide is likely over the next two decades. It is estimated that an additional 100 million people will develop asthma by 2025.[6]

Urbanization has been linked to asthma risk in numerous studies in many parts of the world,[32, 34, 35] although different causes may be responsible for the observed differences between urban and rural areas. Specific factors that contribute to this risk are not clear but may include a "concrete/asphalt environment," changes in diet and physical activity levels, exposures to industrial and motor vehicle pollution, exposures to indoor irritants and allergens, and access to health care, including vaccines and antibiotics. The hygiene hypothesis has been invoked to provide a putative mechanism, namely, that natural exposure to noninfectious microbes, particularly in soil, has been reduced dramatically by living in environments covered with asphalt and concrete as well as the use of modern cleaning practices. In many inner cities, however, good hygiene certainly is not a prominent feature, suggesting that this is not the only determinant. The relationship with poverty is similarly complex.[32]

In high-income, Westernized countries, recent concerns about the health consequences of the urban way of life have resulted in increasing efforts to design cities and communities that promote and support the pursuit of physical and mental health—for example, encouraging walking and cycling, tackling pollution, increasing awareness of inequalities among ethnic groups, and promoting individual responsibility for health.[31] In developing countries, however, the scale and rapidity of the urbanization process often establish conditions of political and administrative instability along with lack of resources, which make direct response to the health consequences the only option.[31] Regarding the specific case of asthma, ongoing research into risk factors associated with living in an urban environment and protective factors associated with rural residence may provide important answers to our understanding of the disease.

CLIMATE CHANGE

It is now widely accepted that climate change is occurring at an accelerating pace as a consequence of human activities, particularly fossil fuel combustion and greenhouse gas emissions from energy supply, transport, agriculture, industry, forestry, waste, and commercial and residential buildings.[36, 37] Because of the long period during which greenhouse gases already released into the atmosphere persist, even if emissions were abruptly reduced to zero, global warming would continue throughout the 21st century and likely persist for hundreds of years.[38] Regardless of whether the world experiences gradual warming or abrupt change, human health likely will be affected adversely by accelerating climate change. Public health officials anticipate more injury, death, and disease related to natural disasters and heat waves; higher rates of food-, water-, and vector-borne infections; and more air pollution–related morbidity and mortality. These impacts not only will affect those with existing respiratory disease but also could affect the incidence and thus the prevalence of respiratory conditions. Health effects will vary regionally depending on latitude, altitude, rainfall and storms, land use patterns, urbanization, transportation, and energy production.

Data on the influence of the weather on asthma are poor and subject to debate. Weather conditions can affect asthma directly, acting on airways, or indirectly, influencing airborne allergens and pollutant levels.[39] Results of the effects of cold air on asthma consistently show an aggravating effect, but the role of humidity, wind, and rainfall is still unclear, possibly because of the complicating effect on the diffusion of pollen and pollutants. In light of current knowledge, air pollution (both ozone and particle) and aeroallergens are the critical environmental factors that are influenced by climate change that may contribute to an increased burden of allergic respiratory disease. Relevant adverse effects of climate change include an earlier start to and increased duration and intensity of the pollen season, increased heat-related ground-level ozone pollution, increased ambient air pollution from natural and anthropogenic sources, and air pollution related to wildfires and increased thunderstorms and extreme rain. A possible positive effect would occur if there was reduced susceptibility to upper respiratory infections resulting from an increase in winter temperature.[38, 39]

Climate change also will affect the burden of disease through secondary large-scale population movements, which is likely to intensify as changing climate leads to the abandonment of flooded or arid environments. Droughts, especially in rural areas, have a tendency to affect migration into cities, increasing urbanization and stressing the socioeconomic conditions exacerbated by high population growth.[36, 37] Increasingly, strategies are needed to reduce the impact of established climate change (adaptation) and to develop responses to reduce the world's dependence on carbon (mitigation).[36, 37]

ACCESS TO MEDICATION

In developing countries, health care systems focus on communicable diseases and injuries, and the infrastructure for the diagnosis and management of asthma may be either poorly available or seen as a low priority. In the 21st century, there are many competing health care priorities, and it may be difficult to convince governments that chronic respiratory diseases, such as asthma, represent a priority in the health care system.

A major obstacle to patient access for asthma treatment in developing countries is the high cost of essential asthma medications.[27] A study in 1998 identified that inhaled beclomethasone (an inhaled corticosteroid) was found to be consistently available in only four out of the eight countries surveyed, and the costs varied more than fivefold. In general, the highest prices were observed in the poorest countries.[40] Other studies that have addressed several important chronic conditions such as asthma [41] similarly have shown large variations in drug prices. A secondary analysis of the data collected from the WHO/Health Action International surveys reported the real transaction prices for 15 essential medications. Results described the gaps between the policies and practices aimed at lowering prices and the actual improvement for access to essential medicines in the developing world.[42] Although medicines may be subsidized in the public sector, availability is often poor and patients still may need to purchase medicines from the private sector, which frequently are unaffordable or unavailable.

ACTIONS REQUIRED TO REDUCE THE GLOBAL BURDEN OF ASTHMA

Despite the transformation in the global health landscape, such that noncommunicable "chronic" diseases are now (and increasingly will be) the predominant health issue facing the world this century, a global response to this situation is likely to be slow and involve more than one initiative. Perception among policy makers fail to see that chronic diseases have become diseases of poor people in most settings,[4] meaning that the gap is widening between the reality of the chronic disease burden worldwide and the response of national governments, civil society, and international agencies.[1] Global health funding overwhelmingly favors infectious disease and the Millennium Development Goals,[3, 43] although it is increasingly recognized that "global health is no longer the house that WHO built,"[3] with the emergence of global health initiatives.[29] Also, there is the danger that the effectiveness of treatment may reduce the burden of the disease to such an extent that funding, the impetus to conduct research, and primary prevention might lessen.[44]

In May 2008, the WHO launched the 2008–2013 Action Plan for the Global Strategy for the Prevention and Control of Non-Communicable Diseases.[45] The objectives include raising the priority accorded to noncommunicable disease, integrating prevention and control into policies across all government departments, and promoting international partnerships. The main shared modifiable risk factors are identified as tobacco use, unhealthy diets, physical inactivity, and harmful use of alcohol. Asthma is one such chronic disease to which these criteria apply. In 1989 the GINA was initiated with the U.S. National Heart, Lung, and Blood Institute, the National Institutes of Health, and the WHO to raise awareness among public health and government officials, health care workers, and the general public that the prevalence of asthma is increasing. The GINA recommended a management program based on the best available scientific evidence to allow doctors to provide effective medical care for asthma tailored to local health care systems and resources.[10]

Working in collaboration with leaders in asthma care from many countries, GINA sponsors World Asthma Day, which has been successful in bringing awareness about the burden of asthma to local health care officials, and to implement programs of effective asthma care. Beginning in 2003, the theme of World Asthma

Day has been the "Global Burden of Asthma." For this event, the GINA commissioned a report to summarize available data on the topic. The report highlighted the following 10 priority actions that are required to reduce the burden of asthma:[6]

1. Recognize asthma as an important cause of morbidity, economic cost, and mortality worldwide.
2. Measure and monitor the prevalence of asthma, and the morbidity and mortality due to asthma throughout the world.
3. Identify and address the economic and political factors that limit the availability of health care.
4. Improve accessibility to essential drugs for the management of asthma in low- and middle-income countries.
5. Identify and address the environmental factors, including indoor and outdoor pollution, that affect respiratory morbidity, including that which is due to asthma.
6. Promote and implement anti-tobacco public health policies to reduce tobacco consumption.
7. Adopt international asthma guidelines for developing countries to ensure that they are practical and realistic in terms of different health care systems. This includes dissemination strategies for their implementation.
8. Integrate asthma guidelines with other global respiratory guidelines for children and adults. In this respect, there is a requirement to merge the key elements of the different respiratory guidelines into an algorithm for use at the first point of entry of a respiratory patient's contact with health services.
9. Promote cost-effective management approaches that have been proven to reduce morbidity and mortality, thereby ensuring optimal treatment is available to as many persons as possible with asthma worldwide.
10. Research the causation of asthma, primary and secondary intervention strategies, and management programs, including those for use in developing countries.

CURRENT INITIATIVES

In the last decade, a number of initiatives have developed. The first of these, a WHO-led strategy known as the Practical Approach to Lung Health (PAL), was developed in 1998. This is a syndromic approach to the management of patients who attend primary health care services for respiratory symptoms. The PAL strategy is aimed at multipurpose health workers, managers in primary health care settings, and managers of tuberculosis-control programs and other respiratory disease-control programs in low- and middle-income countries. The PAL focuses on four priority respiratory diseases in patients age 5 years or more: tuberculosis, acute respiratory infections especially pneumonia, asthma, and chronic obstructive pulmonary disease. The PAL approach provides a framework for both diagnosis and management within the limitations of local resource, but to be relevant and effective, it also needs to be adapted to local conditions. The most recent report identifies 31 countries worldwide at different stages of the PAL development process.[46] Most of the funding for the PAL support at the global level comes from the U.S. Agency for International Development.

A second initiative, also led by the WHO and created in recognition of the impact that chronic respiratory disorders have on the global community, is the

Global Alliance against Chronic Respiratory Diseases (GARD). This is a voluntary alliance of 70 national and international organizations, institutions, and agencies working toward a common vision to improve global lung health according to the local needs. The GARD is dedicated to an integrated approach to chronic respiratory diseases that looks at synergies and proposes a stepwise and integrated program of prevention and control of chronic respiratory diseases, including respiratory allergies, taking into account comorbidities. The GARD focuses specifically on the needs of developing countries and deprived populations.[47]

A third initiative launched under the auspices of the International Union Against Tuberculosis and Lung Disease (IUATLD) is a guideline for the management of asthma in low-income countries first published in 1996 and updated in 2008.[48] This initiative uses standard formats and simple, affordable tools and includes important instructions for the follow-up and monitoring of patients. A report in 2002 identified obstacles in the use of the IUATLD guidelines, including poor availability of equipment, poor access by the population to health services, lack of health insurance, and limited access to drugs.

The fourth initiative is to ensure that patients are treated with appropriate drugs. The IUATLD has established an Asthma Drug Facility (ADF), which is a mechanism to ensure, for qualifying programs, the supply of quality drugs to treat asthma at the lowest possible price. The ADF works with other international agencies that are interested in asthma management, including WHO.[46] To increase affordability and accessibility of asthma medicines in as many countries as possible, the ADF has recommended two essential medicines: (1) beclomethasone 100 mcg/puff, 200 doses, HFA inhaler; and (2) salbutamol 100 mcg/puff, 200 doses, HFA inhaler. The ADF use pooled procurement along with other purchasing and supply strategies to obtain the lowest possible prices. Accompanied by the implementation of standard asthma management, the increased affordability should bring rapid and significant health and cost benefits for patients, their communities, and governments.

CONCLUSION

It is a given that the global burden of chronic diseases including asthma will continue to rise in the foreseeable future. There is an overwhelming need for global action in chronic disease prevention, including developing health care systems that can deal with this epidemic. The WHO has led the way to date with initiatives in chronic disease management, although global health increasingly is being transformed by the entry of private foundations and global funds that operate outside treaty processes and international agencies. Dealing with the fragmentation, duplication, and lack of coordination of global health efforts has been identified as one of the biggest challenges of global health. The ability of health systems to respond effectively to the direct and indirect health effects of climate change also has been identified as a key challenge, especially in many low- and middle-income countries that suffer from disorganized, inefficient, and underresourced health systems.

Although we lack a complete understanding of the factors that cause asthma, cost-effective management approaches that have been proven to reduce morbidity and mortality are available. Until novel public health and pharmacological measures are available to reduce the prevalence of asthma, the priority must be to ensure that

these management approaches are available worldwide to as many people with asthma as possible.

REFERENCES

1. Strong K, Mathers C, Leeder S, Beaglehole R. Preventing chronic diseases: how many lives can we save? *Lancet*. 2005;366:1578–1582.
2. United Nations. Millennium goals. http://www.un.org/milleniumgoals/pdf/ The%20Millenium%20Goals%20Report%202008.pdf. Accessed September 3, 2009.
3. Magnusson R. Rethinking global health challenges: towards a "global compact" for reducing the burden of chronic disease. *Public Health*. 2009;123:265–274.
4. Abegunde D, Mathers C, Adam T, Ortegon M, Strong K. The burden and costs of chronic diseases in low-income and middle-income countries. *Lancet*. 2007; 370:1929–1938.
5. World Health Organization. *Global Alliance against Chronic Respiratory Diseases, General Meeting Report from Seoul, Republic of Korea, 1–2 June 2007*. Geneva, Switzerland: World Health Organization; 2007. http://www.who.int/gard/publications/MtgReport_GMSeoul_Final.pdf. Accessed September 8, 2009.
6. Masoli M, Fabian D, Holt S, Beasley R. Global burden of asthma report. http://www.ginasthma.com/ReportItem.asp?l1=2&l2=2&intId=94. Published 2003. Accessed September 7, 2009.
7. Holgate S. Pathogenesis of asthma. *Clin Exp Allergy*. 2008;38:872–897.
8. Wenzel S. Asthma: defining of the persistent adult phenotypes. *Lancet*. 2006;368: 804–813.
9. Halder P, Pavord I, Shaw D, et al. Cluster analysis and clinical asthma phenotypes. *Am J Respir Crit Care Med*. 2008;178:218–224.
10. Global Initiative for Asthma. Homepage. http://www.ginasthma.com. Accessed September 8, 2009.
11. Pearce N, Douwes J, Beasley R. The rise and rise of asthma: a new paradigm for the new millennium? *J Epidemiol Biostat*. 2000;5:5–16.
12. The International Study of Asthma and Allergies in Childhood (ISAAC) Steering Committee. Worldwide variation in the prevalence of symptoms of asthma, allergic rhinoconjunctivits, and atopic eczema: ISAAC. *Lancet* 1998;351:1225–1232.
13. Pearce N, Ait-Khaled N, Beasley R, et al. Worldwide trends in the prevalence of asthma symptoms: phase three of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax*. 2007;62:758–766.
14. Lai C, Beasley R, Crane J, et al. Global variation in the prevalence and severity of asthma symptoms: phase three of the International Study of Asthma and Allergies in Childhood (ISAAC). *Thorax*. 2009;64:476–483
15. Jansen C, Anto J, Burney P, et al. The European Community Respiratory Health Survey: what are the main results so far? *Eur Respir J*. 2001;18:598–611.
16. ISAAC Steering Committee. Worldwide variations in the prevalence of asthma symptoms: the international study of asthma and allergies in childhood: ISAAC. *Eur Respir J*. 1998;12:315–335.
17. Leung R, Ho, P. Asthma, allergy, and atopy in three south-east Asian populations. *Thorax*. 1994;49:1205–1210.
18. Lai C. Common inhalant allergens in southern China: Proceedings of the 5th West Pacific Allergy Symposium; June 11–14, 1997; Seoul, Korea.
19. Beasley R, Crane J, Lai C, Pearce N. Prevalence and etiology of asthma. *J Allergy Clin Immunol*. 2000;105:S466–S472.

20. Beasley R, Clayton T, Crane J, et al. Association between paracetamol use in infancy and childhood, and risk of asthma, rhinoconjunctivitis, and eczema in children aged 6–7 years: analysis from phase three of the ISAAC program. *Lancet*. 2008;372:1039–1048.
21. Romagnani S. The increased prevalence of allergy and the hygiene hypothesis: missing immune deviation, reduced immune suppression, or both? *Immunology*. 2004;112:352–363.
22. Schaub B, Lauener R, von Mutius E. The many faces of the hygiene hypothesis. *J Allergy Clin Immunol*. 2006;117:969–977.
23. Wijesinghe M, Weatherall M, Perrin K, Crane J, Beasley R. International trends in asthma mortality rates in the 5- to 34-year age-group. *Chest*. 2009;135:1–5.
24. Beasley R, Pearce N, Crane J. Worldwide trends in asthma mortality during the twentieth century. In: Albert Sheffer, ed. *Fatal Asthma*. New York: Marcel Dekker Inc.; 1998.
25. Suissa S, Ernst P, Benayoun S, et al. Low-dose inhaled corticosteroids and the prevention of death from asthma. *N Engl J Med*. 2000;343:332–336.
26. Lanes S, Rodrigues G, Huerta C. Respiratory medications and the risk of asthma death. *Thorax*. 2002;57:683–686.
27. Ait Khaled N, Enarson A, Bissell K, Billo N. Access to inhaled corticosteroids is key to improving quality of care for asthma in developing countries. *Allergy*. 2007;62:230–236.
28. Banatvala N, Donaldson L. Chronic diseases in developing countries. *Lancet*. 2007; 370:2076–2078.
29. World Health Organization Maximizing Positive Synergies Collaborative Group. An assessment of interactions between global health initiatives and country health systems. *Lancet*. 2009;373:2137–2169.
30. Nicolau N, Siddique N, Custovic A. Allergic disease in urban and rural populations: increasing prevalence with increasing urbanization. *Allergy*. 2005;60:1357–1360.
31. Godfrey R, Julien M. Urbanisation and health. *Clin* Med. 2005;5:137–141.
32. Cooper P, Rodrigues L, Cruz A, Barreto M. Asthma in Latin America: a public health challenge and research opportunity. *Allergy*. 2009;64:5–17
33. United Nations. Department of Economic and Social Affairs homepage. http://www.un.org/desa/population/publications/wup2007/2007WUP_Highlights_-web.pdf. Accessed September 7, 2009.
34. Viinanen A, Munhbayarlah S, Zevgee T, et al. Prevalence of asthma, allergic rhinoconjunctivitis and allergic sensitization in Mongolia. *Allergy*. 2005;60:1370–1377.
35. Weinberg E. Urbanization and childhood asthma: an African perspective. *J Allergy Clin Immunol*. 2000;105:224–231.
36. Costello A, Abbas M, Allen A, et al. Managing the health effects of climate change: Lancet and University College London Institute for Global Health Commission. *Lancet*. 2009;373:1693–1733.
37. Ayres J, Forsberg B, Annesi-Maesano I, Environment and Health Committee of the European Respiratory Society, et al. Climate change and respiratory disease: European Respiratory Society position statement. *Eur Respir J*. 2009;34:295–302.
38. Shea K, Truckner R, Weber R, Peden D. Climate change and allergic disease. *J Allergy Clin Immunol*. 2008;122:443–453.
39. D'Amato G, Cecchi L. Effects of climate change on environmental factors in respiratory allergic diseases. *Clin Exp Allergy*. 2008;38:1264–1274.
40. Ait-Khaled N, Auregan G, Bencharif N, et al. Affordability of inhaled corticosteroids as a potential barrier to treatment of asthma in some developing countries. *Int J Tuberc Lung Dis*. 2000;4:268–271.

41. Cameron A, Ewen M, Ross-Degnan D, Ball D, Laing R. Medicine prices, availability, and affordability in 36 developing and middle-income countries: a secondary analysis. *Lancet*. 2009;373:240–249.
42. Vasan A, Kim J. Essential medicines pricing—reform needed. *Lancet*. 2009;373: 191–193.
43. Stuckler S, King L, McKee M. WHO's budgetary allocations and burden of disease: a comparative analysis. *Lancet*. 2008;372:1563–1569.
44. Partridge M. Has ISAAC told us as much as it can? Where now? *Thorax*. 2009; 64:462–463.
45. World Health Organization. *2008–2013 Action Plan for the Global Strategy for the Prevention and Control of Noncommunicable Diseases*. Geneva, Switzerland: World Health Organization; 2008. http://www.who.int/nmh/publications/ncd_action_plan_en.pdf. Accessed September 8, 2009.
46. World Health Organization. Practical approach to lung health. http://www.who.int/tb/publications/pal/en/index.html. Accessed June 9, 2009.
47. Bousquet J, Dahl R, Khaltaev N. Global Alliance against chronic respiratory diseases. *Allergy*. 2007;62:216–233.
48. International Union against Tuberculosis and Lung Disease. Management of asthma: a guide to the essentials of good clinical practice. http://www.theunion.org/asthma-division/asthma.html. Accessed September 8, 2009.

CHAPTER 5

The Challenge of Managing Hypertension

Bernard M. Y. Cheung, MD, PhD, FRCP, and
Kwok Leung Ong, MPhil

INTRODUCTION

"Blood pressure" is the force of blood pushing against the walls of the arteries as the heart pumps out blood. It is needed to deliver cardiac output through the vasculature to the periphery. If this pressure rises and stays high over time, it can damage the body in many ways. High blood pressure (HBP) or hypertension, a serious condition that can lead to coronary heart disease, heart failure, stroke, kidney failure, and other health problems, occurs when the blood pressure (systemic arterial pressure) is consistently elevated (see table 5.1).[1] Since the risk of stroke and myocardial infarction is directly related to the level of blood pressure, reducing the blood pressure is an effective way of reducing the risk of these complications.[2–4] Reduction of blood pressure to optimal levels is important in retarding the deterioration of renal function. Even moderate elevation of arterial blood pressure can lead to shortened life expectancy.

Hypertension is seldom *curable*; the more practical aim is to *control* the blood pressure. Virtually all guidelines on the management of hypertension emphasize the importance of good blood pressure control, and target blood pressures are set. Conventionally, a systolic pressure of 120 mmHg and a diastolic pressure of 80 mmHg are considered ideal. In individuals with renal impairment or diabetes mellitus, the target blood pressure is less than 130/80 mmHg. Blood pressure control is of paramount importance especially among diabetics to prevent or mitigate macrovascular complications of diabetes.[5] The focus of this chapter is to discuss the epidemiology of hypertension and to stress that it is a global disease with potentially serious consequences.

TYPES OF HIGH BLOOD PRESSURE

The word "hypertension" can be classified as either essential (primary) or secondary. Essential or primary hypertension, the most common type of hypertension, means that no medical cause can be found to explain the raised blood pressure.

Table 5.1
WHO/ISH Classification of Blood Pressure

Category	Systolic blood pressure (mmHg)	Diastolic blood pressure (mmHg)
Blood pressure		
Optimal	<120	<80
Normal	<130	<85
High-normal	130–39	85–89
Hypertension		
Grade 1 (mild)	140–159	90–99
Subgroup: borderline	140–149	90–94
Grade 2 (moderate)	160–179	100–109
Grade 3 (severe)	≥180	≥110
Isolated systolic hypertension	≥140	<90
Subgroup: borderline	140–149	<90

Source: Whitworth JA, World Health Organization, International Society of Hypertension Writing Group. 2003 World Health Organization/International Society of Hypertension statement on management of hypertension. *J Hypertens.* 2003;21:1983-1992.
Note: ISH = International Society of Hypertension; WHO = World Health Organization.

About 90 to 95 percent of hypertension is classified as essential hypertension. Secondary hypertension indicates that the HBP is a result of (that is, secondary to) another condition, such as kidney disease or tumors (adrenal adenoma, for example). Most of the mechanisms associated with secondary hypertension generally are understood. Those associated with essential or primary hypertension are far less understood, however. What is known is that cardiac output is raised early in the disease course, with total peripheral resistance (TPR) normal. Over time, cardiac output drops to normal levels, but TPR is increased.

While a normal blood pressure reading of 120/80 mmHg is desirable, it is important to be aware of "borderline hypertension" or prehypertension, defined as mildly elevated blood pressure in which systolic blood pressure is between 120 mmHg and 139 mmHg and diastolic blood pressure is between 80 mmHg and 89 mmHg. Those with borderline hypertension may have a tendency to develop more sustained or higher elevations of blood pressure as they get older. They have a modestly increased risk of developing heart-related (cardiovascular) and stroke-related (cerebrovascular) disease. Therefore, even if the hypertension does not appear to be elevated initially, those with pressures in the high-normal range should take steps to bring their blood pressure down or, at least, to prevent it from rising higher. Certainly those with borderline hypertension should have continuing follow-up work of their blood pressure and monitoring for the complications of hypertension. The goal in treating hypertension is to have a value of less than 140/90 mmHg. For purposes of this chapter, hypertension will be defined as blood pressure greater than or equal to 140/90 mmHg.

Despite the recognition of the dangers of untreated or uncontrolled hypertension, be it primary, secondary, or borderline, the condition is still widely and persistently

undertreated around the world. Millions of people with hypertension remain undiagnosed, thus risking serious damage to major organs of the body. Of those who have been diagnosed, many are not receiving treatment or are being undertreated; that is, their blood pressure is not well controlled.

LOOKING AT THE NUMBERS

Hypertension is exceedingly common, not just in the developed world, but also is prevalent in many developing nations. Indeed, hypertension is not necessarily a disease of the wealthy. In fact, in 2000, one-third of the people with hypertension in the world lived in economically developed countries whereas two-thirds lived in developing countries. In the 21st century, the prevalence of hypertension in adults worldwide is estimated to be 26.4 percent of the population (based on 2000 statistics). By 2025, the prevalence of hypertension is projected to rise to 29.2 percent.[6] Because of population growth, increased longevity, and aging of the population, the total number of adults with hypertension is estimated to increase from 972 million in 2000 to 1.56 billion in 2025, a rise of about 60 percent.

In the United States, HBP is a growing and serious problem. Approximately 30 percent of the population is estimated to have HBP; in raw numbers, hypertension is estimated to affect 65 million adults. More than half of the U.S. elderly population is estimated to have HBP or prehypertension.[7, 8] We analyzed the trends in the prevalence, awareness, treatment, and control of hypertension using the National Health and Nutrition Examination Survey (NHANES) 1999–2004 database. The NHANES database has been valuable for the study of the trends in the health status of a population because of its large sample size, complex sampling design, good quality control, and comprehensive content.[9] Blood pressure information on 14,653 individuals (4,749 in 1999–2000, 5,032 in 2001–2002, and 4,872 in 2003–2004) age 18 and over was analyzed. The following is based on findings from the NHANES study.

Multivariate analysis showed that increasing age, high body mass index (BMI), being non-Hispanic black, and being less educated were significantly associated with an elevated risk of hypertension. In 2003–2004, 75.7 percent of people with hypertension were aware of their diagnosis, 65.1 percent were treated, and 36.8 percent had blood pressure controlled (56.6 percent among those treated) (see table 5.2). Data also show that the age-adjusted awareness and treatment rates had not increased significantly since 1999–2000. The control rate among treated hypertensive people also had not increased significantly. The blood pressure control rate among all people with hypertension increased significantly, however, from 29.2 percent in 1999–2000 to 36.8 percent in 2003–2004 ($p = 0.02$, and after age adjustment, $p = 0.006$). The increase in age-adjusted blood pressure control rate in this six-year period was 8.1 percent (95 percent confidence interval [CI]: 2.4 to 13.8) in absolute terms.

Among the 60 and over age-group, the awareness, treatment, and control rates of hypertension had all increased significantly ($p \leq 0.01$). Factors such as race and ethnicity, gender, and age are known risk factors for hypertension. Between 1999 and 2004, the control rate among those with hypertension increased in both men and women ($p = 0.03$ and 0.05, respectively), and improved significantly in non-Hispanic blacks and Mexican Americans ($p = 0.02$ and 0.03, respectively). In Mexican American men, the control rate increased from 8.7 percent in 1999–2000 to

Table 5.2
Awareness, Treatment, and Control among Participants with Hypertension in the U.S. Population, 1999–2004

Outcome Measures	Percentage, % (SE)		
	1999–2000 (n = 1,530)	2001–2002 (n = 1,500)	2003–2004 (n = 1,614)
Awareness	68.7 (2.0)	70.7 (1.7)	75.7 (2.1)
Treatment	58.2 (2.9)	60.1 (1.6)	65.1 (2.4)
Control			
Among all with hypertension	29.2 (2.3)	32.5 (1.6)	36.8 (2.3)[a]
Among those treated	50.2 (2.2)	54.0 (1.6)	56.6 (2.4)
Among treated hypertensive diabetic subjects	24.1 (3.6)	35.6 (3.9)	37.5 (4.0)

Source: Centers for Disease Control and Prevention, National Center for Health Statistics. National health and nutrition examination survey. http://www.cdc.gov/nchs/nhanes.htm. Accessed September 7, 2009.

Notes: Data are weighted to the U.S. census population.

a. $P<0.01$ for the difference between 1999–2000 and 2003–2004 after adjusting for age.

31.1 percent in 2003–2004 (p = 0.002), which was mainly because of the age-adjusted proportion on treatment increasing from 28.9 percent to 49.8 percent (p = 0.04). Although the overall prevalence of hypertension has not increased significantly between 1999 and 2004, the control rate of hypertension has been increasing, especially in Mexican American men, elderly, and obese people. The improvement in blood pressure control in the United States is encouraging, but much more remains to be done; similar findings in blood pressure improvement also have been reported by other groups.[10–12]

A large part of the problem in diagnosing and treating hypertension is that approximately one-third of those with HBP are unaware of their condition. In general, awareness and control rates of hypertension tend to be better in older people and people with higher BMI, whereas among hypertensive people of younger age and lower BMI there is a greater likelihood of undertreatment. This may be because older and obese people tend to go to the doctor more frequently compared with younger and healthier individuals. Overall, however, the treatment rate among hypertensive patients is only 53.7 percent, which means that approximately half of those with hypertension are not being treated. Clearly much more needs to be done to detect and then treat those with this disease.

In the United States, one of the goals of Healthy People 2010 is to reduce the proportion of adults with HBP to 16 percent and the proportion of adults with HBP whose blood pressure is under control to 50 percent.[13] Perhaps this is unrealistic, but with a concerted effort from health professionals and government agencies, the target of 50 percent blood pressure control rate could be achieved. Recent clinical trials showed that a HBP control rate of 66 to 71 percent could be achieved in some settings and environments.[14–16] To achieve these goals, however, more effective publicity about the disease, improved education campaigns, more concerted efforts

from health professionals, and more effective implementation of clinical guidelines on the management of hypertension are needed.[2]

HYPERTENSION WORLDWIDE

Although hypertension is a common disease worldwide, its prevalence and awareness as a medical problem from country to country varies considerably.[17–21] (see table 5.3). Overall, males have a higher prevalence of hypertension (68.9 percent) compared with women (72.5 percent).[17] In terms of awareness and treatment rates, much work needs to be done to inform and educate individuals about this disease. Awareness rates of hypertension vary from 25.2 percent in South Korea to 75 percent in Barbados. Treatment rate varies from 10.7 percent in Mexico to 66 percent in Barbados. Blood pressure control rate ranges from a dismal 2.3 percent in Mexico to a better (but not great) rate of 38 percent in Barbados.[18] Overall, however, the awareness, treatment, and control rates in developing countries are comparable to those in developed countries.[20]

The situation is not totally bleak. Some progress has been made in controlling blood pressure around the world. In Cuba, for example, a country with limited resources but an excellent health care system, blood pressure control has been achieved through investment in the training of health professionals, an availability of locally manufactured drugs, and a more aggressive policy at the primary-care level.[22] Improvement in blood pressure control has been observed in England [19]; Taiwan, China [21]; Chile [23]; and Finland.[24] In a study of 38 populations from 21 countries included in the World Health Organization's (WHO's) MONICA (Monitoring Trends and Determinants in Cardiovascular Disease) study, data showed that mean blood pressures in both men and women decreased from the

Table 5.3
Worldwide Prevalence and Control of Hypertension by Sex

			Prevalence (%)		Control (%)	
Region	Year	Reference	Men	Women	Men	Women
Barbados	1996	25	25.4	29.8	38.0 (Total)	
China	2002	26	20	17	5% (Total)	
England	2006	27	32	29	24	32
France	1997–1998	28	27.3	14.7	7.4	21.4
Korea, Rep. of	2001	28	31.5	23.7	7.7	14.0
Malaysia	1996	28	31.9	33.9	6.0	6.0
Mexico	1992–1993	25	37.5	28.1	1.1	3.5
Poland	2002	28	29.0	29.0	10.0	14.0
Portugal	2003	28	49.5	38.9	7.2	15.4
Seychelles	2004	28	44.0	36.0	12.0	30.0
Taiwan, China	2002	29	27.1	20.2	21.0	28.5
Turkey	2003	28	27.5	36.1	8.2	8.0
United States	2003–2004	6	30.7	28.2	33.3	35.2

Source: Kearney PM, et al. Worldwide prevalence of hypertension: a systematic review. *J Hypertens.* 2004;22:11–19.

Note: Control rate is based on all hypertensive patients regardless of treatment status.

mid-1980s to mid-1990s.[25] The decrease in blood pressure was observed in people with normal as well as HBP, which suggests that factors besides the use of antihypertensive medication contributed to this populationwide decrease in blood pressure. In China, however, the blood pressure control rate remained almost unchanged from 1991 to 2002, despite an increase in the treatment rate. In fact, HBP tops the list of preventable deaths in China and is responsible for 2.3 million deaths a year.[18] Most of these deaths were caused by strokes and aneurysms followed by heart attacks, conditions that are directly associated with HBP. It appears that blood pressure has become more difficult to control in China, which might be due to (1) few Chinese being tested or treated for HBP, (2) rapid modernization and economic development leading to more stressful lives, and (3) transition to a Western Hemisphere diet.[26] Perhaps a more logical explanation is the fact that dietary sodium is high in China. This issue will be discussed later in this chapter.

RISK FACTORS AND CONTROL OF HYPERTENSION

There is no single identifiable cause of hypertension among the great majority of patients with this disease. In the young and in those with severe rapidly progressive hypertension, secondary causes should be considered, which include possible anatomical, endocrine, and renal causes. The majority of patients with HBP, however, usually have multiple factors leading to hypertension. Age is a known and common risk factor for the disease. As the arteries narrow and stiffen with age, blood has to push harder against the artery walls, contributing to hypertension. Genetic factors, as reflected in a positive family history, also are known to predispose an individual to HBP. Lifestyle factors, including obesity, diet with a high salt intake, diet high in saturated fats and low in fruits and vegetables, stress, tobacco smoking, and alcohol in excess, as well as a sedentary lifestyle, are known risk factors for the development of hypertension. A reduction in any of the above risk factors would do much to prevent further damage to one's organs and would do much to prevent premature death caused by hypertension. Preventive measures focusing on lifestyle changes could result in a decrease in the prevalence of hypertension. Maintaining normal body weight, exercising daily, and eating a healthy diet are all useful preventive measures.

Obesity and Related Lifestyle Factors

Among lifestyle factors, obesity is probably the leading risk factor for hypertension.[27, 28] Approximately 65 to 78 percent of the hypertension in young adults studied in the Framingham Offspring Study, for example, can be attributed to obesity.[29] In the United States, adults with diagnosed hypertension have a high prevalence of central obesity, especially in women (79 percent).[30] Numerous trials show that lifestyle changes, usually involving diet and regular exercise, can lead to a reduction in blood pressure.[31, 32] Although impressive results are obtained under clinical trials conditions, it remains to be seen whether these can be translated successfully to the general population and sustained over time.

Sodium Intake

The association between sodium intake and blood pressure has been demonstrated in epidemiological studies.[33] Salt loading studies [34] as well as studies

limiting salt intake [35] have demonstrated a causal association between sodium intake and blood pressure. The problem is that much of the daily sodium intake is found in processed foods; the average American consumes more than 10 grams of salt a day, but the Institute of Medicine recommends one-third of this amount as optimal.[35] As such, cooperation from the food industry, especially the fast-food industry, to reduce sodium content in their food and beverage products would be an important step in helping to reduce salt consumption in the daily diet.[36]

Other Risk Factors

Hypertension is a known risk factor for cardiovascular disease and also is a component of what is known as the metabolic syndrome. Metabolic syndrome is a name for a group of symptoms that occur together and promote the development of coronary artery disease, stroke, and type 2 diabetes. Symptoms of this syndrome include high triglycerides, insulin resistance, low HDL ("good") cholesterol, and HBP.[37] A substantial proportion of adult hypertensive men (40.9 percent) and women (52.5 percent) in the United States have multiple risk factors for cardiovascular disease, such as obesity, dyslipidemia, elevated blood glucose, and smoking. Because hypertension is only one of a number of risk factors for cardiovascular disease, it is desirable to address all these risk factors rather than to consider the level of blood pressure in isolation. Hence, the level of blood pressure, the presence of other risk factors, target organ damage, and overt cardiovascular disease allow the physician to estimate the cardiovascular risk of a patient. The urgency of treatment depends on the overall degree of cardiovascular risk as well as the level of blood pressure. Despite the knowledge of what should be done, however, population surveys in many parts of the world show that the control of blood pressure is not ideal.

MANAGEMENT OF THE HYPERTENSIVE PATIENT

Antihypertensive Medications

The major drug classes used in the treatment of hypertension include alpha-adrenergic receptor antagonists (alpha-blockers), angiotensin-converting enzyme inhibitors (ACEIs), angiotensin II receptor blockers (ARBs), beta-adrenergic receptor antagonists (beta-blockers), calcium channel blockers (CCBs), and diuretics of the Thiazide type. Other drugs classes include centrally acting agents (moxonidine), direct renin inhibitor (aliskiren), aldosterone antagonist (spironolactone), and older agents (methyldopa, minoxidil). The relative merit of different classes of antihypertensive drugs is debatable, which is why it is important to have well-designed clinical trials to assess the effectiveness of each class of drug on a large, diverse population. The primary goal is the good control of blood pressure. The selection of drug class also is influenced by comorbidities. For example, a hypertensive with angina might receive a beta-blocker or a CCB. A patient with heart failure might receive a diuretic, ACEI, or ARB. A patient with previous myocardial infarction might receive a beta-blocker or ACEI. A patient with diabetic nephropathy might receive an ACEI or ARB. A patient with prostatism might receive an alpha-blocker. Whatever medications are prescribed, they should be acceptable to and well tolerated by the patient. The efficacy and side-effect profile may vary with ethnicity.[38] ACEIs and ARBs

reduce blood pressure less well in black patients, who tend to respond better to Thiazide diuretics. ACEI-induced cough is common among those with hypertension in East Asia, so ACEIs often are changed to ARBs for these patients.[39]

Although blood pressure–lowering drugs on the market have been shown to be efficacious in lowering blood pressure, the cost-effectiveness of antihypertensive medications varies enormously because of the tremendous differences in the pricing of the drugs. Large clinical trials have shown that, in terms of cardiovascular outcome, the newer classes *are not* necessarily superior to diuretics.[40] Moreover, Thiazide diuretics are not costly, and they reduce the incidence of heart failure. Fortunately, many generic formulations of Thiazide diuretics, beta-blockers, CCBs, and ACEIs now are available. These off-patent antihypertensive drugs are relatively cheap, which means that the daily cost of antihypertensive drugs can be made affordable and should not be a barrier to treatment.

Most hypertensive patients require more than one antihypertensive drug to control their blood pressure. Indeed, the recent hypertension management guidelines recommend the initial use of a Thiazide-type diuretic for most patients with uncomplicated hypertension either alone or combined with drugs from other classes, such as ACE inhibitors, ARBs, beta-blockers, and CCBs. The 2002 Allhat study (Antihypertensive and Lipid-Lowering Treatment to Prevent Health Attack Trial), one of the biggest clinical trials conducted by the U.S. federal government, found that diuretics were a cheaper, safer, and equally effective means of treating hypertension compared with other medications.[40] More than 42,000 individuals who were 55 years or older with HBP were randomized to one of four drugs (diuretics, ACE inhibitors, CCBs, and alpha-blockers). This study showed that diuretics not only were equally effective, but also substantially cheaper. The trial was not without its critics, especially among the pharmaceutical industry, which ganged up, attacked, and discredited the study findings. The industry has done little to promote generic diuretics, preferring instead to market their higher priced antihypertensive drugs. Clouding the issue was the publication of an Australian study that was released two months after Allhat, which found that an ACE inhibitor was superior to a diuretic in controlling blood pressure. Despite the findings from the Allhat study, the prescribing of the newer hypertension drugs has increased faster than the use of diuretics.

It is especially important for those individuals with comorbid conditions to keep their blood pressure under control. For example, evidence shows that there is a benefit associated with prescribing an ACE inhibitor for hypertensive patients with diabetes, congestive heart failure, and previous heart attack.[41] The readiness to use multiple drugs may explain the improvement in overall blood pressure control rates from 1999 to 2004. Medication, however, should not be the sole means of addressing the problem. While the prescribing of pharmaceuticals has done much to better control hypertension, lifestyle interventions and modifications should be strongly encouraged for all hypertensive patients.[42]

PRACTICE GUIDELINES

Numerous consensus guidelines are available on the treatment of hypertension. The major international guidelines include the seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood

Pressure (JNC 7),[2] the WHO International Society of Hypertension Writing Group's recommendations,[43] and the European Societies of Hypertension and Cardiology guidelines for the management of arterial hypertension.[44] Not all of the guidelines and recommendations are in uniform agreement. For example, Thiazide diuretics are recommended as the first-line therapy in JNC 7 but not in the WHO or European guidelines. Although these guidelines differ in fine details, there are many areas showing strong agreement. That is, all guidelines agree that blood pressure should be measured carefully and repeatedly before the diagnosis of hypertension can be made. Target organ damage should be identified and, if present, there is a greater urgency for drug treatment to control blood pressure. Recent guidelines emphasize the importance of an overall assessment of a patient's cardiovascular risk factors, identification of cardiovascular complications, and addressing the modifiable risk factors rather than treating blood pressure alone.

IMPROVING COMPLIANCE

Diagnosis of hypertension is not sufficient if the blood pressure is not well controlled. One of the reasons for poor control is poor compliance (concordance).[45] Compliance is poor for many reasons, including (1) the asymptomatic nature of hypertension, which leads individuals to believe incorrectly that they are healthy; (2) the need to take medications daily on a long-term basis; (3) the side effects of medications, which may put off people from taking their pills; and (4) the lack of prescription drug coverage, which may make the costs of treatment and follow-up a barrier to getting needed care. Among the elderly, compliance may be compromised should the patient not fully understand instructions or forget whether medications have been taken as prescribed. Addressing compliance is an important aspect of achieving good blood pressure control. This may require candid discussion and education by the health care provider with the patient about the benefits and risks of drug treatment. The use of once-daily medications and combination pills, and changing the medications in response to side effects, may aid compliance.

ALTERNATIVE MEDICINE

Complementary and alternative medicine treatments are popular among patients with chronic diseases, and hypertension is no exception. Vegetables, such as garlic, onions, carrots, and celery; supplements, such as calcium and vitamin C; and herbs, such as hawthorn and Ginkgo biloba have been taken in the belief that they help to reduce blood pressure.[46, 47] Some patients rely on acupuncture, qigong, relaxation therapy, meditation, and yoga to lower their blood pressure. Relaxation techniques tend to lower blood pressure and may account for part of the beneficial effects of yoga and qigong.[48, 49] These approaches generally have few if any adverse effects, but the effect size is modest compared with drug therapy. More research using robust clinical trial methodology is required before firm recommendations can be given.

CHALLENGES TO BE CONSIDERED

Mild to moderate hypertension is generally an asymptomatic disease. It aptly has been called the "silent killer" because it usually produces no symptoms and

increases gradually and slowly over the years. People with HBP usually have no idea that they have this problem, and they do not go to the doctor specifically because of elevated blood pressure. The detection and treatment of hypertension is thus a major public health challenge. This disease is a common condition worldwide, although its prevalence varies from country to country.

Like other chronic diseases, the risk factors for hypertension are well known—diet (high sodium), lifestyle, and obesity—which means that, in many cases, it is a disease that is potentially preventable and certainly manageable. Because it is an asymptomatic condition, population-based screening for hypertension is certainly a cost effective, noninvasive, easy-to-conduct means to identify individuals with elevated blood pressure. Elevated blood pressure is detectable easily using a sphygmomanometer, which is economical and noninvasive. Although large-scale screening is bound to generate a large amount of false positives, perhaps with this disease, it is better to accept a higher level of false positives rather than fail to work up and diagnose an individual with HBP. Once a diagnosis of hypertension is made, numerous treatment modalities can be implemented easily. That being said, detection of hypertension unfortunately remains unsatisfactory in most parts of the world. Millions of individuals, like ticking time bombs, are walking around with undiagnosed and untreated HBP.

REFERENCES

1. 1999 World Health Organization/International Society of Hypertension Guidelines for the management of hypertension. *J Hypertens*. 1999;17:151–183.
2. Chobanian AV, et al. Seventh report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure. *Hypertension*. 2003;42:1206–1252.
3. Staessen JA, Wang JG, Thijs L. Cardiovascular prevention and blood pressure reduction: a quantitative overview updated until March 1, 2003. *J Hypertens*. 2003;21:1055–1076.
4. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002;360:1903–1913.
5. UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. *BMJ*. 1998;317:703–713.
6. Kearney PM, et al. Global burden of hypertension: analysis of worldwide data. *Lancet*. 2005;365:217–223.
7. Ong KL, et al. Prevalence, awareness, treatment, and control of hypertension among United States adults 1999–2004. *Hypertension*. 2007;49:69–75.
8. Fields LE, et al. The burden of adult hypertension in the United States 1999 to 2000: a rising tide. *Hypertension*. 2004;44:398–404.
9. Centers for Disease Control and Prevention, National Center for Health Statistics. National health and nutrition examination survey. http://www.cdc.gov/nchs/nhanes.htm. Accessed September 7, 2009.
10. Suh DC, et al. Trends in blood pressure control and treatment among type 2 diabetes with comorbid hypertension in the United States: 1988–2004. *J Hypertens*. 2009;27:1908–1916.

11. Plantinga LC, Miller ER III, Stevens LA, et al. Blood pressure control among persons without and with chronic kidney disease: US trends and risk factors 1999–2006. *Hypertension*. 2009;54:47–56.
12. Cutler JA, et al. Trends in hypertension prevalence, awareness, treatment, and control rates in United States adults between 1988–1994 and 1999–2004. *Hypertension*. 2008;52:818–827.
13. US Department of Health and Human Services. *Healthy People 2010*. conference ed. 2 vols. Washington, DC: US Department of Health and Human Services; 2000. http://www.health.gov/healthypeople. Accessed September 7, 2009.
14. Law MR, Wald NJ, Morris JK, Jordan RE. Value of low dose combination treatment with blood pressure lowering drugs: analysis of 354 randomised trials. *BMJ*. 2003;326:1427–1434.
15. Moser M. Rationale for combination therapy in the management of hypertension. *J Clin Hypertens*. 2003;5:17–25.
16. Julius S, et al. Outcomes in hypertensive patients at high cardiovascular risk treated with regimens based on valsartan or amlodipine: the VALUE randomised trial. *Lancet*. 2004;363:2022–2031.
17. Kearney PM, et al. Worldwide prevalence of hypertension: a systematic review. *J Hypertens*. 2004;22:11–19.
18. Wu Y, et al. Prevalence, awareness, treatment, and control of hypertension in China: data from the China National Nutrition and Health Survey 2002. *Circulation*. 2008;118:2679–2686.
19. Falaschetti E, et al. Continued improvement in hypertension management in England: results from the Health Survey for England 2006. *Hypertension*. 2009;53:480–486.
20. Pereira M, et al. Differences in prevalence, awareness, treatment and control of hypertension between developing and developed countries. *J Hypertens*. 2009;27:963–975.
21. Su TC, et al. Evidence for improved control of hypertension in Taiwan: 1993–2002. *J Hypertens*. 2008;26:600–606.
22. Ordunez-Garcia P, et al. Success in control of hypertension in a low-resource setting: the Cuban experience. *J Hypertens*. 2006;24:845–849.
23. Fasce E, et al. Trends in prevalence, awareness, treatment and control of hypertension in urban communities in Chile. *J Hypertens*. 2007;25:1807–1811.
24. Kastarinen MJ, et al. Trends in hypertension care in eastern and south-western Finland during 1982–2002. *J Hypertens*. 2006;24:829–836.
25. Tunstall-Pedoe H, Connaghan J, Woodward M, Tolonen H, Kuulasmaa K. Pattern of declining blood pressure across replicate population surveys of the WHO MONICA project, mid-1980s to mid-1990s, and the role of medication. *BMJ*. 2006;332:629–635.
26. Cheung BM, Thomas GN. The metabolic syndrome and vascular disease in Asia. *Cardiovasc Hematol Disord Drug Targets*. 2007;7:79–85.
27. Wang TJ, Vasan RS. Epidemiology of uncontrolled hypertension in the United States. *Circulation*. 2005;112:1651–1662.
28. Pausova Z. From big fat cells to high blood pressure: a pathway to obesity-associated hypertension. *Curr Opin Nephrol Hypertens*. 2006;15:173–178.
29. Garrison RJ, et al. Incidence and precursors of hypertension in young adults: the Framingham Offspring Study. *Prev Med*. 1987;16:235–251.
30. Ong KL, Tso AW, Lam KS, Cheung BM. Gender difference in blood pressure control and cardiovascular risk factors in Americans with diagnosed hypertension. *Hypertension*. 2008;51:1142–1148.

31. Tuomilehto J, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med*. 2001;344:1343–1350.
32. Lien LF, et al. Effects of PREMIER lifestyle modifications on participants with and without the metabolic syndrome. *Hypertension*. 2007;50:609–616.
33. Law MR, Frost CD, Wald NJ. By how much does dietary salt reduction lower blood pressure? I-Analysis of observational data among populations. *BMJ*. 1991;302:811–815.
34. MacGregor GA, et al. Double-blind study of three sodium intakes and long-term effects of sodium restriction in essential hypertension. *Lancet*. 1989;2:1244–1247.
35. Sacks FM, et al. Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. DASH-Sodium Collaborative Research Group. *N Engl J Med*. 2001;344:3–10.
36. MacGregor GA, Sever PS. Salt: overwhelming evidence but still no action: can a consensus be reached with the food industry? Consensus Action on Salt and Hypertension. *BMJ*. 1996;312:1287–1289.
37. Grundy SM, American Heart Association, National Heart, Lung, and Blood Institute, et al. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. *Circulation*. 2004;109:433–438.
38. McDowell SE, Coleman JJ, Ferner RE. Systematic review and meta-analysis of ethnic differences in risks of adverse reactions to drugs used in cardiovascular medicine. *BMJ*. 2006;332:1177–1181.
39. Woo KS, Nicholls MG. High prevalence of persistent cough with angiotensin converting enzyme inhibitors in Chinese. *Br J Clin Pharmacol*. 1995;40:141–144.
40. Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA*. 2002;288:2981–2997.
41. Gu Q, Paulose-Ram R, Dillon C, Burt V. Antihypertensive medication use among US adults with hypertension. *Circulation*. 2006;113:213–221.
42. Dickinson HO, et al. Lifestyle interventions to reduce raised blood pressure: a systematic review of randomized controlled trials. *J Hypertens*. 2006;24:215–233.
43. Whitworth JA, World Health Organization, International Society of Hypertension Writing Group. 2003 World Health Organization/International Society of Hypertension statement on management of hypertension. *J Hypertens*. 2003;21:1983–1992.
44. Mancia G, et al. 2007 guidelines for the management of arterial hypertension: the Task Force for the Management of Arterial Hypertension of the European Society of Hypertension and of the European Society of Cardiology. *J Hypertens*. 2007;25:1105–1187.
45. Heisler M, et al. When more is not better: treatment intensification among hypertensive patients with poor medication adherence. *Circulation*. 2008;117:2884–2892.
46. Ernst E. Complementary alternative medicine for hypertension: a mini-review. *Wien Med Wochenschr*. 2005;155:386–391.
47. Ried K, et al. Effect of garlic on blood pressure: a systematic review and meta-analysis. *BMC Cardiovasc Disord*. 2008;16;8:13.
48. Cheung BM, et al. Randomised controlled trial of qigong in the treatment of mild essential hypertension. *J Hum Hypertens*. 2005;19:697–704.
49. Dickinson HO, et al. Relaxation therapies for the management of primary hypertension in adults. *Cochrane Database Syst Rev*. 2008;1:CD004935.

CHAPTER 6

Perspectives on the Treatment of Osteoarthritis

Stephen Lyman, PhD, and Joseph T. Nguyen, MPH

As the U.S. population ages, and as life expectancy increases, diseases of old age are much more prevalent now than in the past. Arthritis and other rheumatic conditions are among the most common chronic diseases of adults and account for the leading cause of disability among U.S. adults.[1] Arthritis contributes substantially to disability, poor health-related quality of life, and increased direct and indirect medical costs.[2] Interventions to improve function, decrease pain, and delay disability are required.

Total joint replacements are important surgical interventions for treating severe arthritis of weight-bearing joints. For example, the most common indication for total knee replacement (TKR) is osteoarthritis (OA) of the knee joint. This chapter presents an epidemiologic overview of arthritis and focuses on TKR surgeries to illustrate the growing trend in the treatment of OA. Although TKR is an elective procedure, the quality-of-life benefit to a return to near-normal function and pain-free ambulation is substantial. As a result, demand for this procedure is projected to balloon, particularly with the baby boom generation reaching retirement age within the next year (the eldest boomers will turn 65 in 2011).

EPIDEMIOLOGY OF ARTHRITIS

Arthritis is a group of more than 100 rheumatic diseases and other conditions that can cause pain, stiffness, joint inflammation, and swelling of the joints. It is not a single disease per se, and affects individuals of any age, including children. Generally, however, arthritis is considered a disease of aging as the incidence increases with age. It is estimated that 70 million U.S. adults suffer from some form of arthritis, with an estimated 60 percent of the population over age 65 affected by arthritis.[3] Projected prevalence of self-reported arthritis in the U.S. population age 65 and over is expected to double from 21.4 million in 2005 to 41.1 million in 2030.[4] If left untreated, many types of arthritis can cause irreversible damage to the joints, bones, organs, and skin. The two most common types of arthritis are OA and rheumatoid arthritis.

Osteoarthritis

OA (a degenerative joint disease) is noninflammatory and its onset is subtle and gradual, usually involving one or a few joints, typically the knees, hips, hands, and the spine. The condition results from wear and tear as pressure of gravity causes physical damage to the joints and surrounding tissues leading to pain, tenderness, swelling, and decreased function. *Rheumatoid arthritis* (RA) is an autoimmune disease that occurs when the body's immune system attacks the synovium (cell lining inside the joint). RA is a chronic, usually disabling disease that is characterized by joint pain, stiffness, swelling, and loss of joint function. OA will be the focus of this chapter.

OA is estimated to affect 27 million individuals in the United States.[5] Projections indicate that by 2030, 25 percent of the adult U.S. population (or 67 million individuals) will have physician-diagnosed arthritis. Some of the known unmodifiable risk factors include age and female sex. The association of OA with the female sex is thought to be the result of differences in both the lower leg alignment of females and the hormonal changes associated with menopause reducing bone quality and making joints more susceptible to deformation, which can lead to OA.[6]

Modifiable risk factors include obesity, a primary risk factor for knee OA, with a threefold increased risk of knee OA among those with a body mass index of more than 30 compared with those with a normal body mass index of less than 25.[7, 8] For many years, it was not clear whether being overweight preceded OA or was a consequence of the disease, given the immobility and disability of the disease. Recent studies, however, have shown that being overweight antedates the development of OA.[9] Once OA progresses, being overweight accelerates the progression of the disease, making weight loss an imperative component of any OA prevention effort.

Reduction in body weight can slow the progression of knee OA and also can result in a decrease in pain and improved function and mobility. In the Framingham Study, for example, women who lost an average of 11 pounds decreased the risk for knee OA by 50 percent.[10] Weight loss interventions may be the single most efficacious method of prevention of knee OA and may provide many other health benefits; however, given the difficulty in successful weight loss and maintenance, this strategy may not be particularly effective for most people with OA.[10]

Exercise offers clear cardiovascular and pulmonary benefits, but those who engage in regular moderate to vigorous physical activity or sports are more prone to suffering a knee injury. An increased demand on the knee joint may increase the likelihood of degenerative changes, apart from an acute injury leading to these changes. Although a study of runners found no increased risk of knee OA,[11] other sports requiring jumping and other lateral movements may contribute to increased risk of knee injuries and subsequent OA. That being said, physical activity also can slow the progression of knee OA by providing weight management, strengthening the muscles around the knee joint, maintaining bone density, and maintaining knee joint lubrication.

Having a previous knee injury (that is, meniscus tears, ligament tears, and focal chondral defects) substantially increases the risk of developing knee OA.[12, 13] Meniscus and ligamentous injuries result in instability of the knee joint, which can result in higher forces or loads being placed on the articular cartilage and can accelerate the natural wear-and-tear process that leads to OA.

Ethnic and racial differences are apparent, which may be a result of disparities in access to and use of medical care. Minorities are less likely to seek treatment and, when they do, they usually present with more pain, less function, and reduced quality of life compared with whites.[14]

ECONOMIC COSTS OF OA

The costs of OA vary widely. Xie et al. found that direct costs varied tenfold among studies conducted in the United States.[15] Explanations include different geographic regions studied, different databases used, and perhaps most important, different items included in the cost estimates. Kotlarx et al. conducted multivariate analyses of the relationships between OA and annual health care expenditures borne by patients and insurers.[16] They found that OA was found to contribute substantially to health care expenditures. Among women, OA increased out-of-pocket expenditures by almost $1,400 per year and insurer expenditures by $4,800. Among men, out-of-pocket expenditures were increased by almost $700 per year and insurer expenditures by $4,000. OA raised aggregate annual medical care expenditures by $185.5 billion. Clearly, the economic burden of this disease is substantial.

Hawker et al. studied the economic burden of OA from the perspective of individuals living with the condition. They found that costs incurred were mainly for time lost from employment and leisure as well as for unpaid caregivers.[17] Costs increased with worsening health status and greater OA severity. Manette et al. also found that the direct and indirect costs attributable to OA were substantial, with productivity-related costs being predominant.[18]

TREATMENT

Treatments to manage the disease are designed to control pain and improve function and quality of life while limiting adverse events. Nonpharmacological and pharmacological means to manage OA are available. Exercises as well as medications such as nonsteroidal anti-inflammatory drugs (NSAIDs) often are prescribed in treating the symptoms and adverse effects of OA. Because of the risk of gastrointestinal risk, including ulcer complications, acetaminophen for pain relief is often the first choice of anti-inflammatory agents, even though NSAIDs are considered to be more effective in managing pain. Other medical treatments include short-acting intra-articular injections and symptom-modifying agents such as analgesics. Some patients become refractory to the stronger NSAIDs; other more potent pain relief agents, including opioids and other narcotics, to treat OA-related pain, including Cox-2 inhibitors, have been prescribed. Although the Cox-2 inhibitors such as Vioxx and Celbrex showed a great deal of effectiveness for the suppression of pain, cardiovascular complications led to the removal of these agents from the market.[19]

Corticosteroids and hyaluronic acid injections have been used to provide temporary pain relief in patients with mild to moderate OA. While both types of injections have been shown to provide more pain relief than placebo alone,[20] effectiveness was short term, usually restricted from weeks to a few months. Follow-up injections were given, but concerns about the long-term effects and diminishing effectiveness contraindicates their use indefinitely.[21] These injections may provide pain relief,

but they do not treat the underlying disease. Without treating the underlying cartilage degeneration of OA, progression will not be halted.

Nutraceuticals, or nutritional supplements, have been touted to provide symptomatic relief of OA. Glucosamine sulphate has been evaluated in several placebo-controlled randomized trials and consistently has been found to slow knee OA progression.[22, 23] Randomized trials of glucosamine and chondroitin have shown similar effects of delaying the radiological progression of OA of the knee after daily administration.[24] Neither requires Food and Drug Administration (FDA) approval; both are popular over-the-counter supplements among those with OA.

The breakdown of cartilage is the primary indicator of the progression of OA. The degenerative process may be halted, and perhaps even reversed, using a variety of innovative medical interventions. The most promising disease-modifying OA drugs focus on altering the inflammatory and tissue remodeling processes as they relate to the articular cartilage synovium and subchondral bone. Early trials have been largely unsuccessful in identifying specific agents that will lead to an alteration of the disease process, however. The most promising treatment to date is use of a metalloprotease (MMP) inhibitor such as doxycycline, which has shown less disease progression in OA patients over a 30-month treatment course.[25] Further study is needed to evaluate the safety and efficacy of this treatment modality before FDA approval should be given. Another avenue of potential preventative or therapeutic intervention for knee OA is the use of biologic materials, the most promising of which is the use of stem cells. Much more research needs to be done in this area before widespread adoption of this modality.

For the more advanced cases affecting major joints such as the hips or knees, surgery is usually indicated. Surgical procedures to delay progression of OA, such as anterior cruciate ligament reconstruction, meniscectomy, and meniscal repair, have not been shown to be effective. Once OA has progressed to an advanced state, knee replacement is considered the only effective treatment currently available. This is an elective procedure, but the quality of life benefit to a return to near normal function and pain-free ambulation is substantial.

Total joint replacements are important surgical interventions for treating severe arthritis of weight-bearing joints. The most common indication for TKR is OA of the knee joint. Advances in surgical techniques have made TKR a safe and effective treatment modality for knee OA. Indeed, TKR has been one of the major breakthroughs in orthopedic surgery. Figure 6.1 shows the trends in the number of TKRs performed over the past decades. Given the aging of the population and the burden of obesity in younger individuals, it is estimated that in the not distant future, approximately 3 million Americans each year will undergo TKR.[26] Age-specific trend data (not shown) also indicate that individuals are not waiting until they are Medicare-eligible (age 65 and older) to have one or both knees replaced; rather, younger individuals are electing to have surgery for quality-of-life reasons. Data show a declining average age of TKR recipients, many of whom are still in the workforce with employer-based private health insurance coverage. One study focusing on trends in TKR in Wisconsin, for example, found that the age-adjusted rate for TKR increased the most among the youngest age-group (45–49 years)![27] This trend may reflect an increased prevalence of OA, which in turn may be related directly to the huge increase in the number of individuals who are overweight and obese.

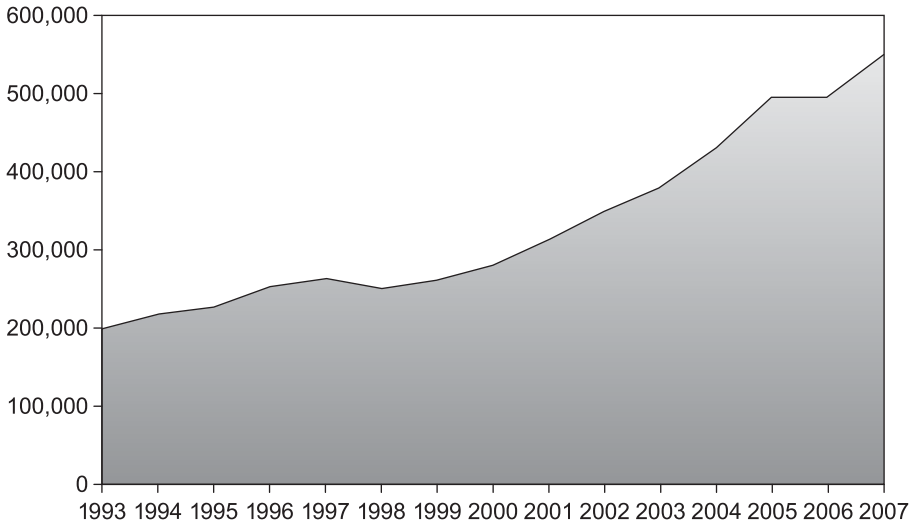

Figure 6.1 Total Knee Replacements by Year
Source: Adapted by authors.

Although TKR is a popular and accepted orthopedic procedure for OA, there are short-term risks and complications, particularly among those with comorbid, chronic conditions. TKR has been shown to improve the quality of life for individuals affected with severe OA of the knee joints. The costs, however, are substantial. Surgical intervention is effective, but it would be prudent to intervene before surgery is necessary by trying to delay and prevent the onset of OA as well as to improve symptoms of OA through nonsurgical methods. Studies have shown that weight loss and physical activity can make a difference not only in alleviating the symptoms of OA but also in delaying the progression of degenerative joint disease, especially among overweight individuals.[28] Both could have a significant effect on reducing the burden of arthritis and thus on the increase in the number of TKRs being performed.

CHALLENGES FOR THE FUTURE

Patients present with symptoms of OA at different stages of their disease course, and with a wide range of OA pathology. A young individual, for example, with a meniscus injury or a ligament rupture may present with OA at a very early stage. These patients often have normal-looking cartilage but more than likely have begun the cascade toward OA especially if no intervention is taken. On the other hand, an older patient with no acute injury history might present with advanced diffuse knee OA that has graduated from a long insidious preclinical stage and now has begun to affect function. It can be argued that both of these patients require treatment to halt the development or progression of their knee OA, but identifying the treatment methods that would be most effective for each is difficult.

OA is a multifactorial disease; with a wide array of promoters and antagonists operating on the cartilage, synovium, and subchondral bone, making it unlikely that there will be one magic bullet or one treatment. OA is not a systemic disease, but rather a localized disease phenomenon. As such, efforts need to be made to identify

those pharmaceutical treatments that would prove to be most efficacious. And, efforts need to be undertaken to address the obesity epidemic in America, which data show contributes significantly to the burden of OA and TKR in the United States.

REFERENCES

1. Centers for Disease Control and Prevention. Prevalence of disabilities and associated health conditions among adults: United States, 1999. *MMWR*. 2001;50:120–125.
2. Centers for Disease Control and Prevention. Trends in aging: United States and worldwide. *MMWR*. 2003;52:101–106.
3. Centers for Disease Control and Prevention. Prevalence of self-reported arthritis or chronic joint symptoms among adults: United States, 2001. *MMWR*. 2002;51: 948–950.
4. Centers for Disease Control and Prevention. Projected prevalence of self-reported arthritis or chronic joint symptoms among persons aged over 65 years: United States, 2005–2030. *MMWR*. 2003;52:489–451.
5. Helmick CG, et al. for the National Arthritis Data Workgroup. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: part 1. *Arthritis Rheum*. 2008;58:15–25.
6. Sirkanth VK, et al. A meta-analysis of sex differences prevalence, incidence, and severity of osteoarthritis. *Osteorathritis Cartilage*. 2005;13:769–781.
7. Hart DJ, et al. The relationship between osteoarthritis and osteoporosis in the general population: the Chingford Study. *Ann Rheum Dis*. 1994;53:158–162.
8. Vrezas I, Elsner G, Bolm-Audorff U, Abolmaali N, Seidler A. Case-control study of knee osteoarthritis and lifestyle factors considering their interaction with physical workload. *Int Arch Occup Environ Health*. 2010;83(3):291–300.
9. Felson DT, et al. Osteoarthritis: new insights. Part 1: the disease and its risk factors. *Ann Intern Med*. 2000;133(8):635–639.
10. Felson DT, el al. Weight loss reduces the risk for symptomatic knee osteoarthritis in women: the Framingham study. *Ann Intern Med*. 1992;6:535–539.
11. Chakravarty EF, et al. Long distance running and knee osteoarthritis: a prospective study. *Am J Prev Med*. 2008;35:133–138.
12. Lohmander LS, et al. The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. *Am J Sports Med*. 2007;35:1756–1769.
13. Oiestad BE, et al. Knee osteoarthritis after anterior cruciate ligament injury: a systematic review. *Am J Sports Med*. 2009;37:1434–1443.
14. Lavernia CJ, et al. Race, ethnicity, insurance coverage, and preoperative status of hip and knee surgical patients. *J Arthroplasty*. 2004;19:978–985.
15. Xie F, Thumbo J, Li SC. True differences or something else? Problems in cost of osteoarthritis studies. *Semin Arthritis Rheum*. 2007;37:127–132.
16. Kotlarx H, et al. Insurer and out-of-pocket costs of osteoarthritis in the US. *Arthritis Rhuem*. 2009;60:3546–3553.
17. Hawker GS, et al. The economic burden of disabling hip and knee osteoarthritis from the perspective of individuals living with this condition. *Rheum*. 2005;44 (12):1531–1537.
18. Manette RV, et al. Direct and indirect costs attributable to osteoarthritis in active subjects. *J Rheum*. 2006;33(6):1152–1158.
19. Avouac J, Gossec L, Dougados M. Efficacy and safety of opioids for osteoarthritis: a meta-analysis of randomized controlled trials. *Osteoarthritis Cartilage*. 2007;15: 957–965.

20. Raynauld JP, et al. Safety and efficacy of long-term intra-articular steroid injections in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled trail. *Arthritis Rheum*. 2003;48:370–377.
21. Neustadt DH. Intra-articular injections for osteoarthritis of the knee. *Cleve Clin J Med*. 2006;73:897–898, 906–911.
22. Pavelka K, et al. Glucosamine sulfate use and delay of progression of knee osteoarthritis: a 3-year, randomized, placebo-controlled, double-blind study. *Arch Intern Med*. 2002;162:2113–2123.
23. McAlindon T. Glucoamine for osteoarthritis: dawn of a new era? *Lancet*. 2001;357(9252):247–248.
24. Sawitzke AD, et al. The effect of glucosamine and/or chondroitin sulfate on the progression of knee osteoarthritis: a report from the glucosamine/chondroitin arthritis intervention trial. *Arthritis Rheum*. 2008;58:3183–3191.
25. Brandt KD, et al. Effects of doxycycline on progression of osteoarthritis: results of a randomized, placebo-controlled, double-blind trial. *Arthritis Rheum*. 2005;52:2015–2025.
26. American Academy of Orthopaedic Surgeons. Total knee replacement. http://orthoinfo.aaso.org/topic.cfm?topic+A00389.
27. Mehrotra C, et al. Trends in total knee replacement surgeries and implications for public health, 1990–2000. *Public Health Rep*. 2005;120:278–282.
28. Rejeski WJ, et al. Obese, older adults with knee osteoarthritis: weight loss, exercise, and quality of life. *Health Psychol*. 2002;21:419–426.

This page intentionally left blank

SECTION 2

TREATMENT AND PREVENTION OF CANCERS

This page intentionally left blank

CHAPTER 7

Screening and Early Detection of Disease: Focus on Colon Cancer

Lyndal Trevena, MBBS (Hons), MPhilPH, PhD

INTRODUCTION

The issue of whether an individual benefits from early detection of disease is not as straightforward as it may seem. Intuitively, it might seem obvious that early detection would be beneficial and that early intervention would successfully change the course of the disease progression. In reality, this assumption is open to debate. For some diseases, a preclinical phase may be so short that the disease is not likely to be detected by screening. Or, even if detected, many options are not available to effect a cure. Also, not every preclinical case progresses to clinical disease. Most men at time of death, for example, have cancerous prostate cells, which were not detected and did not cause harm. Not all diseases in a preclinical phase progress to a clinical phase.

Many questions need to be answered before a decision can be made about the wisdom of screening for disease. It is first important to understand that screening tests are performed on *healthy* individuals. They are asymptomatic—that is, they have not yet developed symptoms of illness. Also, not every disease lends itself to screening. Can the disease be detected early in the first place? For some diseases, such as lung cancer and ovarian cancer, screening tests are not yet available to detect these respective cancers at an early and more treatable stage. Furthermore, no screening test is 100 percent perfect. False positives (test is positive but there is no disease) must be followed up with additional testing to rule in or rule out disease. The medical, psychological, and economic consequences of false positive test results can be substantial. Will the screening test be acceptable to individuals—that is, how invasive is the test? How costly is the test? And, most important, if an individual tests positive, are treatments available? There is little advantage to screening if individuals who are screened derive no benefit. Should everyone be screened for the particular disease or should just high-risk people be targeted for screening? For example, sickle cell anemia is much more prevalent among African Americans. Should all Americans be screened for this disease or would it be more economically prudent to screen only those at high risk?

Early detection of disease implies that a disease can be detected at an earlier, treatable stage. The assumption is that intervention or treatment is available and can be more effectively applied if the disease is detected at an earlier stage, thus making a difference in outcome—that is, a reduction in mortality. It also implies that a biologic point exists in the natural history of a disease before which treatment would benefit an individual more than if the disease was detected at a later stage. Using cervical cancer as an example, early detection of this cancer followed by effective intervention has been shown to be hugely beneficial in reducing the progression of carcinoma in situ to invasive cervical cancer and thus mortality.

In 1968, Wilson and Jungner published their seminal paper on the principles of screening amid a great deal of controversy surrounding the early detection of disease. These principles have been refined and debated many times over the past 40 years and include the following:[1]

- The screening program should respond to a recognized need.
- The objectives of screening should be defined at the outset.
- A target population should be defined.
- Screening program effectiveness should be supported by scientific evidence.
- The program should integrate education, testing, clinical services, and program management.
- Quality should be assured, with mechanisms to minimize potential risks of screening.
- The program should ensure informed choice, confidentiality, and respect for autonomy.
- The program should promote equity and access to screening for the entire target population.
- Program evaluation should be planned from the outset.
- The overall benefits of screening should outweigh the harm.

DEFINITIONS OF SCREENING

What then do international experts mean when they use the term "screening"? Wilson and Jungner originally suggested that the term refers to "medical investigation, which does not arise from a patient's request for advice for specific complaints."[2] The UK's Screening Committee recently has determined that "[s]creening is a process of identifying apparently healthy people who may be at increased risk of a disease or condition. They can then be offered information, further tests and appropriate treatment to reduce their risk and/or any complications arising from the disease or condition."[3] Regardless of the wording within definitions, one of the fundamental concepts of screening is that it involves testing or assessing asymptomatic, healthy people. As discussed later in this chapter, this definition is crucial to the debate about whether screening, in some instances, might do more harm than good.

CANCER SCREENING AS A STRATEGY FOR PREVENTION

Much of the debate in Wilson and Jungner's day was about screening for inborn errors of metabolism. New tests for the early detection of cancer quickly emerged, however, primarily because of advances in medical technology that enable those with cancer, especially those in the early stages of cancer, to receive treatment (surgical, radiation, pharmaceuticals). Many countries, for example, now have well-organized

screening programs and policies for cervical, breast, and colon cancer. Cervical intraepithelial neoplasia detected by a Papanicolaou (Pap) smear effectively can identify those women with cervical cancer or precancer; mammography is widely used for breast cancer detection; and fecal occult blood testing (FOBT) or colonoscopy is widely used to detect colon cancer. But, it is important to acknowledge that there are controversies associated with screening testing.

One of the key principles embedded within the definitions of screening is that the potential benefits of screening should outweigh the harms of testing. Given that most people in the general population *will not* have cancer (or the disease of interest), it is important that any benefits from detecting and treating disease early through screening are weighed against the harms of follow-up testing in healthy people who have a false positive result (see appendix 7. A for a definition of terms used in screening). Because population-based mass screening programs often involve testing and follow-up treatment for millions of asymptomatic people in the community, the physical, psychological, and economic burden of invasive follow-up testing needs to be compared with the number of lives saved through screening. Simply detecting disease at an earlier point in time without an effective treatment to reduce disease-specific mortality can result in *lead-time bias.* In other words, you are not prolonging life expectancy, but simply are increasing the amount of time the patient lives with the diagnosis. Another common pitfall of screening is that it may detect disease that is not clinically significant, thereby leading to overdiagnosis and overtreatment in some people.

Screening can be implemented in a number of different ways. This chapter speaks only of *population screening,* which are programs that are large, organized, and systematically implemented. Such programs define a target a group deemed to be at sufficient risk of disease (often defined by age, gender, or family history). They also may have systems for recall and reminders, as well as registers. Quality assurance systems need to be in place to quantify regular reporting and evaluation against population-level health outcomes. In some instances, however, screening is implemented in the form of *case finding.* Tests or inquiries are performed at the time a patient may present for other reasons. They often are instigated proactively by the clinician. Examples might include performing a chlamydia test on a sexually active woman under the age of 25, or taking a blood pressure measurement on an otherwise healthy adult.

With this background in mind, this chapter takes a closer look at colon cancer screening as an example, and illustrates many of these issues through an analysis of current policy and practice.

THE GLOBAL BURDEN OF CANCER

Cancer causes approximately one in every eight deaths (12.6 percent) worldwide, following cardiovascular disease (29 percent) and infectious/parasitic diseases (16.2 percent) as a major cause of mortality.[4, 5] In high-income countries, lung, colorectal, breast, and stomach cancers account for 4 of the top 10 causes of death,[4] whereas infectious/parasitic and perinatal causes continue to dominate mortality in low-income nations. Cancer generally is more common in developed countries, with the exception of cervical cancer, which is most common in Africa, India, and Southeast Asia. Lung cancer continues to be the most common cause of cancer death worldwide, followed by colorectal cancer and stomach cancer.[4] As

communicable disease control improves, the burden attributable to noncommunicable diseases will increase and projections estimate that cancer in 2030 will be the most common cause of death worldwide.[4]

On a more positive note, cancer mortality rates have fallen in recent decades, but not to the same extent as cardiovascular deaths.[6] The decline is influenced mainly by decreases in the number of deaths of breast, prostate, and colorectal cancer cancers thought to be due to early detection and treatment. Deaths due to lung cancer have declined in men but increased in women due to changes in tobacco consumption.

COLORECTAL CANCER: PRELIMINARY FACTS

Adenocarcinoma accounts for more than 95 percent all colorectal cancers, usually developing within large adenomatous polyps that become neoplastic over time. These cancers are the most likely to be detectable and treated by screening, with leiomyosarcoma, lymphoma, melanoma, and neuro-endocrine cancers of the colon occurring in only 2 to 3 percent of cases. Survival following colorectal cancer diagnosis is highest for people diagnosed with localized disease, for which the five-year survival rate is a robust 87.2 percent. Once the disease spreads regionally, the five-year survival rate falls to 67 percent and drops to just 12.1 percent for those with distant metastases. Advancing age; a history of colorectal polyps or previous colorectal, endometrial, or ovarian cancer; or a family history of colorectal cancer are known to increase the risk of colorectal cancer. The risk imparted by a family history of colorectal cancer depends on the age of onset of relatives affected with colorectal cancer, the number of affected relatives, and whether or not there is polyposis or a family history of additional cancers that may indicate the presence of a familial cancer syndrome. For those affected with colorectal cancer, approximately 20 to 30 percent have a family history of colorectal cancer.[7] An inherited contribution to colorectal cancer is likely for approximately 15 percent of people with colorectal cancer.

Most people (approximately 95 percent in most countries) are at *average risk* of developing colorectal cancer through having either no family history or the presence of just one first-degree relative who had colorectal cancer diagnosed at age 55 years or older or one second-degree relative diagnosed at any age. For this group, in many countries, screening with FOBT is recommended at least every two years from age 50.[7] *Moderate risk* can be defined as having either a first-degree relative affected under age 55 years, two first-degree relatives, or one first-degree and one second-degree relative affected at any age (on one side of the family), without the presence of high-risk features such as those that characterize high-risk people. Screening colonoscopy is often advised every five years (provided no adenomatous polyps were present) either from age 50 or from 10 years earlier than the youngest diagnosis of colorectal cancer in the family (whichever is first). Flexible sigmoidoscopy plus double contrast barium enema or computed tomography (CT) colonography (a new imaging technique for the large bowel requiring bowel preparation like colonoscopy and insufflation of the bowel) may be offered if colonoscopy is contraindicated for some reason. Guidelines also may advise offering FOBT in the intervening years.[8]

Potentially high risk is often defined as follows: (1) three or more first-degree relatives or a combination of first- or second-degree relatives on the same side of the family; (2) two or more first- or second-degree relatives affected on the same side of the family *plus* any of the following high-risk features: one person with multiple

Table 7.1
Stages of Colorectal Cancer—National Cancer Institute

Stage 0: Carcinoma in situ.
Stage I (Dukes A): Carcinoma contained within the wall of the colon.
Stage II (Dukes B): Stage IIA—carcinoma extends beyond wall of colon and possibly within adjacent tissue. Stage IIB—carcinoma extends into nearby organs or through the peritoneum.
Stage III (Duke's C): Stage IIIA—intramural carcinoma and involvement of up to three lymph nodes. Stage IIIB—spread to as many as three nearby lymph nodes and has spread beyond the middle tissue layers of the colon wall; or to nearby tissues around the colon or rectum; or beyond the colon wall into nearby organs and/or through the peritoneum. Stage IIIC—cancer has spread to four or more nearby lymph nodes and has spread to or beyond the middle tissue layers of the colon wall; or to nearby tissues around the colon or rectum; or to nearby organs or through the peritoneum.
Stage IV (Duke's D): Spread to nearby lymph nodes and has spread to other parts of the body, such as the liver or lungs.

primary colorectal cancers or colorectal cancer diagnosed before age 50; or (3) people at risk of or confirmed to have a known inherited predisposition to colorectal cancer, such as Lynch Syndrome or Familial Adenomatous Polyposis (FAP), or they are part of a family in which a gene mutation predisposing to these conditions has been identified. Familial Adenomatous Polyposis describes an inherited medical syndrome that puts affected individuals at enormous risk for the early development of colorectal cancer unless they take preventive action. FAP occurs when a person is born with a mutation in a specific gene called the adenomatous polyposis coli (APC) gene. FAP is an autosomal dominant inherited disorder, characterized by the development of hundreds of colonic adenomatous polyps. If left untreated, approximately 75 percent of people with FAP will develop colorectal cancer by the age of 40.[9] Several studies have demonstrated the efficacy of colonoscopy screening in reducing the incidence of, and mortality from, colorectal cancer in these groups.[10–12] Lynch Syndrome and FAP are thought to cause between 1 and 4 percent of colorectal cancer cases and to occur in less than 1 percent of the population.[13, 14] Screening for Lynch Syndrome usually occurs along with one to two yearly colonoscopies from age 25 or five years younger than the earliest diagnosis in the family. Annual screening for FAP is advised from 10 to 14 years. Features of the family history of colorectal cancer that indicate possible Lynch Syndrome are multiple affected relatives, especially in the presence of people diagnosed young (under age 50), or relatives with multiple primary colorectal cancers.[15]

Once diagnosed, treatment of colorectal cancer depends on the type of cancer, stage of disease, the location of the cancer, and available resources (see table 7.1). Surgery may involve local resection or colectomy with or without ileostomy. Radiation options include external radiotherapy internal therapy using radioactive needles, wires, and catheters. Chemotherapeutic options may include systemic or regional regimens and more recently the use of chemo-embolization.

THE GLOBAL BURDEN OF COLORECTAL CANCER

Worldwide in 2007, there were approximately 1.2 million cases and 630,000 deaths due to colorectal cancer, which accounts for approximately 8 percent of all

cancer deaths. Colorectal cancer affects both men and women and its incidence increases with age. Differences in population age distributions (for high- compared with low-income countries) are a likely reason for this cancer type to be more common in developed nations. It is more common in Europe, North America, Australia, New Zealand, and parts of South America, and notably among males in China.[16] Although colorectal cancer mortality rates per 100,000 population are similar in the United States and China, in absolute terms, the total number of colorectal cancer deaths per year is far greater in China due to the size of its population. A reduction in colorectal cancer globally needs to take into account these overall numbers (not just rates), if substantial gains are to be made in colorectal cancer control. In the United States, statistics show a reduction of 3.4 percent in colorectal cancer deaths between 2001 and 2006. This compares with a 1.7 percent reduction in colorectal cancer deaths between 1984 and 2001 and, in part, may be due to some colorectal cancer screening in the latter period and due to new cancer treatments.

STRATEGIES FOR REDUCING THE GLOBAL BURDEN OF COLORECTAL CANCER: PRIMARY PREVENTION

Primary prevention aims to prevent the development of disease at the outset. Examples include immunization and lifestyle changes. Unlike some other cancers, however, relatively little is known about the biological mechanism and pathogenesis of colorectal cancer, making the implementation of effective primary prevention strategies difficult. Most scientists agree that an interaction between gene and environment is likely. The World Health Organization estimates that 13 percent of colorectal cancer deaths worldwide are attributable to three modifiable risk factors: (1) overweight and obesity, which account for 11 percent; (2) physical inactivity, which accounts for 15 percent; and (3) low fruit and vegetable intake, which accounts for 2 percent of colorectal cancer deaths.[6] Downstream reduction on cancer incidence is likely to be greatest at the primary level of prevention and, not incidentally, the cost-effectiveness of such strategies is often highest. It therefore makes sense to target any primary prevention strategies for colorectal cancer prevention toward maintaining a healthy weight, exercising regularly, and eating the recommended amount of fruit and vegetables. Some other factors have been suggested for targeted primary prevention, but these are not as well substantiated. Following is a summary of the current evidence for primary prevention strategies for colorectal cancer.

Established primary prevention strategies include the following:

- **Increased physical activity.** An inverse relationship has been suggested between physical activity and colorectal cancer recently quantified in a meta-analysis of 52 studies. The authors found that colorectal cancer incidence could be reduced by 24 percent through increased physical activity. Further research is needed to clarify the type, intensity, and duration of physical activity that will give the greatest benefit.[17]
- **High-fiber, low-fat diet.** A number of trials have attempted to document the effect of high fiber and low fat on colorectal cancer incidence, but they have been hampered by methodological limitations such as low dose, inadequate duration of follow-up, and low power.[18] Since such dietary changes are unlikely to cause any harm, they often

are recommended, but there is insufficient evidence at this time to warrant large-scale public health strategies to promote such an approach to colorectal cancer prevention.

Emerging or unsubstantiated primary prevention strategies include the following:

- **Calcium supplementation.** Two well-conducted randomized placebo-controlled trials on 1,246 subjects over three to four years showed a moderate protective effect for dietary supplementation of at least 1,200 mg elemental calcium daily for the reduction of adenomatous polyps (Odds Ratio [OR] 0.74; Confidence Interval [CI] 0.58, 0.95). The study was not able to show any effect on colorectal cancer itself.[19]
- **Vitamin D.** There is some evidence to indicate that vitamin D (particularly in combination with calcium) can have a protective effect against colorectal cancer. Some confusion arose with the publication of the Women's Health Initiative study, which appeared to conflict with this emerging body of evidence (see below). Subsequent analysis suggests that estrogen may interact with vitamin D and reduce its effect on colorectal cancer.[20] Further work on this issue is ongoing.
- **Antioxidant supplementation.** A systematic review of 20 randomized controlled trials showed that beta-carotene, vitamin A, vitamin C, vitamin E, or their combinations taken as supplements did not reduce colorectal cancer incidence. In fact, there is some evidence to indicate that supplementation increased all-cause mortality (14 percent intervention versus 11 percent placebo), but the reason for this result is unclear.[21]
- **Green tea.** A review of 52 studies of more than 1.6 million participants showed insufficient and conflicting evidence about the effect of three to five cups of green tea per day on the incidence of gastrointestinal cancer.[22]
- **Postmenopausal hormone therapy.** The Women's Health Initiative trial showed a 44 percent reduction in colorectal cancer with the use of combined estrogen-progestin replacement therapy after menopause but not in the estrogen-only group. Additionally, however, results showed a 26 percent increase in invasive breast cancer, a 29 percent increase in coronary heart disease, a 41 percent increase in stroke, and a twofold increase in thromboembolic events. Because the harms outweigh the benefits, this is not recommended as a strategy for colorectal cancer risk reduction.[23]
- **Aspirin or nonsteroidal anti-inflammatory drugs (NSAIDs).** There is fair to good evidence suggesting that aspirin and NSAIDs reduce the incidence of adenomatous polyps, but they must be taken in high doses and over a long period of time (10 years or more). The U.S. Preventive Services Task Force (USPSTF) recommends against their use for the prevention of colorectal cancer in people of average risk, however, because evidence indicates that aspirin and NSAIDs in these doses can cause gastrointestinal bleeding, renal impairment and other complications. Thus, the harms outweigh the benefits.[24]

STRATEGIES FOR REDUCING THE GLOBAL BURDEN OF COLORECTAL CANCER: SECONDARY PREVENTION THROUGH SCREENING AND EARLY DIAGNOSIS

Secondary prevention screening options for colorectal cancer include the following:

- **Fecal occult blood testing.** Biennial FOBT has been used as the preferred screening strategy for colorectal cancer in several countries (Australia, United Kingdom, and parts of Canada). Results from three large randomized clinical trials show that biennial FOBT reduces colorectal cancer mortality by approximately 16 percent.[25] The sensitivity

and specificity of the FOBT test is such that it is likely that some individuals will test positive even though they do not have cancer (false positives) and will have to undergo an unnecessary colonoscopy. Those with higher risk of colorectal cancer (for example, males and those with a family history) are more likely to benefit from screening than those who are at less risk. Although more colorectal cancer deaths probably would be averted in the higher risk groups, the risk of false positives remains fairly constant among all groups. In some cases, the harms of screening outweigh the benefits—for example, screening of younger individuals with no family history would not be recommended, which is why Australia recommends screening for those over the age of 50 and the United Kingdom recommends screening for those over age 55. The USPSTF undertook a large systematic review of the main screening tests for colorectal cancer and found that the newer immunochemical FOBTs have the best sensitivity (range 61 percent to 91 percent) and specificity (97 percent to 98 percent).[26]

- **Flexible sigmoidoscopy.** The USPSTF systematic review found insufficient evidence about the effect of community screening with flexible sigmoidoscopy on colorectal cancer mortality and limited data on complication rates in this setting.[26] The best estimate of complication rates pooled from six studies (n = 126,985) was 0.34 serious complications per 1,000 procedures (CI 0.06 to 1.9 per 1,000 procedures; test for heterogeneity p = 0.26). There is no evidence available to estimate the accuracy of flexible sigmoidoscopy or the optimal interval for screening by this method.
- **Colonoscopy**. The removal of polyps and the identification and treatment of colorectal cancers following a positive FOBT is responsible for the reduction in colorectal cancer mortality noted in the major trials reported earlier. There have been no randomized controlled trials of colonoscopy alone as a screening test, and no trials have directly compared the outcomes of FOBT versus screening colonoscopy. Its utility as a replacement test for biennial FOBT screening therefore remains uncertain.[27] Evidence is also insufficient evidence to estimate the accuracy of colonoscopy, but the Whitlock review identified several studies that confirm colonoscopy misses some polyps and also may miss colorectal cancer.[26] In addition, the major concern about colonoscopy as a stand alone screening test in an average risk populations is the well-documented chance of a serious complication. Whitlock pooled the results of 12 studies (n =57,742) and estimates 2.8 serious complications (including perforations, hemorrhage, diverticulitis, cardiovascular events, severe abdominal pain, and death) per 1,000 procedures (CI 1.5 to 5.2 per 1,000 procedures, test for heterogeneity p =0.13). Most of these complications appear to coincide with polypectomy. This complication rate is approximately 10 times higher than the estimates for flexible sigmoidoscopy. Colonoscopy, however, is expensive and requires time, equipment, and most of all, expertise.
- **CT Colonography.** This newer screening option for colorectal cancer appears to have equivalent sensitivity as standard colonoscopy for large adenomas and colorectal cancer, provided that there is adequate quality control. Uncertainties remain for smaller polyps and the referral threshold for follow-up colonoscopy. For example, based on a referral threshold of any polyp 6 mm or greater, the Whitlock review estimates that between 1 in 3 and 1 in 18 people screened with CT colonography would be referred for follow-up colonoscopy. The risk of perforation with screening CT colonography in asymptomatic people is very low, with no perforations reported in 14,238 screenees. Uncertainty remains about the long-term effects of radiation exposure with CT colonography.

VARIATIONS IN COLORECTAL CANCER SCREENING RECOMMENDATIONS AND PRACTICE

The USPSTF provides recommendations about each of these screening options based on their benefits-harms framework (see table 7.2). Recommendations include

Table 7.2
U.S. Preventive Services Task Force (USPSTF) Framework

	Net benefit				
Quality of Evidence	Substantial		Moderate	Small	Zero/Negative
Good	A	B	C		D
Fair	B	B	C	D	
Poor	I	I	I	I	

Key:

A: The USPSTF strongly recommends that clinicians provide the screening to eligible patients. It found good evidence that the screening improves important health outcomes and concludes that benefits substantially outweigh harms.

B: The USPSTF recommends that clinicians provide the screening to eligible patients. It found at least fair evidence that the screening improves important health outcomes and concludes that benefits outweigh harms.

C: The USPSTF does not make recommendations for or against routine provision of the screening. It found at least fair evidence that the screening can improve health outcomes but concluded that the balance of benefits and harms is too close to justify a general recommendation.

D: The USPSTF recommends against routinely providing the screening to asymptomatic patients. It found at least fair evidence that the screening is ineffective or that harms outweigh benefits.

I: The USPSTF concludes that the evidence is insufficient to recommend for or against routinely providing the screening. Evidence that the screening is effective is lacking, or poor quality, or conflicting, and the balance of benefits and harms cannot be determined.

Source: U.S. Preventative Services Task Force. Screening for colorectal cancer. http://www.ahrq.gov/clinic/uspstf/uspscolo.htm

screening for colorectal cancer using FOBT, sigmoidoscopy, *or* colonoscopy in adults beginning at 50 years of age and continuing until age 75. They recommend against routine screening above this age. In choosing which test to perform, patients need to weigh the benefits and harms of each. Conversely, the American Cancer Society suggests that screening begin at 50 years of age but that it end "at a point where curative therapy would not be offered due to life-limiting morbidity." Their guidelines more explicitly suggest that screening recommendations be related to personal and family history, but they do not offer specific guidance on this. The guidelines prefer tests that are effective in detecting and treating cancer and adenomatous polyps (for example, colonoscopy) over those that simply detect cancer (for example, FOBT). They more positively endorse the use of stool DNA (deoxyribonucleic acid) testing, screening CT colonography, and double-contrast barium enema, whereas the USPSTF states that evidence is sufficient for these or does not recommend them.

In the United Kingdom, the National Health Service (NHS) Bowel Cancer Screening Programme (BCSP) invites men and women in their 60s to be screened for bowel cancer every two years. Participants are sent a FOBT kit, which they complete at home and send to a laboratory for testing. Anyone with a positive result is

referred for a colonoscopy and any necessary treatment. People age 70 and over can request a kit.

In Australia, the national BCSP also offers people turning 50, 55, or 65 years of age between January 2008 and December 2010 the option of screening with a FOBT in the privacy of their own home. The policy for recurrent screening beyond the initial offer round is uncertain (see appendix 7. B for further country-specific guidelines). Inconsistencies exist in Australia between the government's evidence-based guidelines and the implementation of the government-funded screening program.[28] The guidelines recommend a risk-appropriate screening policy that includes three to five yearly colonoscopy for patients with high-risk family histories and biennial FOBT for average-risk members of the community over the age of 50 years.

Estimates of the costs of these screening options vary depending on the protocol and the in-country health service structure, and they change over time. According to one U.K. estimate, biennial FOBT screening costs 1,000 British pounds for every life saved and the five-yearly colonoscopy costs 3,000 British pounds for every life saved.[29] Another estimate from the United States found that screening costs between US$10,000 and US$25,000 for every life saved and that no one strategy appeared to be more optimal than others at that time.[30]

INFORMED CHOICE FOR CANCER SCREENING: DEBATES AND CONTROVERSIES

There is debate over informing consumers about possible limitations as well as the benefits of screening tests. Some argue that those participating in screening have a right to know the possible limitations as well as the benefits of screening tests. Others fear that providing such information would lead to a fall in participation rates. Evidence from Australia provides interesting findings. Randomized trials of evidence-based decision aids for mammography screening in women over 70 years and in people age 50 to 69 who are eligible for FOBT screening showed that people were better informed but no less likely to participate in screening after viewing the decision aids that contained information about the potential benefits *and* harms of screening.[31, 32] The exception to this was prostate cancer screening using the prostate specific antigen (PSA) test where the evidence of benefit is uncertain.[33]

One conclusion from these studies is that individuals value cancer screening so highly that they will often choose to be screened even if the potential benefit does not clearly outweigh the harms. Factors most likely to reduce participation are not cognitive but often are practical, such as busy lifestyle, the unpleasantness or inconvenience of the test, or other competing health factors. In some countries, however, new quality frameworks, such as performance indicators and incentives, sometimes are not well aligned with the patient-centered model of informed choice for preventive health care activities. For example, earlier this year, the National Health and Hospital Reforms Commission proposed that the Australian government consider a range of indicators within the next generation of the Australian Health Care Agreements. These included the proportion of 50- to 69-year-old women having a mammogram in the previous two years. Thus, there are financial and performance incentives for practitioners to increase screening participation rates, which may be at odds with allowing women the option to choose not to be screened at a particular point in time.

PITFALLS IN COMMUNICATION ABOUT SCREENING IN CLINICAL PRACTICE

FOBT kits can be obtained independently by the consumer and do not directly involve health care providers. As such, consumers have to rely on the information provided with the kit or on information obtained from the Internet, friends, or relatives. Information obtained from sources other than the physician may emphasize the potential benefits of screening and deemphasize or even omit information about the potential downsides of screening, including the complications of follow-up testing and the limitations in accuracy of the test. A balanced focus of information is needed.[34] To address the potential harms of misleading or inaccurate information, evidence-based decision aids for cancer screening have been developed to provide balanced information about the pros and cons of screening. A pragmatic approach to screening discussions needs to be adopted.

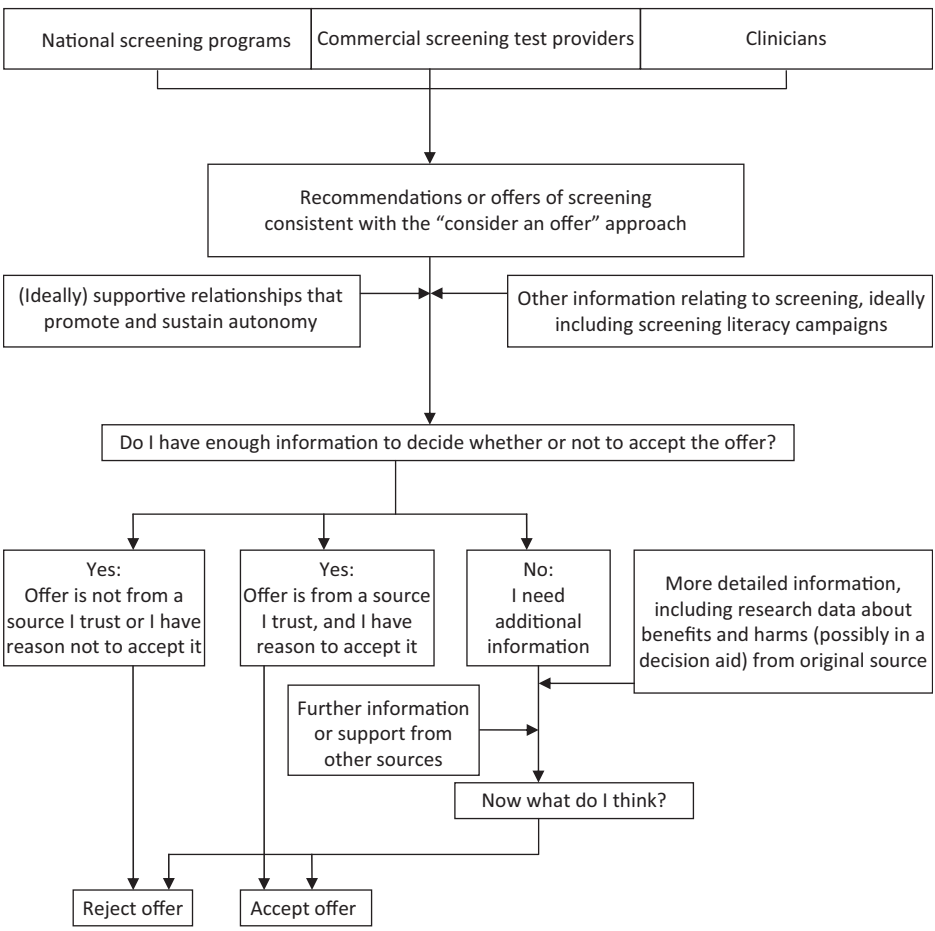

Figure 7.1 Communicating about screening in practice
Source: Entwistle V, Carter S, Trevena L, et al. Communicating about screening. *Br Med J.* 2008;337:a1591.

Australia has experimented with ways of communicating the benefits and harms of screening. The "Consider An Offer" approach relies on individuals feeling that they have sufficient information to accept or reject the offer to be screened and that the information comes from a trustworthy source.[35] (see figure 7.1) The "Consider An Offer" approach should include (1) an open explanation and discussion of the rationale for screening; (2) an individual assessment of the screening test (including consideration of the potential risks and benefits) and the trustworthiness of those making the claims; (3) access to additional information, if so requested; and (4) acknowledgment that the recommendation for screening might reasonably be refused.

HEALTH SYSTEMS AND IMPLEMENTATION

Currently, there is limited evidence for effective primary prevention strategies for colorectal cancer, although the promotion of increased physical activity within many communities is likely to have widespread benefits and improved health outcomes. The introduction of colorectal cancer screening programs, however, is a substantial investment for a health system and important issues of access and equity need to be considered. For many countries, colorectal cancer screening will not be a priority. Appendix 7. B summarizes the type of health system and current colorectal cancer screening programs in the top 10 countries for colorectal cancer global burden.

CONCLUSION

Governments need to recognize and appropriately fund colorectal cancer screening. The capacity for the existing health system to provide follow-up colonoscopy services for patients with a positive FOBT and treatment of screen-detected colorectal cancer is fundamental to the effective implementation of screening programs. Indeed, one could argue that it is unethical to screen individuals if follow-up testing and treatment are *not* available.

Governments face some tough decisions over the implementation of colorectal cancer screening programs at a population level. Recently, the American Cancer Society released a statement acknowledging that the benefits of some screening has been overstated with overtreatment of small cancers, while missing other potentially deadly cancers. This particularly seems to be the case for breast and prostate cancer for which there is no easy way to treat early cancers and precancers. Colorectal cancer, however, may be different in that FOBT is a low-cost highly sensitive and specific test designed to detect colon cancer, with follow-up colonoscopy having the capacity to detect and treat precancerous polyps as well as the cancer. As discussed, FOBT accuracy is limited, however, and false positive rates need to be weighed against the potential benefit in reducing colorectal mortality. These benefits will be greatest in high-risk groups for whom the prevalence of colorectal cancer is greatest. Many governments have adopted a uniform recommendation, which potentially places an unreasonable risk on those who may benefit the least. Perhaps the way of the future is risk-appropriate screening strategies that minimize the potential harms to low-risk individuals but maximize the potential benefits of screening for those at higher risk. Given the highly charged political climate and strong community sentiment around cancer screening, it also will be important to engage the general population and media in debating these issues.

Appendix 7. A
Common Epidemiological Terms Used in Screening

Sensitivity is the proportion of people with the disease who have a positive test—in other words, the true positive rate.
Specificity is the proportion of people without the disease who have a negative test—in other words, the true negative rate.
False positives are those people who have a positive test result but do not have the disease.
Lead time bias refers to an overestimation of the survival time for a disease. It is caused by a backward shift of the starting point of disease because earlier diagnosis makes it appear as though there is a greater survival benefit from screening.
Inconsequential disease is "disease" detected by screening but not likely to have any clinical significance to the patient if it had remained undetected. For example, some types of ductal carcinoma in situ detected at screening mammography would not be of clinical significance but, because it is detected through screening asymptomatic women, they are faced with the dilemma of whether or not to have treatment.
Interval cancers are those that arise between screening rounds.

Appendix 7. B
Statistics by Country for Colorectal Cancer

Country	Colorectal Cancer Deaths in 2004	Health System Type	Colorectal Cancer Screening Status
China [36]	107,600	Hospital-based care: user pays for services	No systematic screening or recommendations
United States [37]	62,100	Mainly private plus some safety net	FOBT, flex sig, or colonoscopy recommended for 50- to 74-year-olds by USPSTF (user pays)
Japan [36]	41,700	Self-financed	No systematic screening but consensus statement and guidelines recommending FOBT
Russian Federation	38,700	Self-financed, opportunistic	FOBT recommended by European consensus but not systematically implemented
India	36,000	Self-financed	No systematic screening or local guidelines
Germany	31,400	Government funded	Systematic biennial FOBT
Indonesia	24,200	Self-financed	No systematic screening or recommendations
France	21,600	Government funded	Systematic biennial FOBT

(Continued)

Appendix 7. B
(Continued)

Country	Colorectal Cancer Deaths in 2004	Health System Type	Colorectal Cancer Screening Status
United Kingdom	19,500	Government funded	Systematic biennial FOBT
Italy	18,800	Government funded	Systematic biennial FOBT

Source: Adapted from Statistics by country for colorectal cancer. www.cureresearch.com/c/colorectal/stats-country.html
Note: FOBT = fecal occult blood testing; USPSTF = U.S. Preventive Services Task Force.

REFERENCES

1. Andermann A, Blancquaert I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ*. 2008;86(4):241–320.
2. Wilson J, Jungner G. *Principles and Practice of Screening for Disease*. Geneva, Switzerland: World Health Organization; 1968. Public Health Paper No.34.
3. National Health Service. UK National Screening Committee: UK screening portal. http://www.screening.nhs.uk/screening. Published 2009.
4. World Health Organization. *The Global Burden of Disease: 2004 Update*. Geneva, Switzerland: World Health Organization; 2008.
5. Lopez A, Mathers C, Ezzati M, Jamison D, Murray C. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet*. 2006;367:1747–1757.
6. Danaei G, Vander Hoorn S, Lopez A, Murray C, Ezzati M. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *Lancet*. 2005;366:1784–1793.
7. National Health and Medical Research Council, Australian Cancer Network, Royal Australian College of General Practitioners. *Clinical Practice Guidelines for the Prevention, Early Detection and Management of Colorectal Cancer: A Guide for General Practitioners*. Sydney, Australia: Department of Health and Ageing; 2008.
8. Cockburn J, Paul C, Tzelepis F. Screening for bowel cancer among NSW adults with varying levels of risk: a community survey. *Aust N Z J Public Health*. 2002;26:236–241.
9. Belchetz L, Berk T, Bapat B, Cohen Z, Gallinger S. Changing causes of mortality in patients with familial adenomatous polyposis. *Dis Colon Rectum*. 1996;39:384–387.
10. Jarvinen H, Aarnio M, Mustonen H. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology*. 2000;118:829–834.
11. Australia Institute of Health and Welfare. *National Bowel Cancer Screening Program Monitoring Report.* Canberra, Australia: Australia Institute of Health and Welfare; 2008.
12. de Jong A, Hendriks Y, Kleibeuker J. Decrease in mortality in Lynch syndrome families because of surveillance. *Gastroenterology*. 2006;130:665–671.
13. Aaltonen L, Johns L, Jarvinen H. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumours. *Clin Cancer Res*. 2007;13:356–361.

14. Peel D, Ziogas A, Fox E. Characterization of hereditary nonpolyposis colorectal cancer families from a population-based series of cases. *J Natl Cancer Inst*. 2000;92:1517–1522.
15. Rodriguez-Bigas M, Boland C, Hamilton S. A National Cancer Institute workshop on hereditary nonpolyposis colorectal cancer syndrome: meeting highlights and Bethesda guidelines. *J Natl Cancer Inst*. 1997;89:1758–1762.
16. International Agency for Research on Cancer. GLOBOCAN. http://www-dep.iarc.fr/. Published 2002.
17. Wolin K, Yan Y, Colditz G, Lee I-M. Physical activity and colon cancer prevention: a meta-analysis. *Br J Cancer*. 2009;100:611–616.
18. Asano T, McLeod R. Dietary fibre for the prevention of colorectal adenomas and carcinomas. *Cochrane Database Syst Rev*. 2002;1:CD003430. doi: 10.1002/14651858.CD003430.
19. Weingarten M, Zalmanovici A, Yaphe J. Dietary calcium supplementation for preventing colorectal cancer and adenomatous polyps. *Cochrane Database Syst Rev*. 2008;1:CD003548. doi: 10.1002/14651858.CD003548.pub4.
20. Ding E, Mehta S, Fawzi W, Giovannucci E. Interaction of estrogen therapy with calcium and vitamin D supplementation on colorectal cancer risk: reanalysis of Women's Health Initiative randomized trial. *Int J Cancer*. 2008;122:1690–1694.
21. Bjelakovic G, Nikolova D, Simonetti R, Gluud C. Antioxidant supplements for preventing gastrointestinal cancers. *Cochrane Database Syst Rev*. 2008;3:CD004183. doi: 10.1002/14651858.CD004183.pub3.
22. Boehm K, Borrelli F, Ernst E, et al. Green tea (Camellia sinensis) for the prevention of cancer. *Cochrane Database Syst Rev*. 2009;3:CD005004. doi: 10.1002/14651858.CD005004.pub2.
23. Chlebowski R, Wactawaski-Wende J, Ritenbaugh C, et al. Estrogen plus progestin and colorectal cancer in postmenopausal women. *N Engl J Med*. 2004;350:991–1004.
24. US Preventative Services Task Force. Routine aspirin or non-steroidal anti-inflammatory drugs for the primary prevention of colorectal cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2007;146:361–364.
25. Hewitson P, Glasziou P, Irwig L, Towler B, Watson E. Screening for colorectal cancer using the fecal occult blood test, Hemoccult. *Cochrane Database Syst Rev*. 2007;10(4):112.
26. Whitlock E, Lin J, Liles E, Beil T, Fu R. Screening for colorectal cancer: a targeted, updated systematic review for the US Preventive Services Task Force. *Ann Intern Med*. 2008;149:638–658.
27. Bossuyt P, Irwig I, Craig J, Glasziou P. Comparative accuracy: assessing new tests against existing diagnostic pathways. *Br Med J*. 2006;332:1089–1092.
28. National Health and Medical Research Council. *Guidelines for the Prevention, Early Detection and Management of Colorectal Cancer: A Guide for General Practitioners*. Sydney, Australia: Department of Health and Ageing; 2000.
29. Scholefield J. ABC of colorectal cancer. *Br Med J*. 2000;321:1004–1006.
30. Pignone M, Saha S, Hoerger T, Mandelblatt J. Cost-effectiveness analyses of colorectal cancer screening: a systematic review for the US Preventive Services Task Force. *Ann Intern Med*. 2002;137(2):96–104.
31. Mathieu E, Barratt A, Davey H, McGeechan K, Howard K, Houssami N. Informed choice in mammography screening: a randomized trial of a decision aid for 70-year-old women. *Arch Intern Med*. 2007;167(19):2039–2046.
32. Trevena L, Irwig L, Barratt A. Randomised trial of a self-administered decision aid for colorectal cancer screening. *J Med Screen*. 2008;15:76–82.

33. Gattellari M, Ward J. Does evidence-based information about screening for prostate cancer enhance consumer decision-making? a randomised controlled trial. *J Med Screen*. 2003;10(1):27–39.
34. Gøtzsche P, Hartling O, Nielsen M, Brodersen J, Jørgensen K. Breast screening: the facts: or maybe not. *Br Med J*. 2009;338.
35. Entwistle V, Carter S, Trevena L, et al. Communicating about screening. *Br Med J*. 2008;337:a1591.
36. Sung J, Young J, Sao Y, et al. Asia Pacific consensus recommendations for colorectal cancer screening. *Gut*. 2008;57:1166–1176.
37. Agency for Healthcare Research and Quality. USPSTF recommendations for colorectal cancer screening 2008. http://www.ahrq.gov/clinic/uspstf08/colocancer/colors.htm.

CHAPTER 8

Advances in Breast Cancer: Reducing Risk and Improving Health Outcomes

Elena B. Elkin, PhD, and Victoria S. Blinder, MD, MPH, MS

Breast cancer is the most common nonskin cancer in American women, with almost 200,000 new cases of invasive disease diagnosed annually.[1] The World Health Organization estimates that approximately 1.2 million women worldwide are diagnosed with breast cancer each year.[2] The breast cancer mortality rate in the United States has declined in recent decades, from a peak of more than 33 deaths per 100,000 women to about 23 per 100,000 women in 2006.[3] Despite considerable scientific progress in breast cancer prevention, screening, and treatment, about 40,000 women in the United States die of breast cancer each year.[1] The morbidity and mortality associated with breast cancer have a tremendous economic impact, accounting for more than $8 billion in direct medical care [4] and more than $80 billion in the value of lives lost annually.[5]

This chapter presents an overview of notable advances in breast cancer prevention, screening, and treatment. Special attention is given to topics of particular interest and importance, which include racial disparities in breast cancer incidence and outcomes; strategies for predicting and reducing breast cancer risk; controversy regarding screening mammography guidelines and the age at which women should begin routine screening; emerging technologies in breast cancer treatment; and the issues faced by breast cancer survivors.

BREAST CANCER EPIDEMIOLOGY

Incidence

The average American woman faces a one in eight chance of developing breast cancer in her lifetime.[6] Breast cancer incidence increases with age, and the median age at diagnosis is 61 years.[7] Between 2000 and 2005, the incidence of breast cancer in women ages 40 to 79 declined by almost 9 percent, from 284.3 cases per 100,000 to 259.4 cases per 100,000.[8] A decline in incidence was seen across age groups, but it was greatest in women ages 55 to 69, who had a 13 percent reduction in the rate of new breast cancers. Among subtypes of breast cancer, the decrease in

incidence was more pronounced for hormone receptor–positive tumors than for hormone receptor–negative tumors. Observing these patterns, experts speculated that the decline in breast cancer incidence was attributable to a decrease in the use of postmenopausal hormone replacement therapy (HRT) following the release of findings of serious adverse health consequences of HRT, a hypothesis supported by several empiric studies.[8–10]

Risk Factors

Given the substantial public and personal health burden of breast cancer, much research has been devoted to identifying risk factors for this disease. Many of the known risk factors for breast cancer are not modifiable—for example, age, ethnicity, early age at menarche, late age at menopause, a history of benign breast disease, a family history of breast cancer, and specific genetic mutations associated with breast cancer. Early age at menarche (age 12 or younger) compared with later menarche (age 13 and older) is associated with an elevated risk of developing breast cancer. Breast cancer risk is two times greater for those who experience menopause at age 55 or older than for those who experience menopause at age 45 or younger.

Women with specific mutations in the BRCA1 and BRCA2 genes face a substantially increased risk of breast and ovarian cancers. In the general population, the frequency of a known mutation in one of these genes is estimated to be extremely low—less than 0.5 percent.[11–12] Women of Ashkenazi Jewish descent, however, are significantly more likely to have a mutation; one study estimated that 2 to 3 percent of Ashkenazi Jewish women carry at least one of the three mutations responsible for the majority of BRCA1/2-associated cancers.[13] Although fewer than 5 percent of all breast cancers are attributable to a BRCA1/2 mutation,[14] the lifetime risk of breast cancer among mutation carriers is about 60 percent, compared with 12 percent for the general population.

A meta-analysis of more than 20 pedigree studies estimated that the average cumulative risk of breast cancer by age 70 was 65 percent (95 percent confidence interval [CI] 44 to 78 percent) in women with a BRCA1 mutation and 45 percent (95 percent CI 31 to 56 percent) in BRCA2-mutation carriers.[15] Cancers associated with BRCA1/2 mutations tend to develop at younger ages,[16] and cancers attributed specifically to a BRCA1 mutation are more commonly negative for hormone receptors (estrogen and progesterone) and for human epidermal growth factor receptor-2 (HER2) overexpression.[17] These "triple-negative" cancers are associated with a particularly poor prognosis.[18]

Hormones play a major role in the etiology of breast cancer. Estrogen, in particular, stimulates cell division, and can increase the chances of a normal cell becoming a cancer cell. Breast cancer risk associated with exogenous estrogen exposure was addressed in the Women's Health Initiative (WHI), a large, prospective study initiated by the National Institutes of Health in 1991 to investigate the most common causes of morbidity and mortality in postmenopausal women. More than 160,000 women participated in the study, which included an observational cohort and several clinical trials designed to assess the effects of HRT, calcium, and vitamin D supplementation and diet on the risks of heart disease, osteoporosis and fractures, and cancers of the breast and colon. A placebo-controlled trial that evaluated

the impact of HRT with a combination of estrogen and progestin was stopped early when an elevated incidence of breast cancer and cardiovascular disease was observed among women in the HRT arm, in the absence of significant health benefits.[19] As a result of these findings, the U.S. Food and Drug Administration (FDA) issued a "black box" warning, and the use of HRT among postmenopausal women declined dramatically.

Analyses of population-based cancer surveillance data linked an observed decrease in breast cancer incidence to the decline in HRT use following publication of the WHI results.[8–10] An analysis of WHI participants suggested that while breast cancer risk increased with the duration of combined estrogen and progestin treatment, breast cancer incidence declined in the postintervention phase of the trial, supporting a causal relationship between HRT use and breast cancer risk.[20]

Not all women with one or more of the above risk factors will necessarily develop breast cancer, although they do face a greater-than-average risk of the disease. There are several other risk factors that can be modified, potentially reducing one's risk of breast cancer. A complex relationship has been demonstrated between obesity and breast cancer. Increased rates of breast cancer have been noted among obese postmenopausal women, attributed to the higher levels of circulating estrogen that result from increased peripheral adipose tissue. The opposite relationship has been observed in premenopausal women, and the etiologic mechanism of this relationship is not well understood.[21–22] Evidence regarding the impact of dietary consumption of fat and calcium on breast cancer risk is less clear, and studies of these relationships are limited by difficulty in measuring the level of exposure or intake.[23–26]

A number of tools have been developed to predict a woman's risk of breast cancer,[27] but perhaps the most important of these is the National Cancer Institute (NCI) Breast Cancer Risk Assessment Tool, based on the model developed by Gail et al.[28–29] Available online (http://www.cancer.gov/bcrisktool/), the tool predicts a woman's risk of developing breast cancer based on her age, the number of first-degree relatives with a history of breast cancer, number of prior breast biopsies, atypical hyperplasia in a biopsy specimen, age at menarche, and age at first live birth. The latter two risk factors are related to a woman's exposure to endogenous estrogen over the course of her lifetime.

Ethnic and Racial Disparities

Observed ethnic and racial disparities in breast cancer are a source of great concern and the focus of numerous investigations and interventions. Race and ethnicity have been associated with disparities in breast cancer incidence, stage, treatment, survival, and mortality. In particular, black women have had consistently poorer outcomes compared with white women. Although white women have a higher incidence of breast cancer (see figure 8.1), black women have a higher breast cancer mortality rate, and recent declines in breast cancer mortality have been larger for white women than for black women (see figure 8.2).

Minority women tend to be diagnosed with more advanced disease and experience greater morbidity and mortality related to their cancers.[30] Similar to disparities in mortality, evidence of disparities in stage distribution is clearest in

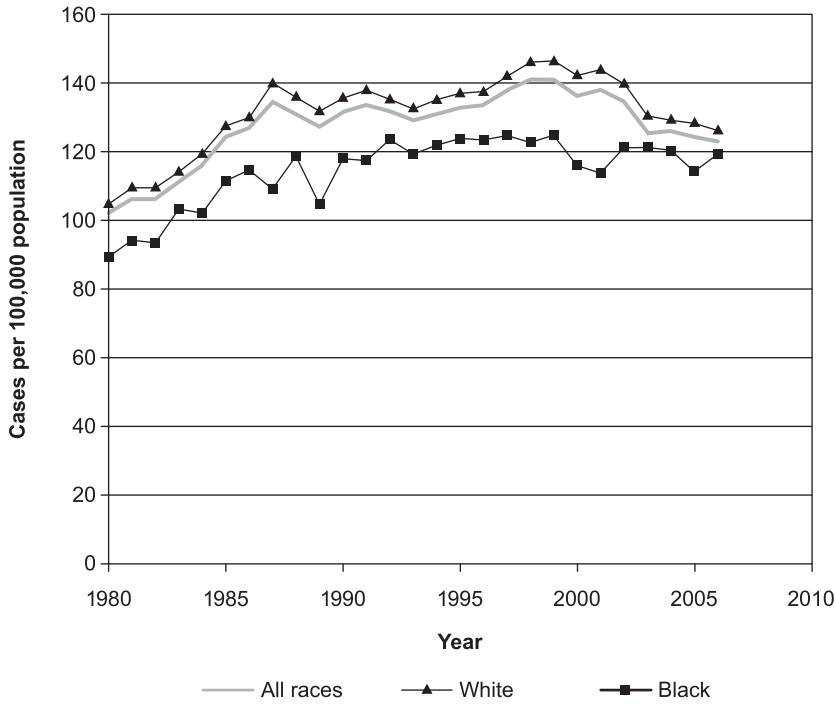

Figure 8.1 Breast cancer incidence age-adjusted to the 2000 U.S. population. [Source: Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence—SEER 9 Regs Limited-Use, Nov 2008 Sub (1973–2006), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2009, based on the November 2008 submission.]

comparisons of black women and white women. Among women diagnosed with breast cancer in 1999–2005, 62 percent of white women were diagnosed with a localized cancer, that is, cancer confined to the breast, whereas only 50 percent of black women were diagnosed with localized disease (see figures 8.3a and 8.3b). The proportion of women diagnosed with a distant-stage cancer was twice as high in black women compared with white women (4 percent versus 8 percent). Although late-stage diagnoses represent a minority of new breast cancer cases for women of both races, this racial disparity is especially important because of the poor prognosis for women whose disease has spread to distant sites.

Beyond mortality and survival, disparities in other breast cancer outcomes have been observed. Quality of care may be lower and a greater financial burden has been described in nonwhite populations.[31] Ethnic disparities have been cited in social avoidance, fear of recurrence, concern about the family's future, and coping mechanisms such as fatalism versus self-efficacy.[31–32]

Whether racial disparities in breast cancer outcomes are explained by differences in biology or in socioeconomic status is a subject of some debate. Socioeconomic characteristics, which are associated with race and ethnicity, influence breast cancer screening, stage at presentation, and treatment, and these factors may explain the higher breast cancer mortality rate observed in African American women compared with white women.[30, 33–35] Race, however, remains an independent

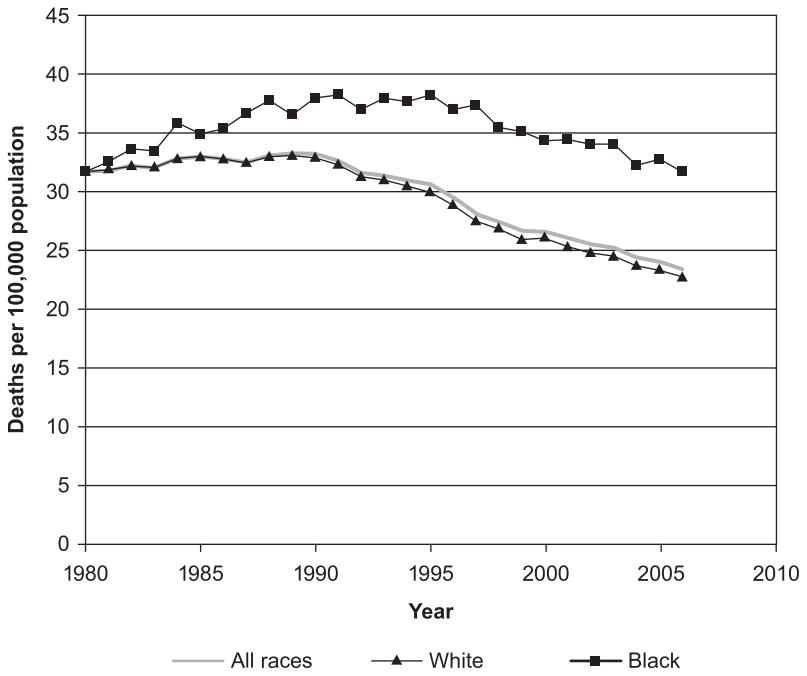

Figure 8.2 Breast cancer mortality age-adjusted to the 2000 U.S. population. [Source: Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Mortality—Total U.S. (1969–2006), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2009. Underlying mortality data provided by NCHS (www.cdc.gov/nchs).]

predictor of poorer survival and an increased risk of breast cancer mortality, controlling for a variety of socioeconomic characteristics and tumor features.[36–37]

Racial disparities in breast cancer mortality persist even in clinical trials where participants receive standardized treatment and are stratified by stage.[38–39] Therefore, it is likely that both socioeconomic and biologic factors contribute to observed differences in breast cancer mortality. For example, the basal-like phenotype, or "triple negative" breast cancer, which is associated with increased mortality and decreased response to therapy, occurs more frequently in premenopausal African-American women than in other women.[40–41] Attempts to isolate the contributions of socioeconomic versus biologic factors to racial disparities in breast cancer often are impaired by limited information regarding specific individual characteristics. Obesity, which has become more prevalent in blacks than whites, may be one such factor.[42–43]

Male Breast Cancer

Breast cancer is not solely a disease of women, although the overwhelming majority of cases are diagnosed in females. Each year about 1,900 men in the United States are diagnosed with breast cancer, accounting for 1 percent of all breast cancer cases.[1] Some of the risk factors for male breast cancer are similar to risk factors for female breast cancer, such as age, family history, BRCA1/2 mutations,

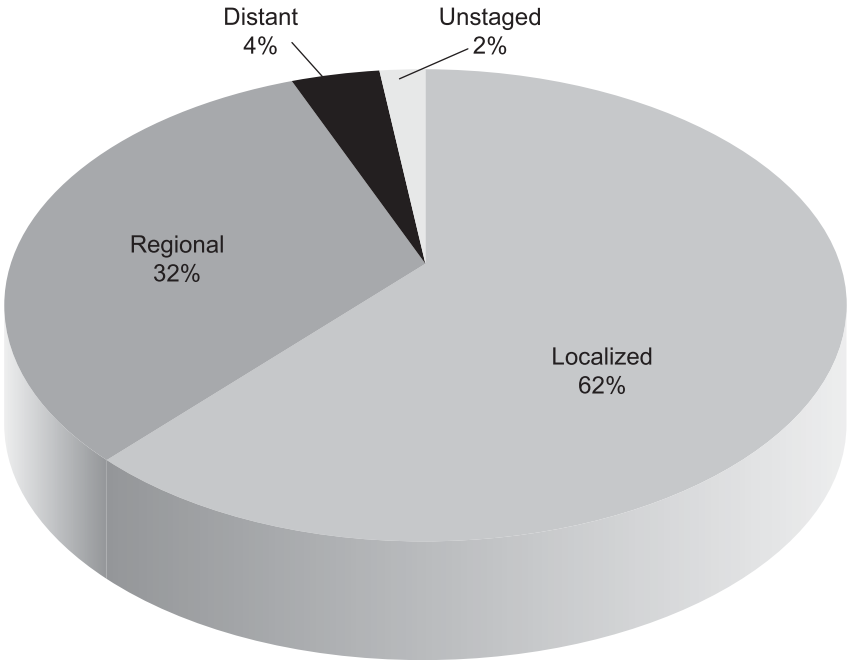

Figure 8.3a Breast Cancer Stage Distribution in White Women Diagnosed 1999–2005. (Source: SEER Cancer Statistics Review 1975–2006.)

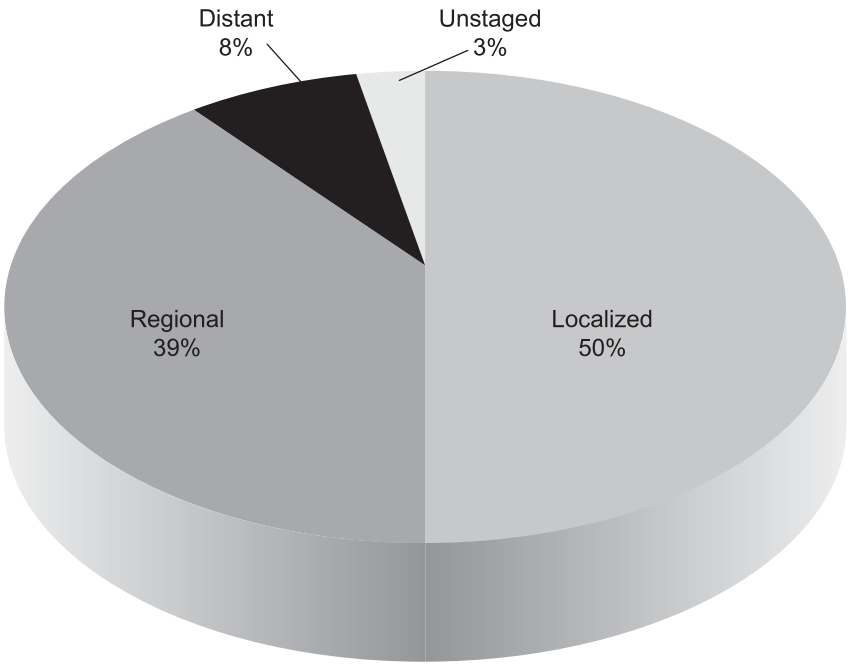

Figure 8.3b Breast Cancer Stage Distribution in Black Women Diagnosed 1999–2005. (Source: SEER Cancer Statistics Review 1975–2006.)

exogenous estrogen exposure, and obesity.[44] Other risk factors in men include increased endogenous estrogen due to a genetic abnormality such as Klinefelter's syndrome, and decreased endogenous androgen due to a testicular abnormality such as undescended testes, surgical removal of the testes, or medically induced androgen deprivation used in the treatment of prostate cancer. Men who work in hot environments or have occupational exposure to petroleum and exhaust fumes also may be at increased risk of developing breast cancer. Compared with female breast cancer, breast cancer in men tends to be diagnosed at an older age and later stage. The poorer prognosis observed in male breast cancer patients is primarily attributable to their later stage at diagnosis.[44]

BREAST CANCER RISK REDUCTION

Because so many of the risk factors for breast cancer are not modifiable, much attention has been devoted to other means of breast cancer prevention. A number of clinical interventions, both surgical and medical, can be effective in reducing breast cancer risk, and the benefits of these interventions may be greatest in women who face an elevated risk of developing the disease.

Chemoprevention

An increased understanding of the role of hormones in breast cancer etiology, the development and validation of models to predict breast cancer risk, and the observation that women who received tamoxifen for early stage breast cancer had a decreased risk of developing a new cancer in the opposite breast all motivated the investigation of antiestrogens for breast cancer prevention. In 1992, the National Surgical and Adjuvant Breast and Bowel Project (NSABP) initiated a trial of tamoxifen, a selective estrogen-receptor modulator (SERM) and a mainstay of systemic breast cancer therapy since the late 1970s. This study focused on many factors related to the prevention of breast cancer in women who were at increased risk—that is, over age 60, a history of lobular carcinoma in situ (LCIS), or a five-year predicted breast cancer risk of at least 1.66 percent.[45]

In total, more than 13,000 women were randomized to receive tamoxifen or a placebo for five years. After seven years of follow-up, tamoxifen decreased the relative risk (RR) of invasive breast cancer by 43 percent, corresponding to an absolute risk reduction of 1.8 percent.[45] While this finding was encouraging, there were significant adverse effects associated with this drug, including endometrial cancer (0.8 percent incidence with tamoxifen compared with 0.26 percent with placebo, RR 2.53) and cardiovascular events, including stroke (RR 1.59), pulmonary embolism (RR 3.1), and deep venous thrombosis (RR 1.6). With the exception of the latter, these adverse events primarily were limited to participants over 50 years of age. Similar results were found in three other studies of tamoxifen for breast cancer prevention from the same time period.[46–48]

More recently, similar reductions in breast cancer incidence were found among postmenopausal women taking raloxifene, another SERM. Raloxifene was first studied in the treatment of osteoporosis in postmenopausal women, and its impact on breast cancer risk was evaluated as a secondary endpoint in a four-year placebo-controlled trial. Like tamoxifen, raloxifene binds to the estrogen receptor on the cell

surface. While tamoxifen stimulates uterine cells but has antiestrogen effects on breast epithelial cells, raloxifene is an antagonist of both cell types. Thus, a hypothesized advantage of raloxifene is its lack of endometrial stimulation and therefore its carcinogenic potential. The relative risk of developing invasive breast cancer was 0.28 (95 percent CI, 0.17–0.46) in favor of the treatment arm.[49–50] This benefit was confirmed in the Study of Tamoxifen and Raloxifene (STAR), a randomized controlled trial comparing the two agents in postmenopausal women with an elevated risk of breast cancer.[51]

The STAR trial found that tamoxifen and raloxifene were equally effective in reducing the risk of invasive breast cancer. The incidence of noninvasive cancer was somewhat lower in women who received tamoxifen, but this difference was not statistically significant. Adverse-effect profiles favored raloxifene, which was associated with a lower risk of uterine cancer (RR 0.62, 95 percent CI, 0.35–1.08), thromboembolic events (RR 0.70, 95 percent CI, 0.54–0.91), cataracts (RR 0.79, 95 percent CI, 0.68–0.92), and cataract surgeries (RR 0.82, 95 percent CI, 0.68–0.99). No significant differences were found in the risks of fracture, ischemic heart disease, stroke, invasive cancer of other sites, or mortality. Based on these findings, the American Society of Clinical Oncology (ASCO) recommended that physicians consider prescribing a SERM for women at elevated risk of developing breast cancer.[52]

Although tamoxifen was developed and initially studied in patients with breast cancer, raloxifene is recommended only for breast cancer risk reduction in women without a history of breast cancer. It is not recommended as breast cancer therapy in either the adjuvant or metastatic setting. Tamoxifen may be prescribed to premenopausal women, and either tamoxifen or raloxifene may be used in postmenopausal women.

Despite clear evidence that both tamoxifen and raloxifene reduce the risk of invasive breast cancer, uptake of these therapies has been slow. Several studies have found low rates of interest in and use of tamoxifen for chemoprevention, even among women at elevated risk of breast cancer and even when women are well informed about the benefits of risk-reducing therapy.[53–55] Physician experience and familiarity likely play a role in women's use of breast cancer chemoprevention. One survey found that a minority of primary physicians and gynecologists had ever prescribed tamoxifen for chemoprevention, and those who did were more likely to have had a family member diagnosed with breast cancer, to have a patient who requested information about tamoxifen, or to believe that chemopreventive tamoxifen had a favorable balance of benefits to risks and that eligibility for tamoxifen was easy to determine.[56]

Risk-Reducing Surgery

Surgical approaches to breast cancer prevention, including bilateral mastectomy with or without bilateral oophorectomy, have been evaluated predominantly in gene-mutation carriers. In a prospective observational study of more than 400 BRCA1/2 mutation carriers identified from 11 medical centers, bilateral mastectomy reduced breast cancer incidence by 95 percent in women who had prior or concurrent oophorectomy and by 90 percent in women with intact ovaries.[57] In a retrospective single-institution study of about 600 women with a family history of

breast cancer, bilateral mastectomy reduced the risk of breast cancer by 80 to 90 percent.[58]

An increasing number of women diagnosed with unilateral breast cancer are opting for contralateral mastectomy, with the aim of preventing a new primary tumor in the opposite breast.[59–61] Several studies suggest that such surgery is effective in reducing the risk of contralateral breast cancer,[62] but there is little or no evidence of a long-term survival benefit, even in women with a known BRCA1/2 mutation.[63–64]

BREAST CANCER SCREENING

Mammography

Screening mammography represents a notable public health achievement and an important advance in the use of population screening to reduce the morbidity and mortality associated with cancer. The efficacy of screening mammography was first demonstrated in the 1960s. The Health Insurance Plan (HIP) study was the first large-scale randomized clinical mammography trial. Approximately 60,000 women ages 40 to 64 were randomized to have screening mammograms and clinical breast exams or not. The trial found a significant reduction in breast cancer mortality among those screened compared with those not screened.[65] Extended follow-up of this trial,[66–68] and the results of subsequent randomized trials, have shown 10 to 30 percent mortality reductions among women ages 50 to 69.[69–73] In women younger than age 50, there is less evidence for the effectiveness of screening, although one randomized trial demonstrated a significant mortality reduction in women ages 40 to 49.[74] Two randomized trials in Canada, however, found no mortality reductions in any age-group.[75–78] A meta-analysis of eight randomized trials suggested that screening mammography reduces breast cancer mortality by about 15 percent in women ages 39–49.[79]

Despite some controversy regarding trial design and analysis,[80–85] international consensus emerged on the merits of annual or biennial screening mammography for women ages 50 to 69,[86–87] and U.S. guidelines for women in this age-group have been fairly consistent over time and across advisory panels and agencies. Guidelines for women ages 40 to 49, however, have been a source of considerable controversy.[88–91] In 1989, a joint recommendation from the NCI, the American Cancer Society (ACS), and other organizations advocated screening for women ages 40 to 75, based on the age range of women included in the first randomized trials of screening mammography. Later, when additional trials were completed and subgroup analysis of the early trials was performed, the benefits of screening women in their 40s came into doubt.

In 1993 an international panel convened by NCI concluded that there was not sufficient evidence to recommend routine screening for women under age 50, and the panel advised that the NCI guideline be changed accordingly.[87] This advice, however, was rejected by the National Cancer Advisory Board (NCAB), whose members voted to continue recommending routine screening for women in their 40s. When the NCI director then overturned the NCAB's decision, heated debate ensued, prompting both a congressional investigation of the NCI's decision-making

process and fierce criticism of the Canadian trial that found no mortality benefit of annual screening mammography for women under age 50.[76] At the same time, the ACS modified its original recommendation, advising women in their 40s to make individual decisions about screening mammography in consultation with their physicians. Two years later, the ACS went back to its earlier recommendation for routine screening mammograms in women ages 40 to 49, consistent with the U.S. Preventive Services Task Force's (USPSTF) guideline.

In 1997, in a joint statement with the ACS, the NCI reverted to its earlier position, recommending regular screening mammography starting at age 40. These two agencies and the USPTF maintained a similar stance with respect to mammography for women in their 40s until 2009, when the USPSTF revised its guideline. Since the task force's previous update in 2002, new evidence regarding the benefits and risks of routine screening mammography had emerged. In particular, a randomized trial of more than 150,000 in England showed only a small difference in the breast cancer death rate after 10 years for women who were screened, and this difference was not statistically significant.[92] Investigators in Sweden published information from extended follow-up of another randomized trial also failing to find a statistically significant benefit of screening in women under age 50.[93]

Based on their review of the evidence from all randomized trials, the USPSTF revised its guideline, recommending that routine screening mammography start at age 50, not at age 40, for women at average risk of breast cancer, and that women over age 50 be tested every other year instead of every year.[94] In explaining the new recommendation, the panel cited the small and uncertain benefit of routine screening for women in their 40s, as well as the potential harms of over diagnosis—that is, treating screen-detected cancers that would not have been diagnosed or would not have caused morbidity or mortality in the absence of screening, and the potential harms of false-positive screening results, which are more common in women under age 50 than in older women.[94]

In addition to the updated meta-analysis of randomized trials,[79] the revised guideline was informed by a projection of the population impact of different screening strategies, which suggested that beginning routine screening mammography at age 40 rather than age 50 was associated with only a modest increase in benefit, but a substantial increase in the number of false-positive screens and the number of unnecessary biopsies.[95] The new screening recommendations were met with confusion, anger, and opposition from many, including breast cancer survivors and advocates, radiologists and other physicians, and healthy women in their 40s hesitant to forego their annual screening mammograms. Against the backdrop of a vigorous debate on health care reform, the guideline change engendered vocal outrage from politicians and others concerned that the USPSTF's action represented the first of many efforts to explicitly ration medical care.

A less contentious but similarly ambiguous question has been the value of regular mammography for women ages 70 and older. Compared with their younger counterparts, older women face a greater risk of developing breast cancer, but they also face a greater risk of death from other causes. Of the eight randomized trials of screening mammography, only one study included women in this age-group, which found a 24 to 27 percent reduction in breast cancer mortality among women ages 70 to 74. The result was not statistically significant.[96] Nonrandomized studies of

screening mammography in women ages 70 and older also have shown mortality reductions between 6 and 40 percent.[97–99] Although regular screening mammography for older women may reduce the risk of late-stage diagnosis and thereby improve prognosis,[100–102] the resulting average gain in life expectancy may be small compared with the potential benefits of screening in women under age 70 and with the costs and potential harms of screening.[103]

Given the lack of conclusive evidence of the value of regular screening mammography in women ages 70 and older, recommendations for this age-group have been mixed. In 1992, an NCI panel recommended that screening mammography decisions for women ages 75 and older be based on clinical judgment.[104] The USPSTF concluded in 1996 that evidence was insufficient to recommend either for or against routine mammography in women ages 70 and older,[105] and subsequent updates of these guidelines did little to clarify an upper age limit for screening.[106] Similarly, the ACS has not specified an age at which routine screening mammography should be terminated.[107]

Much of the controversy regarding screening mammography guidelines stems from concern about the methods employed in the design and analysis of the relevant screening trials.[80–85] The randomized controlled trial remains the gold standard for estimating the efficacy of a medical intervention, but other techniques have been used to estimate the population health benefits of screening mammography. Notably, computer simulation models developed by the NCI-sponsored Cancer Intervention and Surveillance Modeling Network (CISNET) suggest that almost half of the observed decline in breast cancer mortality between 1975 and 2000—from 48.3 deaths per 100,000 women (ages 30 to 79) in 1975 to 38 deaths per 100,000 by 2000—was attributable to populationwide screening mammography.[108] Simulating actual U.S. screening patterns from 1990 to 2000 compared with the alternative of no screening, one model estimated that screening mammography was associated with a gain of 1.7 million quality-adjusted life years.[109]

Despite the well-publicized benefits of screening mammography, about one-third of eligible women in the United States report having no recent mammogram, and this proportion varies widely by state and by ethnic groups.[110] After increasing for about 25 years, rates of screening mammography began to decline in 2000. From 2000 to 2005, two national surveys found a significant decrease in the percentage of women age 40 and older who reported having a mammogram in the prior two years.[111–112] Causes of the observed drop in screening mammography rates are not clear. Analyses of national survey data and Medicare claims found that women who lived in areas with greater mammography capacity—more machines per population—were more likely to have had a recent mammogram.[113] This suggests that the recent decrease in screening mammography rates may be associated, at least in part, with the documented decline in the number of mammography facilities in the United States for individual women making decisions about their health; for experts formulating policy at the population level, the benefits of screening mammography must be weighed against the potential harms, specifically the consequences of inaccurate mammogram readings. False-negative results—cancers missed by screening—are associated with treatment delays and litigation. Although radiologists' perceived risk of a malpractice action is generally greater than the actual risk, delayed diagnosis of breast cancer was the most common reason for medical

malpractice lawsuits filed in the United States in 1995 and in 2002. From a public health perspective, this has two important consequences. First, delayed diagnosis and treatment may have a detrimental effect on health outcomes.[114] Second, fear of litigation may dissuade radiologists from offering mammography, thereby reducing access to mammography services.[115–116]

False-positive results are abnormal findings on screening mammography in women who do not have breast cancer. Because breast abnormalities identified in a screening mammogram require further testing to confirm or rule out a diagnosis, false-positive findings generate substantial, and arguably unnecessary, health service utilization (including additional imaging and often a breast biopsy, which may be viewed as both costly and inconvenient). The potential distress and anxiety associated with false-positive screening results have been a source of particular concern. In a meta-analysis of the long-term outcomes of false-positive mammograms, most studies of this issue found elevated levels of distress and worry among women who had a false-positive result.[117] Evidence of heightened anxiety and fear was mixed, and no study found significantly greater levels of depression. Interestingly, an overwhelming majority of surveyed adults who received false-positive cancer screening exams (including mammography) indicated that they still were glad they had the tests,[118] and women who expressed generalized concern about overdiagnosis and overtreatment of breast cancer did not seem specifically concerned about false-positive results.[119] The potential harms of screening, particularly those due to the increased false-positive rate in women ages 40 to 49, were an important factor in the USPSTF's 2009 decision to revise its guideline, recommending against routine screening in women under age 50.[94]

Other Imaging Modalities

Although mammography remains the only guideline-recommended method for populationwide breast cancer screening, and the only screening modality supported by evidence of a mortality reduction, the use of newer imaging technologies has been advocated for screening women at high risk of breast cancer. In 2007, ACS modified its breast cancer screening guidelines recommending magnetic resonance imaging (MRI) as an adjunct to mammography in women with at least a 20 percent lifetime risk of breast cancer—a group that includes women with BRCA1/2 mutations and women previously treated for Hodgkin's lymphoma who are at increased risk of breast cancer because of prior radiation therapy. This recommendation was based on findings from a published report demonstrating the ability of breast MRI to detect cancers missed by standard mammography in the contralateral breast of women with a diagnosis of unilateral breast cancer.[120] Notably, however, this study provided no evidence that breast MRI, used as a screening modality or a diagnostic tool, reduced breast cancer mortality. Moreover, because the study was conducted in a group of patients already diagnosed with breast cancer, it provided no information about the use of breast MRI in healthy women at average risk of breast cancer.

Breast ultrasound also has been promoted as a cancer screening modality, particularly for women under age 50 in whom typically denser breast tissue limits the sensitivity of conventional mammography. When used as an adjunct to screening mammography in a high-risk population, breast ultrasound increased the sensitivity

of a single prevalence screen from 50 percent with mammography alone to 77 percent with both modalities combined.[121] This improvement in the detection of cancerous lesions, however, came at the expense of a decline in the specificity of screening (from 95 percent with mammography alone to 89 percent with mammography and ultrasound combined), resulting in an increase in false-positive findings. Perhaps more important, the study was not designed to demonstrate a difference in health outcomes, and information on cancer mortality was not collected.

Full-field digital mammography (FFDM), another technological advance in breast imaging, offers greater contrast resolution than conventional screen-film mammography. Like breast ultrasound, FFDM is more effective at detecting cancerous tumors in younger women and those with denser breasts. FFDM, however, does not appear to increase the false-positive rate of screening. In a trial of almost 50,000 asymptomatic women randomized to screening with conventional screen-film mammography or FFDM, there was no statistically significant difference in accuracy between the two modalities overall, but FFDM was more accurate in women under age 50, premenopausal or perimenopausal women, and women with heterogeneously dense or extremely dense breasts.[122] On the basis of this evidence, FFDM was approved by the FDA for breast cancer screening and diagnosis, and by 2009, more than half of all mammography facilities certified by the FDA had at least one digital unit.[123]

Breast MRI, ultrasound, and FFDM may offer important advantages over conventional mammography, especially for subgroups of women defined by their age, breast density, or breast cancer risk. In addition to a lack of evidence that these modalities result in improved health outcomes relative to mammography, adoption of these new imaging modalities as screening tools is hampered by important logistical challenges. Some of these modalities, MRI in particular, are expensive and require substantial capital investment as well as qualified technologists and radiologists with sufficient training in image interpretation.

ADVANCES IN BREAST CANCER TREATMENT

The treatment of breast cancer typically involves multiple modalities addressing various aspects of the disease, including surgery and radiation therapy for local-regional control and systemic therapies for preventing or halting the spread of the cancer to distant sites of the body. In all of these areas, scientific progress has facilitated the development of new therapeutic technologies and treatment strategies.

Local-Regional Therapy: Surgery

Surgical resection—alone or in combination with radiation therapy—remains the primary mode of local treatment for early stage breast cancer. After 20 years of follow-up, two large randomized trials comparing mastectomy and breast-conserving therapy (that is, lumpectomy and radiation) in women with tumors less than 2 cm (stage T1) and no palpable axillary lymph nodes found no difference between groups in long-term breast cancer survival.[124–125] Compared with mastectomy, breast-conserving therapy has been associated with higher ratings of body image, but quality of life after surgery appears to be similar with the two procedures and is influenced by the use of breast reconstruction.[126]

The selection of primary therapy should ideally be a function of tumor features, such as size and location, as well as patient characteristics and preferences. Because breast-conserving therapy requires radiation to attain efficacy comparable to that of mastectomy, it is not appropriate for women who have had prior radiation to the breast or chest wall, or who have diffuse suspicious microcalcifications, widespread disease, or positive margins evaluated on pathology.[127]

The use of breast-conserving therapy increased steadily during the 1980s and 1990s,[128–130] likely due to the widespread use of screening mammography and the resulting increase in the proportion of small breast tumors. More recently however, investigators at the Mayo Clinic observed an increase in the proportion of early stage breast cancer patients treated with mastectomy.[131] In this cohort and in other studies, mastectomy was significantly more common among women who received diagnostic breast MRI.[132–133] MRI is able to detect additional foci of disease that are not identified with mammography alone, in many cases leading to a change in surgical management favoring a more extensive procedure. It is not yet clear whether the use of MRI in preoperative staging results in improved health outcomes. Of particular concern is the high false-positive rate associated with MRI. Synthesizing data from 19 studies, Houssami et al. estimated that one-third of additional breast lesions identified only by MRI represent false-positive findings.[133]

Axillary Lymph Node Evaluation

Axillary lymph node stage is an important marker of prognosis and a determinant of subsequent treatment decisions in women with early stage breast cancer. Although axillary lymph node dissection (ALND) has been the standard method of axillary staging, it is associated with an increased risk of long-term adverse effects, including pain, numbness, swelling, and weakness of the ipsilateral arm, all of which may substantially reduce quality of life.[134] In the absence of a clear therapeutic benefit to ALND, sentinel node biopsy (SNB) has emerged as an alternative method of axillary staging. An SNB involves lymphatic mapping with a dye or radioactive tracer to identify the first lymph node to receive drainage from the primary tumor site. If the SNB reveals axillary lymph node involvement or if the sentinel node cannot be identified, then ALND is recommended.[127]

Comparisons of SNB and ALND suggest that they are concordant in 97 to 100 percent of cases.[135–138] In terms of clinical outcomes, SNB is associated with a lower rate of lymphedema, greater arm function, and superior quality of life, compared with ALND.[134] Because the success rate of SNB varies with surgeon experience and with patient and tumor characteristics, SNB is the preferred method of axillary lymph node staging only if the surgical team is sufficiently experienced and if the patient is an appropriate candidate.[139]

Radiation Therapy

Radiation therapy is a fundamental part of treatment for women with early stage breast cancer who choose breast-conserving surgery. Radiation may be recommended after mastectomy for women with positive surgical margins or other high-risk features. For women with advanced disease, radiation is often used with palliative intent to slow the growth of metastases. In women receiving breast-conserving therapy for

early stage disease, radiation typically is delivered by external beam to the whole breast. While improving both local control of disease and overall survival,[140] external beam whole-breast irradiation (WBI) also is associated with toxicities that may cause treatment delays and diminish quality of life and that may increase the risk of death from nonbreast cancer causes.[141–142]

Recent scientific advances have led to the development of several radiation therapy techniques, collectively referred to as accelerated partial breast irradiation (APBI), that aim to reduce the risk of acute radiation toxicities without compromising the effectiveness of treatment. For example, intensity-modulated radiation therapy (IMRT) uses computer-optimized treatment planning and computer-controlled intensity modulation to maximize the radiation dose to cancerous tissue, while minimizing the dose to surrounding normal tissue.[143] Breast cancer patients treated with IMRT experience fewer acute toxicities and achieve similar rates of local disease control, although long-term clinical outcomes are uncertain.[144–145] Another APBI technique involves brachytherapy, the implantation of radioactive seeds in the lumpectomy cavity. Observational studies suggest that breast brachytherapy, and specifically the delivery of high-dose radiation implants with a patented balloon catheter (Mammosite), can achieve outcomes similar to other APBI techniques.[146–149]

Systemic Therapy

Systemic therapy is a major component of treatment for women with early stage breast cancer and the dominant treatment modality for women with advanced disease. The use of systemic therapy in the adjuvant setting is directed by tumor and patient characteristics, with some form of systemic therapy recommended for nearly all women. In the metastatic setting, patients may receive multiple treatment regimens in successive attempts to control the spread of disease, with the selection of specific regimens determined by tumor and patient characteristics.

In the past 20 years, advances in pharmacotherapy and biotechnology have expanded the arsenal of agents available for reducing the risk of systemic breast cancer recurrence and halting the growth of lesions once the disease has metastasized. The most notable recent developments in systemic cancer therapy involve targeted approaches to treatment, using tumor markers and germline genetic information to identify women most likely to benefit from or least likely to be harmed by the various agents with demonstrated antitumor activity in breast cancer patients. As improvements in breast cancer treatment have led to increased rates of long-term survival among patients with both localized and metastatic disease, the focus of clinical research and practice has shifted to minimizing toxicity and improving quality of life in both the long term and the short term. Many clinical trials now include an assessment of quality of life as well as toxicity, and these considerations have come to the forefront of clinical care, particularly in the metastatic setting.

Hormonal Therapy

More than half of all women with breast cancer are postmenopausal at diagnosis, and 80 percent of these women have hormone receptor–positive disease. Thus, systemic endocrine therapy plays a large role in breast cancer treatment. Tamoxifen, the standard of care for postmenopausal women with hormone-responsive breast

cancer, is prescribed to women with metastatic breast cancer to halt the progression of advanced disease as well as to women in the adjuvant setting to prevent disease recurrence. Meta-analysis of clinical trials with more than 10 years of follow-up shows a 41 percent relative reduction in the risk of disease recurrence and a 34 percent relative reduction in the risk of death associated with five years of adjuvant tamoxifen in postmenopausal women with estrogen receptor (ER)-positive disease.[150] Tamoxifen, however, does have potentially serious side effects in addition to hot flashes and vaginal dryness risks include venous thromboembolism, stroke, and endometrial cancer.[151]

Aromatase inhibitors (AIs), specifically anastrozole, letrozole, and exemestane, have emerged as alternatives to tamoxifen in women with ER-positive or progesterone receptor (PR)-positive breast cancer. Evidence from several randomized trials suggests that AIs offer longer disease-free survival, compared with tamoxifen, when taken for five years following surgery.[152–154] Like tamoxifen, AIs are associated with hot flashes and sexual dysfunction, but have a very low risk of thromboembolic events and do not appear to increase the risk of stroke or endometrial complications.[151] AIs have been shown to reduce bone density and increase the risk of fracture and, therefore, may be contraindicated in women with osteoporosis.[155]

Chemotherapy

For women with metastatic disease, the use of chemotherapy depends primarily on hormone receptor status, HER2 status, and patient preference. Chemotherapy often is deferred in favor of less toxic endocrine therapies when the latter are likely to be effective. In such cases, chemotherapy may be administered once the disease has become resistant to endocrine therapies. First-line choices usually include single agents that have been shown to halt the progression of metastatic breast cancer, including taxanes, anthracyclines, and antimetabolites.[127] Combination chemotherapy generally is not used in the metastatic setting, given the increased risk of toxicity and a lack of clear evidence that it improves outcomes. The exception to this is triple-negative breast cancer, in which there is some evidence to indicate that combination chemotherapy improves progression-free survival and objective response rates.[156–157]

In the adjuvant setting, systemic chemotherapy has been shown to reduce the risk of disease recurrence. Thus, the expected benefit of chemotherapy corresponds with the magnitude of a woman's recurrence risk, which is a function of lymph node status, tumor size, grade, hormone-receptor status, and HER2 status. Given the numerous permutations of these tumor characteristics, clinical practice guidelines provide useful algorithms for physicians to assign patients to treatment.[127] To facilitate shared decision making, physicians and patients may use a decision aid to help women understand the benefits and risks of adjuvant chemotherapy. *Adjuvant!* is one such tool that can be accessed directly by patients online.[158–159]

Adjuvant chemotherapy almost always is administered when lymph node metastases are identified. In the absence of lymph node involvement, decisions regarding adjuvant chemotherapy are more complex, especially for women with a small primary tumor. A notable advance in breast cancer treatment is the development of tumor gene expression profiles (GEP) to guide adjuvant chemotherapy decisions.

GEP involves analysis of tumor genes using DNA (deoxyribonucleic acid) microarray or real-time polymerase chain reaction (RT-PCR) technology.[160] With each of the GEP tests commercially available in the United States (including the OncotypeDX and Mammaprint assays), tumor gene signatures from selected candidate genes are used to estimate a patient's risk of disease recurrence, based on proprietary risk prediction algorithms.[161–162] GEP has been recommended as a tool for risk-stratifying patients who would not be candidates for adjuvant chemotherapy based solely on other tumor features, such as lymph node involvement and size.[163–164]

Although several different chemotherapy regimens may be used in the adjuvant setting, most contain an anthracycline and a taxane.[127] Other regimens involve combinations of agents such as cyclophosphamide, methotrexate, and 5-fluorouracil. A meta-analysis of clinical trials suggests that anthracycline-based chemotherapy reduces annual breast cancer mortality by about 38 percent in women diagnosed before age 50 and by 20 percent in women ages 50 to 69.[150] A seminal randomized controlled trial evaluated the impact of treatment schedule, comparing the administration of adriamycin (an anthracycline), cyclophosphamide, and paclitaxel (AC-T) every two weeks to the same combination administered every three weeks.[165] The dose-dense schedule significantly improved disease-free survival and soon became a standard adjuvant strategy for node-positive breast cancer.

It is important to stress that the benefit of adjuvant chemotherapy corresponds with remaining life expectancy, and a woman's risk of death from other causes. This is particularly important in considering adjuvant chemotherapy for older women and those with multiple or serious comorbid conditions. Indeed, numerous studies have found that the use of adjuvant chemotherapy declines with age.[166–170] However, in both randomized trials and observational analyses, the adjuvant chemotherapy is effective in reducing the risk of recurrence and extending disease-free and overall survival in older women.[167–168, 171] Physicians may be reluctant to recommend adjuvant chemotherapy for their older patients even when functional status, comorbidity, and life expectancy suggest that it is a reasonable option.[172]

Targeted Biologic Therapy and Other Agents

Approximately 20 to 30 percent of breast cancers overexpress the HER2 protein, a product of the HER2/*neu* oncogene. Trastuzumab, a monoclonal antibody, has demonstrated antitumor effects in HER2-overexpressing breast cancers, increasing progression-free survival in women with HER2-positive metastatic breast cancer and improving both disease-free and overall survival in women with HER2-positive early stage breast cancer.[173–175] The ASCO, the College of American Pathologists (CAP), and the National Comprehensive Cancer Network (NCCN) recommend routine testing of all newly diagnosed breast cancers with immunohistochemistry (IHC), fluorescence in situ hybridization (FISH), or a combination of the two.[163–164] IHC and FISH vary with regard to their accuracy and performance characteristics, and the laboratories that perform these tests may vary in their procedures, accreditation, and proficiency. Trastuzumab is indicated only for women with a positive HER2 test result; women with HER2-negative tumors should not receive trastuzumab.[163–164]

Other agents approved for use in the treatment of metastatic breast cancer include targeted and antiangiogenic therapies, such as lapatinib and bevacizumab,

as well as chemotherapeutic agents that have been manipulated to minimize certain toxicities, such as nanoparticle albumin-bound paclitaxel or Abraxane®, which has a much lower risk of hypersensitivity reactions than standard paclitaxel. The benefits of these newer agents, particularly in relation to their high costs, are a subject of ongoing investigation, and the role of these agents in the adjuvant treatment of breast cancer is evolving.

Bisphosphonates

In breast cancer patients with bone metastases, bisphosphonates generally have been used to reduce the risk of fractures and other skeletal complications. Recent evidence suggests, however, that bisphosphonates may exert antitumor effects in women with early stage disease. In a randomized trial of more than 1,800 premenopausal women with hormone receptor–positive breast cancer, the bisphosphonate, zoledronic acid, combined with hormonal therapy was associated with an absolute increase in disease-free survival of 3.2 percentage points and a 36 percent reduction in the risk of disease progression.[176] Findings showed that women who received zoledronic acid were more likely to experience bone and joint pain, but there were no documented cases of osteonecrosis of the jaw—a rare but potentially severe adverse effect of bisphosphonate therapy. Although the fracture rate in this study was too low to make statistical comparisons between groups, results of other randomized trials show that zoledronic acid can safely prevent bone loss in premenopausal women receiving chemotherapy,[177] and in postmenopausal women receiving hormonal therapy for early stage breast cancer.[178]

ADVERSE EFFECTS OF TREATMENT

Secondary Malignancies

As increasing numbers of patients are cured of breast cancer, the risk of secondary malignancies—nonbreast cancers attributed to an effect of breast cancer treatment—persists. Secondary malignancies include chemotherapy-related leukemias, which primarily are associated with anthracyclines. This link has received increased attention as efforts have been made to assess the need for anthracycline-based therapy in different subtypes of breast cancer, and nonanthracycline options have gained favor in certain clinical situations.[179–180] Anthracyclines are among the most effective breast cancer therapies, however. Debate regarding their role has prompted efforts to determine which patients are most likely to benefit from their use and which patients may forego anthracycline-based treatment without a significant loss of clinical benefit. Other types of malignancies, most commonly soft-tissue sarcomas, can result from radiation therapy. Even surgical technique may be associated with a small risk of secondary malignancy in the form of Stewart-Treves Syndrome, the development of angiosarcomas in the setting of chronic lymphedema.[181–182]

Premature Ovarian Failure

Adjuvant chemotherapy causes premature permanent menopause in about two-thirds of premenopausal breast cancer patients. The risk of premature ovarian failure

is influenced by patient age and chemotherapy regimen.[183–185] The closer a woman is to natural menopause, the more likely it is that chemotherapy will induce menopause, and regimens that include an alkylating agent, such as cyclophosphamide, are more frequently associated with premature menopause than other regimens.[184] In addition to compromising fertility, premature ovarian failure commonly induces menopausal symptoms, such as hot flashes and vaginal dryness, and bone loss, potentially leading to osteopenia or osteoporosis.

Other Treatment Effects

Breast cancer survivors are at increased risk of other late effects of the disease and its treatment. Although surgical complications of axillary surgery have declined with the use of sentinel node biopsy, some patients are not eligible for this procedure and must have traditional ALND. Both carry risks of lymphedema and cellulitis, although the risk is higher after ALND.[186–187] Some patients may experience neurologic effects, such as numbness in the medial aspect of the arm as a result of surgical nerve damage and postmastectomy pain.[188–189] Brachial plexopathy is a rare complication of radiation therapy, but a more common adverse neurological effect is peripheral neuropathy attributed to chemotherapy, particularly the taxanes used in the adjuvant and metastatic settings.[190–192] Although the cardiac complications of radiation therapy have improved with the advent of more precise radiation techniques that can spare the heart and surrounding vasculature, chemotherapy and biologic therapy (most notably the anthracyclines and trastuzumab) have been associated with cardiomyopathy. Anthracycline-associated cardiomyopathy is dose related, may have a delayed onset, and typically is not reversible.[193] In contrast, trastuzumab-related cardiomyopathy can occur regardless of the cumulative dose administered and typically is reversible, such that administration of this drug may be suspended while cardiac function is medically optimized with appropriate medications. Trastuzumab may be administered again once cardiac function has returned to normal.[194–195]

BREAST CANCER SURVIVORSHIP

As a result of an increase in breast cancer incidence, the dissemination of screening mammography, improvements in treatment, and demographic shifts toward an older population, there are now more breast cancer survivors than ever before: an estimated 2.4 million in the United States, 57 percent of whom are women under the age of 65.[196] In recognition of the increasing number of cancer survivors and in an effort to help promote and support cancer survivorship research, the NCI established the Office of Cancer Survivorship (OCS) in 1996. Since its inception, the OCS has distributed more than $28 million in research grants and currently funds more than 100 projects to investigate a broad range of issues, such as quality of life, disease- and treatment-related cognitive deficits, vocational rehabilitation, exercise and weight loss, and the psychosocial effects of treatment.[197–198] Among the areas of particular emphasis in the OCS mission statement are health care disparities, economic outcomes (including work-related survivorship issues), optimization of follow-up care, and the development of study instruments to assess long-term quality of life and health outcomes.[197]

Awareness is growing among patients and physicians of the need to address the unique challenges facing cancer survivors. This has led to a surge in the investigation of issues such as the long-term medical complications of breast cancer and its treatment, the psychological burden of disease, cognitive changes associated with treatment, quality of life, and financial and employment-related concerns. In 2005, the Institute of Medicine (IOM) published its groundbreaking report, *From Cancer Patient to Cancer Survivor: Lost in Transition*.[198] In this authoritative text, the IOM's Committee on Cancer Survivorship charged oncologists with the responsibility of developing a "Survivorship Care Plan" for each patient. Professional organizations and patient advocacy groups have encouraged the use of survivor care plans in clinical practice, but there is no general consensus regarding the appropriate form and content of these plans, and little evidence regarding their effectiveness in communicating information to either primary care providers or survivors.

Guidelines are clearer with respect to recommended follow-up care and surveillance for disease recurrence in women who were treated with curative intent for early stage disease. The NCCN recommends annual mammography as well as a clinical history and physical exam every four to six months for the first five years and annually thereafter.[127] Women receiving adjuvant endocrine therapy should have follow-up care that is tailored to their regimen, including annual gynecologic exams for women taking tamoxifen who have not had a hysterectomy, and bone density monitoring for women taking an AI. Laboratory and imaging studies other than mammography are not recommended in the absence of symptoms or another clinical indication. ASCO has published similar guidelines, emphasizing regular clinician visits for physical exams and standard breast imaging.[199]

PUBLIC HEALTH PRIORITIES: DIRECTIONS FOR THE FUTURE

More than a century of scientific research has been translated into numerous advances for breast cancer risk reduction, screening, and treatment. As a result, more women than ever are surviving breast cancer, and living longer after their diagnosis. In the current century, the major public health challenges in breast cancer involve improving access to screening and treatment, reducing disparities in breast cancer treatment and outcomes, expanding resources for survivors, and using pharmacogenomic technologies to maximize the benefit of therapeutic interventions while minimizing their cost and harm.

As research continues on existing and new technologies for breast cancer risk reduction, screening, and treatment, an ongoing challenge for physicians and scientists is the communication of this information to breast cancer patients, survivors, and healthy women at risk of developing breast cancer. The outcry prompted by the USPSTF's recent change in screening mammography guidelines illustrates the importance of conveying scientific information to individuals and to the general public in a comprehensible way. Clearly communicating the risks and benefits of mammography is not simple, but it is essential for the dissemination of safe and effective population-based screening strategies.

The increasingly wide array of options for breast cancer risk reduction, screening, and treatment may demand more complex approaches to decision making. Particularly when available scientific evidence leaves questions unresolved, or when

patients are likely to vary in their preferences for the trade-off between benefits and risks of an intervention, a shared decision-making process involving both physician and patient, may be more satisfying and effective than the paternalistic approach common in earlier eras.[200] While collaborative approaches to decision making likely appeal to many patients and physicians, the constraints of time and reimbursement in modern medical practice may limit opportunities to implement such approaches on a widespread basis.

The interests of current and future breast cancer patients are well represented by some of the largest and most vocal disease advocacy organizations, and breast cancer research receives a substantial share of funding for medical research. In 2008, 13 percent or $725 million of the $5.6 billion allocated by the National Institutes of Health for cancer research was devoted to breast cancer.[201] While the pursuit of scientific breakthroughs to advance our understanding of breast cancer etiology and to expand the modalities available to detect and treat it remains a priority, the immediate challenges of delivering interventions—interventions that reduce risk, identify disease early, and halt its progression—must be addressed in order to further reduce the burden of breast cancer in the United States and around the world.

REFERENCES

1. American Cancer Society. *Cancer Facts and Figures 2009*. Atlanta, GA: American Cancer Society; 2009.
2. Ferlay J, Bray F, Pisani P, Parkin DM. *GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide, IARC CancerBase No. 5 version 2.0*. Lyon, France: International Agency for Research on Cancer; 2004.
3. Horner MJ, Ries LAG, Krapcho M, et al. *SEER Cancer Statistics Review: 1975–2006*. Bethesda, MD: National Cancer Institute; 2009.
4. National Cancer Institute. *Cancer Trends Progress Report: 2007 Update*. Bethesda, MD: National Cancer Institute/National Institutes of Health/US Department of Health and Human Services; 2007.
5. Yabroff KR, Bradley CJ, Mariotto AB, Brown ML, Feuer EJ. Estimates and projections of value of life lost from cancer deaths in the United States. *J Natl Cancer Inst*. 2008;100:1755–1762.
6. American Cancer Society. *Breast Cancer Facts and Figures: 2007–2008*. Atlanta, GA: American Cancer Society; 2008.
7. Ries LAG, Melbert D, Krapcho M, et al. *SEER Cancer Statistics Review: 1975–2004*. Bethesda, MD: National Cancer Institute; 2007.
8. Coombs NJ, Cronin KA, Taylor RJ, Freedman AN, Boyages J. The impact of changes in hormone therapy on breast cancer incidence in the US population. *Cancer Causes Control*. 2010;21(1):83–90.
9. Glass AG, Lacey JV Jr, Carreon JD, Hoover RN. Breast cancer incidence, 1980–2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. *J Natl Cancer Inst*. 2007;99:1152–1161.
10. Ravdin PM, Cronin KA, Howlader N, et al. The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med*. 2007;356:1670–1674.
11. Antoniou AC, Pharoah PD, McMullan G, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer*. 2002;86:76–83.

12. Whittemore AS, Gong G, John EM, et al. Prevalence of BRCA1 mutation carriers among US non-hispanic whites. *Cancer Epidemiol Biomarkers Prev*. 2004;13:2078–2083.
13. McClain MR, Nathanson KL, Palomaki GE, Haddow JE. An evaluation of BRCA1 and BRCA2 founder mutations penetrance estimates for breast cancer among Ashkenazi Jewish women. *Genet Med*. 2005;7:34–39.
14. Nathanson KL, Wooster R, Weber BL. Breast cancer genetics: what we know and what we need. *Nat Med*. 2001;7:552–556.
15. Antoniou A, Pharoah PD, Narod S, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet*. 2003;72:1117–1130.
16. Easton DF, Narod SA, Ford D, Steel M. The genetic epidemiology of BRCA1: Breast Cancer Linkage Consortium. *Lancet*. 1994;344:761.
17. Atchley DP, Albarracin CT, Lopez A, et al. Clinical and pathologic characteristics of patients with BRCA-positive and BRCA-negative breast cancer. *J Clin Oncol*. 2008;26:4282–4288.
18. Anders CK, Carey LA. Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. *Clin Breast Cancer*. 2009;9(Suppl 2):S73–S81.
19. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002;288:321–333.
20. Chlebowski RT, Kuller LH, Prentice RL, et al. Breast cancer after use of estrogen plus progestin in postmenopausal women. *N Engl J Med*. 2009;360:573–587.
21. Healy LA, Ryan AM, Rowley S, et al. Obesity increases the risk of postmenopausal breast cancer and is associated with more advanced stage at presentation but no impact on survival. *Breast J*. 2010;16(1):95–97.
22. Pike MC, Wu AH, Spicer DV, Lee S, Pearce CL. Estrogens, progestins, and risk of breast cancer. *Ernst Schering Found Symp Proc*. 2007;1:127–150.
23. Chen P, Hu P, Xie D, Qin Y, Wang F, Wang H. Meta-analysis of vitamin D, calcium and the prevention of breast cancer. *Breast Cancer Res Treat*. 2010;121(2):469–477.
24. Howe GR, Hirohata T, Hislop TG, et al. Dietary factors and risk of breast cancer: combined analysis of 12 case-control studies. *J Natl Cancer Inst*. 1990;82:561–569.
25. Lof M, Weiderpass E. Impact of diet on breast cancer risk. *Curr Opin Obstet Gynecol*. 2009;21:80–85.
26. Prentice RL. Measurement error and results from analytic epidemiology: dietary fat and breast cancer. *J Natl Cancer Inst*. 1996;88:1738–1747.
27. Jacobi CE, de Bock GH, Siegerink B, van Asperen CJ. Differences and similarities in breast cancer risk assessment models in clinical practice: which model to choose? *Breast Cancer Res Treat*. 2009;115:381–390.
28. Gail M, Brinton L, Byar D, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*. 1989;81:1879–1886.
29. Gail M, Costantino J, Bryant J, et al. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. *J Natl Cancer Inst*. 1999;91:1829–1846.
30. Baquet CR, Commiskey P. Socioeconomic factors and breast carcinoma in multicultural women. *Cancer*. 2000;88:1256–1264.

31. Ashing-Giwa KT, Padilla G, Tejero J, et al. Understanding the breast cancer experience of women: a qualitative study of African American, Asian American, Latina and Caucasian cancer survivors. *Psychooncology*. 2004;13:408–428.
32. Carver CS, Smith RG, Petronis VM, Antoni MH. Quality of life among long-term survivors of breast cancer: different types of antecedents predict different classes of outcomes. *Psychooncology*. 2006;15:749–758.
33. Bach PB, Schrag D, Brawley OW, Galaznik A, Yakren S, Begg CB. Survival of blacks and whites after a cancer diagnosis. *JAMA*. 2002;287:2106–2113.
34. Gold HT, Do HT, Dick AW. Correlates and effect of suboptimal radiotherapy in women with ductal carcinoma in situ or early invasive breast cancer. *Cancer*. 2008;113:3108–3115.
35. Gotay CC, Holup JL, Pagano I. Ethnic differences in quality of life among early breast and prostate cancer survivors. *Psycho-oncology*. 2002;11:103–113.
36. Newman LA, Mason J, Cote D, et al. African-American ethnicity, socioeconomic status, and breast cancer survival: a meta-analysis of 14 studies involving over 10,000 African-American and 40,000 white American patients with carcinoma of the breast. *Cancer*. 2002;94:2844–2854.
37. Menashe I, Anderson WF, Jatoi I, Rosenberg PS. Underlying causes of the black-white racial disparity in breast cancer mortality: a population-based analysis. *J Natl Cancer Inst*. 2009;101:993–1000.
38. Albain KS, Unger JM, Crowley JJ, Coltman CA Jr, Hershman DL. Racial disparities in cancer survival among randomized clinical trials patients of the Southwest Oncology Group. *J Natl Cancer Inst*. 2009;101:984–992.
39. Polite BN, Cirrincione C, Fleming GF, et al. Racial differences in clinical outcomes from metastatic breast cancer: a pooled analysis of CALGB 9342 and 9840, Cancer and Leukemia Group B. *J Clin Oncol*. 2008;26:2659–2665.
40. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer registry. *Cancer*. 2007;109:1721–1728.
41. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295:2492–2502.
42. Brawley OW. Is race really a negative prognostic factor for cancer? *J Natl Cancer Inst*. 2009;101:970–971.
43. Ward E, Jemal A, Cokkinides V, et al. Cancer disparities by race/ethnicity and socioeconomic status. *CA Cancer J Clin*. 2004;54:78–93.
44. Fentiman IS, Fourquet A, Hortobagyi GN. Male breast cancer. *Lancet*. 2006;367:595–604.
45. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. *J Natl Cancer Inst*. 2005;97:1652–1662.
46. Cuzick J, Forbes J, Edwards R, et al. First results from the International Breast Cancer Intervention Study: a randomised prevention trial. *Lancet*. 2002;360:817–824.
47. Powles T, Eeles R, Ashley S, et al. Interim analysis of the incidence of breast cancer in the Royal Marsden Hospital tamoxifen randomised chemoprevention trial. *Lancet*. 1998;352:98–101.
48. Veronesi U, Maisonneuve P, Costa A, et al. Prevention of breast cancer with tamoxifen: preliminary findings from the Italian randomised trial among hysterectomised women. *Lancet*. 1998;352:93–97.

49. Cummings SR, Eckert S, Krueger KA, et al. The effect of raloxifene on risk of breast cancer in postmenopausal women: results from the Multiple Outcomes of Raloxifene Evaluation randomized trial. *JAMA*. 1999;281:2189–2197.
50. Cummings SR, Duong T, Kenyon E, Cauley JA, Whitehead M, Krueger KA. Serum estradiol level and risk of breast cancer during treatment with raloxifene. *JAMA*. 2002;287:216–220.
51. Vogel VG, Costantino JP, Wickerham DL, et al. Effects of tamoxifen vs raloxifene on the risk of developing invasive breast cancer and other disease outcomes: the NSABP Study of Tamoxifen and Raloxifene P-2 trial. *JAMA*. 2006;295:2727–2741.
52. Visvanathan K, Chlebowski RT, Hurley P, et al. American society of clinical oncology clinical practice guideline update on the use of pharmacologic interventions including tamoxifen, raloxifene, and aromatase inhibition for breast cancer risk reduction. *J Clin Oncol*. 2009;27:3235–3258.
53. Bastian LA, Lipkus IM, Kuchibhatla MN, et al. Women's interest in chemoprevention for breast cancer. *Arch Intern Med*. 2001;161:1639–1644.
54. Melnikow J, Paterniti D, Azari R, et al. Preferences of Women Evaluating Risks of Tamoxifen study of preferences for tamoxifen for breast cancer risk reduction. *Cancer*. 2005;103:1996–2005.
55. Fagerlin A, Zikmund-Fisher BJ, Smith DM, et al. Women's decisions regarding tamoxifen for breast cancer prevention: responses to a tailored decision aid. *Breast Cancer Res Treat*. 2010;119(3):613–620.
56. Armstrong K, Quistberg DA, Micco E, Domchek S, Guerra C. Prescription of tamoxifen for breast cancer prevention by primary care physicians. *Arch Intern Med*. 2006;166:2260–2265.
57. Rebbeck TR, Friebel T, Lynch HT, et al. Bilateral prophylactic mastectomy reduces breast cancer risk in BRCA1 and BRCA2 mutation carriers: the PROSE Study Group. *J Clin Oncol*. 2004;22:1055–1062.
58. Hartmann LC, Schaid DJ, Woods JE, et al. Efficacy of bilateral prophylactic mastectomy in women with a family history of breast cancer. *N Engl J Med*. 1999;340:77–84.
59. Arrington AK, Jarosek SL, Virnig BA, Habermann EB, Tuttle TM. Patient and surgeon characteristics associated with increased use of contralateral prophylactic mastectomy in patients with breast cancer. *Ann Surg Oncol*. 2009;16:2697–2704.
60. McLaughlin CC, Lillquist PP, Edge SB. Surveillance of prophylactic mastectomy: trends in use from 1995 through 2005. *Cancer*. 2009;115(23):5404–5412.
61. Tuttle TM, Jarosek S, Habermann EB, et al. Increasing rates of contralateral prophylactic mastectomy among patients with ductal carcinoma in situ. *J Clin Oncol*. 2009;27:1362–1367.
62. McDonnell SK, Schaid DJ, Myers JL, et al. Efficacy of contralateral prophylactic mastectomy in women with a personal and family history of breast cancer. *J Clin Oncol*. 2001;19:3938–3943.
63. Brekelmans CT, Seynaeve C, Menke-Pluymers M, et al. Survival and prognostic factors in BRCA1-associated breast cancer. *Ann Oncol*. 2006;17:391–400.
64. Brekelmans CT, Tilanus-Linthorst MM, Seynaeve C, et al. Tumour characteristics, survival and prognostic factors of hereditary breast cancer from BRCA2-, BRCA1- and non-BRCA1/2 families as compared to sporadic breast cancer cases. *Eur J Cancer*. 2007;43:867–876.
65. Shapiro S, Strax P, Venet L. Periodic breast cancer screening in reducing mortality from breast cancer. *JAMA*. 1971;215:1777–1785.

66. Shapiro S. Evidence on screening for breast cancer from a randomized trial. *Cancer*. 1977;39(Suppl 6):2772–2782.
67. Shapiro S, Venet W, Strax P, Venet L. *Periodic Screening for Breast Cancer: The Health Insurance Plan Project and Its Sequelae, 1963–1986*. Baltimore, MD: Johns Hopkins University Press; 1988.
68. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst*. 1982;69:349–355.
69. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *BMJ*. 1988; 297:943–948.
70. Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg Breast Cancer Screening Trial: preliminary results on breast cancer mortality for women aged 39–49. *J Natl Cancer Inst Monogr*. 1997;22:53–55.
71. Frisell J, Lidbrink E, Hellstrom L, Rutqvist LE. Followup after 11 years: update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat*. 1997;45:263–270.
72. Roberts MM, Alexander FE, Anderson TJ, et al. Edinburgh trial of screening for breast cancer: mortality at seven years. *Lancet*. 1990;335:241–246.
73. Tabar L, Fagerberg G, Duffy SW, Day NE. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. *J Epidemiol Community Health*. 1989;43:107–114.
74. Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmo Mammographic Screening Program. *J Natl Cancer Inst Monogr*. 1997:63–67.
75. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening study 2: breast cancer detection and death rates among women aged 50 to 59 years. *Can Med Assoc J*. 1992;147:1477–1488.
76. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening study 1: breast cancer detection and death rates among women aged 40 to 49 years. *Can Med Assoc J*. 1992;147:1459–1476.
77. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening study: update on breast cancer mortality. *J Natl Cancer Inst Monogr*. 1997: 37–41.
78. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening study 2: 13–year results of a randomized trial in women aged 50–59 years. *J Natl Cancer Inst*. 2000;92:1490–1499.
79. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: an update for the US Preventive Services Task Force. *Ann Intern Med*. 2009;151:727–737, W237–W242.
80. Freedman DA, Petitti DB, Robins JM. On the efficacy of screening for breast cancer. *Int J Epidemiol*. 2004;33:43–55.
81. Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet*. 2000;355:129–134.
82. Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet*. 2001;358:1340–1342.
83. Begg CB. The mammography controversy. *Oncologist*. 2003;7:174–176.
84. Duffy SW. Interpretation of the breast screening trials: a commentary on the recent paper by Gotzsche and Olsen. *Breast*. 2001;10:209–212.
85. Olsen O, Gotzsche PC. Screening for breast cancer with mammography. *Cochrane Database Syst Rev*. 2001;(4):CD001877.

86. Boyle P, Autier P, Bartelink H, et al. European code against cancer and scientific justification: third version. *Ann Oncol*. 2003;14:973–1005.
87. Fletcher SW, Black WC, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer. *J Natl Cancer Inst*. 1993;85:1644–1656.
88. Ernster VL. Mammography screening for women aged 40 through 49: a guidelines saga and a clarion call for informed decision making. *Am J Public Health*. 1997;87:1103–1106.
89. Taubes G. The breast-screening brawl. *Science*. 1997;275:1056–1059.
90. Wells J. Mammography and politics of randomized controlled trials. *Br Med J*. 1998;317:1224–1230.
91. Antman K, Shea S. Screening mammography under age 50. *JAMA*. 1999;281:1470–1472.
92. Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet*. 2006;368:2053–2060.
93. Bjurstam N, Bjorneld L, Warwick J, et al. The Gothenburg Breast Screening Trial. *Cancer*. 2003;97:2387–2396.
94. US Preventive Services Task Force. Screening for breast cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2009;151:716–726.
95. Mandelblatt JS, Cronin KA, Bailey S, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Ann Intern Med*. 2009;151:738–747.
96. Tabar L, Vitak B, Chen HH, et al. The Swedish Two-County Trial twenty years later: updated mortality results and new insights from long-term follow-up. *Radiol Clin North Am*. 2000;38:625–651.
97. Jonsson H, Tornberg S, Nystrom L, Lenner P. Service screening with mammography of women aged 70–74 years in Sweden: effects on breast cancer mortality. *Cancer Detect Prev*. 2003;27:360–369.
98. Lenner P, Jonsson H. Excess mortality from breast cancer in relation to mammography screening in northern Sweden. *J Med Screen*. 1997;4:6–9.
99. van Dijck JA, Verbeek AL, Beex LV, et al. Breast-cancer mortality in a non–randomized trial on mammographic screening in women over age 65. *Int J Cancer*. 1997;70:164–168.
100. McCarthy EP, Burns RB, Freunds KM, et al. Mammography use, breast cancer stage at diagnosis, and survival among older women. *J Am Geriatr Soc*. 2000;48:1226–1233.
101. Smith-Bindman R, Kerlikowske K, Gebretsadik T, Newman J. Is screening mammography effective in elderly women? *Am J Med*. 2000;108:112–119.
102. Solin LJ, Schultz DJ, Legoretta AP, Goodman RL. Downstaging of breast carcinomas in older women associated with mammographic screening. *Breast J*. 1999;5:94–100.
103. Kerlikowske K, Salzmann P, Phillips KA, Cauley JA, Cummings SR. Continuing screening mammography in women aged 70 to 79 years: impact on life expectancy and cost-effectiveness. *JAMA*. 1999;282:2156–2163.
104. Screening recommendations of the forum panel. *J Gerontol*. 1992;47(Spec No):5.
105. US Preventive Services Task Force. *Guide to Clinical Services*. Baltimore, MD: Williams and Wilkins; 1996.
106. US Preventative Services Task Force. Screening for breast cancer: recommendations and rationale. *Ann Intern Med*. 2002;137:344–346.

107. Leitch AM, Dodd GD, Costanza M, et al. American Cancer Society guidelines for the early detection of breast cancer: update 1997. *CA Cancer J Clin*. 1997; 47:150–153.
108. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med*. 2005;353:1784–1792.
109. Stout NK, Rosenberg MA, Trentham-Dietz A, Smith MA, Robinson SM, Fryback DG. Retrospective cost-effectiveness analysis of screening mammography. *J Natl Cancer Inst*. 2006;98:774–782.
110. Miller JW, King JB, Ryerson AB, Eheman CR, White MC. Mammography use from 2000 to 2006: state-level trends with corresponding breast cancer incidence rates. *AJR Am J Roentgenol*. 2009;192:352–360.
111. Centers for Disease Control and Prevention. Use of mammograms among women aged $\geq$ 40 years United States, 2000–2005. *MMWR*. 2007;56:49–51.
112. Breen N, A Cronin K, Meissner HI, et al. Reported drop in mammography: is this cause for concern? *Cancer*. 2007;109:2405–2409.
113. Elkin EB, Ishill NM, Snow JG, et al. Geographic access and the use of screening mammography. *Med Care*. 2010;48(4):349–356.
114. Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet*. 1999;353:1119–1126.
115. Cypel YS, Sunshine JH, Ellenbogen PH. The current medical liability insurance crisis: detailed findings from two ACR surveys in 2003 and 2004. *J Am Coll Radiol*. 2005;2:595–601.
116. Elmore JG, Taplin SH, Barlow WE, et al. Does litigation influence medical practice? the influence of community radiologists' medical malpractice perceptions and experience on screening mammography. *Radiology*. 2005;236:37–46.
117. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med*. 2007;146:502–510.
118. Schwartz LM, Woloshin S, Fowler FJ Jr, Welch HG. Enthusiasm for cancer screening in the United States. *JAMA*. 2004;291:71–78.
119. Schwartz LM, Woloshin S, Sox HC, Fischhoff B, Welch HG. US women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. *BMJ*. 2000;320:1635–1640.
120. Lehman CD, Gatsonis C, Kuhl CK, et al. MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. *N Engl J Med*. 2007;356:1295–1303.
121. Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA*. 2008;299:2151–2163.
122. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med*. 2005;353:1773–1783.
123. US Food and Drug Administration. *Mammography Quality Standards Act and Program: 2009 Scorecard Statistics*. Washington, DC: US Food and Drug Administration; 2009.
124. Fisher B, Anderson S, Bryant J, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. *N Engl J Med*. 2002;347:1233–1241.
125. Veronesi U, Cascinelli N, Mariani L, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. *N Engl J Med*. 2002;347:1227–1232.

126. Pockaj BA, Degnim AC, Boughey JC, et al. Quality of life after breast cancer surgery: what have we learned and where should we go next? *J Surg Oncol*. 2009;99:447–455.
127. National Comprehensive Cancer Network. NCCN Clinical practice guidelines in oncology: breast cancer.
128. Foote RL, Johnson RE, Donohue JH, et al. Trends in surgical treatment of breast cancer at Mayo Clinic, 1980–2004. *Breast*. 2008;17:555–562.
129. Nattinger AB, Gottlieb MS, Veum J, Yahnke D, Goodwin JS. Geographic variation in the use of breast-conserving treatment for breast cancer. *N Engl J Med* 1992;326:1102–1107.
130. Nattinger AB, Hoffmann RG, Kneusel RT, Schapira MM. Relation between appropriateness of primary therapy for early-stage breast carcinoma and increased use of breast-conserving surgery. *Lancet*. 2000;356:1148–1153.
131. Katipamula R, Degnim AC, Hoskin T, et al. Trends in mastectomy rates at the Mayo Clinic, Rochester: effect of surgical year and preoperative magnetic resonance imaging. *J Clin Oncol*. 2009;27:4082–4088.
132. Bleicher RJ, Ciocca RM, Egleston BL, et al. Association of routine pretreatment magnetic resonance imaging with time to surgery, mastectomy rate, and margin status. *J Am Coll Surg*. 2009;209:180–187, 294–295.
133. Houssami N, Hayes DF. Review of preoperative magnetic resonance imaging (MRI) in breast cancer: should MRI be performed on all women with newly diagnosed, early-stage breast cancer? *CA Cancer J Clin*. 2009;59:290–302.
134. Sclafani LM, Baron RH. Sentinel lymph node biopsy and axillary dissection: added morbidity of the arm, shoulder and chest wall after mastectomy and reconstruction. *Cancer J*. 2008;14:216–222.
135. Veronesi U, Paganelli G, Galimberti V, et al. Sentinel-node biopsy to avoid axillary dissection in breast cancer with clinically negative lymph-nodes. *Lancet*. 1997;349:1864–1867.
136. Albertini JJ, Lyman GH, Cox C, et al. Lymphatic mapping and sentinel node biopsy in the patient with breast cancer. *JAMA*. 1996;276:1818–1822.
137. Krag D, Weaver D, Ashikaga T, et al. The sentinel node in breast cancer: a multicenter validation study. *N Engl J Med*. 1998;339:941–946.
138. Veronesi U, Paganelli G, Viale G, et al. Sentinel lymph node biopsy and axillary dissection in breast cancer: results in a large series. *J Natl Cancer Inst*. 1999;91:368–373.
139. Lyman GH, Giuliano AE, Somerfield MR, et al. American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. *J Clin Oncol*. 2005;23:7703–7720.
140. Clarke M, Collins R, Darby S, et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet*. 2005;366:2087–2106.
141. Browall M, Ahlberg K, Karlsson P, Danielson E, Persson LO, Gaston-Johansson F. Health-related quality of life during adjuvant treatment for breast cancer among postmenopausal women. *Eur J Oncol Nurs*. 2008;12:180–189.
142. Early Breast Cancer Trialists' Collaborative Group. Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: an overview of the randomised trials. *Lancet*. 2000;355:1757–1770.
143. Vicini F, Sharpe M, Kestin L, Wong J, Remouchamps V, Martinez A. The use of intensity modulated radiation therapy in the treatment of breast cancer:

evolving definition, misdirected criticism, and untoward effects. *Int J Radiat Oncol Biol Phys*. 2004;58:1642–1644.
144. Livi L, Buonamici FB, Simontacchi G, et al. Accelerated partial breast irradiation with IMRT: new technical approach and interim analysis of acute toxicity in a phase III randomized clinical trial. *Int J Radiat Oncol Biol Phys*. 2009;77(2): 509–515.
145. Bhatnagar AK, Beriwal S, Heron DE, et al. Initial outcomes analysis for large multicenter integrated cancer network implementation of intensity modulated radiation therapy for breast cancer. *Breast J*. 2009;15:468–474.
146. Harper JL, Watkins JM, Zauls AJ, et al. Six-year experience: long-term disease control outcomes for partial breast irradiation using MammoSite balloon brachytherapy. *Am J Surg*. 2009;199(2):204–209.
147. Nelson JC, Beitsch PD, Vicini FA, et al. Four-year clinical update from the American Society of Breast Surgeons MammoSite brachytherapy trial. *Am J Surg*. 2009;198:83–91.
148. Wadasadawala T, Budrukkar A, Chopra S, et al. Quality of life after accelerated partial breast irradiation in early breast cancer: matched pair analysis with protracted whole breast radiotherapy. *Clin Oncol*. 2009;21:668–675.
149. Wadasadawala T, Sarin R, Budrukkar A, Jalali R, Munshi A, Badwe R. Accelerated partial-breast irradiation vs conventional whole-breast radiotherapy in early breast cancer: a case-control study of disease control, cosmesis, and complications. *J Cancer Res Ther*. 2009;5:93–101.
150. Early Breast Cancer Trialists' Collaborative Group. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet*. 2005;365:1687–1717.
151. Perez EA. Safety profiles of tamoxifen and the aromatase inhibitors in adjuvant therapy of hormone-responsive early breast cancer. *Ann Oncol*. 2007;18 (Suppl 8):viii–35.
152. Howell A, Cuzick J, Baum M, et al. Results of the ATAC (Arimidex, Tamoxifen, alone or in combination) trial after completion of 5 years' adjuvant treatment for breast cancer. *Lancet*. 2005;365:60–62.
153. Thurlimann B, Keshaviah A, Coates AS, et al. A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. *N Engl J Med*. 2005;353:2747–2757.
154. Lin NU, Winer EP. Advances in adjuvant endocrine therapy for postmenopausal women. *J Clin Oncol*. 2008;26:798–805.
155. Perez EA, Weilbaecher K. Aromatase inhibitors and bone loss. *Oncology*. 2006;20:1029–1039, 39–40, 42, 48.
156. Thomas ES, Gomez HL, Li RK, et al. Ixabepilone plus capecitabine for metastatic breast cancer progressing after anthracycline and taxane treatment. *J Clin Oncol*. 2007;25:5210–5217.
157. Bunnell C, Vahdat L, Schwartzberg L, et al. Phase I/II study of ixabepilone plus capecitabine in anthracycline-pretreated/resistant and taxane-resistant metastatic breast cancer. *Clin Breast Cancer*. 2008;8:234–241.
158. Siminoff LA, Gordon NH, Silverman P, Budd T, Ravdin PM. A decision aid to assist in adjuvant therapy choices for breast cancer. *Psycho-oncology*. 2006;15:1001–1013.
159. Ravdin PM, Siminoff LA, Davis GJ, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol*. 2001;19:980–991.

160. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. *N Engl J Med*. 2009;360:790–800.
161. Buyse M, Loi S, Van't Veer L, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. *J Natl Cancer Inst*. 2006;98:1183–1192.
162. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351:2817–2826.
163. National Comprehensive Cancer Network. *NCCN Clinical Practice Guidelines in Oncology™ Breast Cancer (Version 1.2009)*. National Comprehensive Cancer Network, Inc.; 2009.
164. Wolff AC, Hammond ME, Schwartz JN, et al. American Society of Clinical Oncology/College of American Pathologists—guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer. *J Clin Oncol*. 2007;25:118–145.
165. Citron ML, Berry DA, Cirrincione C, et al. Randomized trial of dose-dense versus conventionally scheduled and sequential versus concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: first report of Intergroup Trial C9741/Cancer and Leukemia Group B trial 9741. *J Clin Oncol*. 2003;21:1431–1439.
166. Du XL, Key CR, Osborne C, Mahnken JD, Goodwin JS. Discrepancy between consensus recommendations and actual community use of adjuvant chemotherapy in women with breast cancer. *Ann Intern Med*. 2003;138:90–97.
167. Elkin EB, Hurria A, Mitra N, Schrag D, Panageas KS. Adjuvant chemotherapy and survival in older women with hormone receptor-negative breast cancer: assessing outcome in a population-based, observational cohort. *J Clin Oncol*. 2006;24:2757–2764.
168. Giordano SH, Duan Z, Kuo YF, Hortobagyi GN, Goodwin JS. Use and outcomes of adjuvant chemotherapy in older women with breast cancer. *J Clin Oncol*. 2006;24:2750–2756.
169. Mandelblatt JS, Hadley J, Kerner JF, et al. Patterns of breast carcinoma treatment in older women: patient preference and clinical and physician influences. *Cancer*. 2000;89:561–573.
170. Merchant TE, McCormick B, Yahalom J, Borgen P. The influence of older age on breast cancer treatment decisions and outcome. *Int J Radiat Oncol Biol Phys*. 1996;34:565–570.
171. Muss HB, Woolf S, Berry D, et al. Adjuvant chemotherapy in older and younger women with lymph node–positive breast cancer. *JAMA*. 2005;293:1073–1081.
172. Hurria A, Naeim A, Elkin E, et al. Adjuvant treatment recommendations in older women with breast cancer: a survey of oncologists. *Crit Rev Oncol Hematol*. 2007;61:255–260.
173. Piccart-Gebhart MJ, Procter M, Leyland-Jones B, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med*. 2005; 353:1659–1672.
174. Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med*. 2005;353:1673–1684.
175. Slamon D, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med*. 2001;344:783–792.
176. Gnant M, Mlineritsch B, Schippinger W, et al. Endocrine therapy plus zoledronic acid in premenopausal breast cancer. *N Engl J Med*. 2009;360:679–691.

177. Hershman DL, McMahon DJ, Crew KD, et al. Zoledronic acid prevents bone loss in premenopausal women undergoing adjuvant chemotherapy for early-stage breast cancer. *J Clin Oncol*. 2008;26:4739–4745.
178. Brufsky AM, Bosserman LD, Caradonna RR, et al. Zoledronic acid effectively prevents aromatase inhibitor–associated bone loss in postmenopausal women with early breast cancer receiving adjuvant letrozole: Z-FAST study 36-month follow-up results. *Clin Breast Cancer*. 2009;9:77–85.
179. Jones SE, Savin MA, Holmes FA, et al. Phase III trial comparing doxorubicin plus cyclophosphamide with docetaxel plus cyclophosphamide as adjuvant therapy for operable breast cancer. *J Clin Oncol*. 2006;24:5381–5387.
180. Slamon D, Eiermann W, Robert N. Phase III randomized trial comparing doxorubicin and cyclophosphamide followed by docetaxel with doxorubicin and cyclophosphamide followed by docetaxel and trastuzumab with docetaxel, carboplatin and trastuzumab in HER2 positive early breast cancer patients: BCIRG 006 study: San Antonio Breast Cancer Symposium; December 8–11, 2005; San Antonio, TX.
181. Stewart FW, Treves N. Classics in oncology: lymphangiosarcoma in postmastectomy lymphedema: a report of six cases in elephantiasis chirurgica. *CA Cancer J Clin*. 1981;31:284–299.
182. West JG, Qureshi A, West JE, et al. Risk of angiosarcoma following breast conservation: a clinical alert. *Breast J*. 2005;11:115–123.
183. Burstein HJ, Winer EP. Reproductive issues. In: Harris JR, ed. *Diseases of the Breast*. Philadelphia, PA: Lippincott; 2000:1051–1059.
184. Partridge AH, Burstein HJ, Winer EP. Side effects of chemotherapy and combined chemohormonal therapy in women with early-stage breast cancer. *J Natl Cancer Inst. Monogr*. 2001;30:135–142.
185. Shapiro C, Recht A. Drug therapy: side effects of adjuvant treatment of breast cancer. *N Engl J Med*. 2001;344:1997–2008.
186. Canavese G, Catturich A, Vecchio C, et al. Sentinel node biopsy compared with complete axillary dissection for staging early breast cancer with clinically negative lymph nodes: results of randomized trial. *Ann Oncol*. 2009;20:1001–1007.
187. McLaughlin SA, Wright MJ, Morris KT, et al. Prevalence of lymphedema in women with breast cancer 5 years after sentinel lymph node biopsy or axillary dissection: objective measurements. *J Clin Oncol*. 2008;26:5213–5219.
188. Crane-Okada R, Wascher RA, Elashoff D, Giuliano AE. Long-term morbidity of sentinel node biopsy versus complete axillary dissection for unilateral breast cancer. *Ann Surg Oncol*. 2008;15:1996–2005.
189. Del Bianco P, Zavagno G, Burelli P, et al. Morbidity comparison of sentinel lymph node biopsy versus conventional axillary lymph node dissection for breast cancer patients: results of the sentinella-GIVOM Italian randomised clinical trial. *Eur J Surg Oncol*. 2008;34:508–513.
190. Galecki J, Hicer-Grzenkowicz J, Grudzien-Kowalska M, Michalska T, Zalucki W. Radiation-induced brachial plexopathy and hypofractionated regimens in adjuvant irradiation of patients with breast cancer–a review. *Acta Oncol*. 2006;45:280–284.
191. Kuroi K, Shimozuma K, Ohashi Y, et al. Prospective assessment of chemotherapy-induced peripheral neuropathy due to weekly paclitaxel in patients with advanced or metastatic breast cancer study. *Support Care Cancer*. 2008; Epub ahead of print.

192. Vahdat L. Choosing a taxane for adjuvant treatment of breast cancer: more than a flip of the coin? *Nat Clin Pract Oncol*. 2008;5:570–571.
193. Jensen BV. Cardiotoxic consequences of anthracycline-containing therapy in patients with breast cancer. *Semin Oncol*. 2006;33:S15–S21.
194. Ewer SM, Ewer MS. Cardiotoxicity profile of trastuzumab. *Drug Saf*. 2008;31: 459–467.
195. Suter TM, Cook-Bruns N, Barton C. Cardiotoxicity associated with trastuzumab (Herceptin) therapy in the treatment of metastatic breast cancer. *Breast*. 2004;13:173–183.
196. National Cancer Institute. Estimated US cancer prevalence counts: who are our cancer survivors in the US? http://dccps.nci.nih.gov/ocs/prevalence/index.html. Published 2007. Accessed July 10, 2007.
197. National Cancer Institute, Office of Cancer Survivorship. About cancer survivorship research: history. http://dccps.nci.nih.gov/ocs/history.html. Published 2007. Accessed July 10, 2007.
198. Hewitt M, Greenfield S, Steovall E. *From Cancer Patient to Cancer Survivor: Lost in Transition*. Washington, DC: National Academies Press; 2006.
199. Khatcheressian JL, Wolff AC, Smith TJ, et al. American Society of Clinical Oncology 2006 update of the breast cancer follow-up and management guidelines in the adjuvant setting. *J Clin Oncol*. 2006;24:5091–5097.
200. Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). *Soc Sci Med*. 1997;44: 681–692.
201. National Institutes of Health. Estimates of funding for various research, condition, and disease categories 2009. http://report.nih.gov/rcdc/categories/. Accessed December 16, 2009.
202. Surveillance Epidemiology and End Results, National Cancer Institute, Division of Cancer Control and Population Sciences. SEER*Stat Database: SEER 9 Regs Limited-Use, November 2008 Sub (1973–2006). Updated November 2008. Accessed April 2009.

CHAPTER 9

Advances in Lung Cancer Prevention, Screening, and Treatment: A Comparative Effectiveness Research Perspective

Madhu Mazumdar, PhD

INTRODUCTION

Lung cancer, the leading cause of cancer mortality worldwide, typically exhibits symptoms only after the disease has spread to other organs, making it difficult to cure patients with such advanced disease. The overall prognosis of this cancer is poor when compared with other cancers, such as breast or colon, and is dependent on where the cancer is located, the size and type of tumor, and the overall health status of the patient. The two types of lung cancers (small-cell lung cancer [SCLC] and non-small-cell lung cancer [NSCLC]) grow and spread in different ways and also have different treatment options.

The National Cancer Institute estimates that 1 in every 14 men and women in the United States will be diagnosed with lung cancer at some point in their lifetime; 70 percent of those diagnosed are over age 65 years. Lung cancer develops slowly over time; only 3 percent of lung cancer cases occur in those younger than age 45 years. In the United States, an estimated 213,380 new lung cancer diagnoses were made in 2007, and there were more than 160,000 deaths. Whereas lung cancer incidence and mortality rates among American males have been declining since the mid-1990s, among American women, only recently have incidence rates leveled off although mortality rates continue to increase. For both males and females, however, long-term survival after a lung cancer diagnosis has changed little, with overall five-year relative survival increasing only slightly from 12.4 percent for 1974–1976 diagnoses to 15 percent for 1996–2002 diagnoses.[1] This slight improvement in survival reflects the difficulty in diagnosing lung cancer at an early, treatable stage.

This chapter focuses on advances in lung cancer prevention, screening, and treatment from an efficacy viewpoint. Concern for generalizability of the clinical

This research was supported by two grants: Center for Education and Research on Therapeutics grant (CERTs), Agency of Health Research and Quality RFA-HS-05-14, and Clinical Translational Science Center (CTSC) grant (UL1-RR024996). Ms. Sarah Laura R'bibo is acknowledged for her assistance with literature search.

outcomes obtained in research studies across a variety of settings and patient populations as well as concerns about increasing health care cost issues have demanded that research focus on the effectiveness and efficiency of treatment regimens in addition to being efficacious. A national initiative for comparative effectiveness research (CER) is gaining momentum to aid clinical decision making. A discussion of CER provides a review of ongoing research and initiatives in this area, and highlights the gaps in information and research. Overall, there is still a long way to go in finding cure for lung cancer and in being able to bring the best possible care to patients of all race, gender, and socioeconomic status.

CAUSES AND CORRELATES OF LUNG CANCER

Primary lung cancer is heterogeneous in its clinical presentation, histopathology, and treatment response. Conventionally, lung cancer has been divided into NSCLC and SCLC. NSCLC are the most common lung cancers, accounting for 80 percent of this type of cancer, and include adenocarcinomas (the most common type of NSCLC), squamous cell carcinomas (representing 30 percent of NSCLC), and large cell carcinomas (the least common type). The most localized NSCLC cancers can be removed surgically with relatively high rates of five-year survival. SCLC is the most aggressive and rapidly growing of all lung cancers and is strongly related to cigarette smoking (only 1 percent of these tumors are found in nonsmokers who develop lung cancer). SCLC metastasizes rapidly to many sites and is most often diagnosed after having spread extensively. Median survival time is only two to four months after diagnosis if untreated; this type of lung cancer, if detected and diagnosed in an early stage, is quite responsive to radiation therapy and chemotherapy. But, everything is relative. Despite treatment, only 5 to 10 percent of patients are still alive five years after diagnosis.

As with most cancers, cancer stage is the most significant predictor of survival. The tumor, node, metastasis (TNM) staging system is based on clinical assessment of tumor size, pathologic diagnosis of tumor spread to lymph nodes, and clinical identification of metastases. This system, however, does not completely account for the tremendous heterogeneity among patients with lung cancer, as evidenced by the variable prognosis and response to treatment for patients within the same staging group.

Although cancers of the lung have proven difficult to diagnose early and treat successfully, the cause of most of these cancers is well known. Epidemiologic studies have demonstrated that cigarette smoking is the major risk factor in the development of lung cancer,[2] with a striking dose-response relationship.[3] It is estimated that 80 to 90 percent of lung cancer incidence can be attributed to cigarette smoking. The remaining 10 to 20 percent of lung cancer is attributed to other causes, such as secondhand smoke, active smoking of other tobacco products, and exposure to other carcinogens, such as asbestos, radon, radiation therapy, combustion products, and various other exposures in occupational, environmental, or medical settings.[4] Clearly, having individuals not smoke would do much to reduce the incidence of lung cancer.

Even with the high attributable risks due to cigarette smoke exposure, only 10 to 15 percent of all smokers will develop lung cancer, suggesting that there are host differences in susceptibility to lung carcinogens.[5] Efforts toward predicting which smokers are at highest risk are under way.[6] This knowledge will enable medical practitioners and researchers to focus screening and chemoprevention efforts more

precisely to appropriate subgroups of patients and to offer insights into the biologic mechanisms of smoking prevention.[7] Multimodality approaches to smoking cessation, including telephone and Web-based counseling and pharmacotherapy, have been found to be effective but depend on the determination of the person for wanting to quit and are related to the individual's psychological well-being.

ADVANCES IN LUNG CANCER TREATMENT

Treatments of lung cancer include many modalities, with surgery, chemotherapy, and radiation therapy being the primary ones. Guidelines on the role of each of these therapies are available, but the diversity in outcomes points to the heterogeneous nature of this disease.[8, 9] Clinical trials with targeted therapies have begun, but completion and mature follow-up are needed to assess their success.[10, 11] Novel statistical designs representing a paradigm shift from the traditional approach have been developed for these clinical trials.[12, 13]

Clinical management and outcome of recurrent lung cancer have demonstrated benefits of various treatments over no treatment. Treatment modality is directed mostly by lung cancer type, and stage determines disease survival.[14] For example, a surgical resection of stage I NSCLC offers five-year survival rates of more than 50 percent.[15] But, for any given chemo- or radiotherapy regimen, which is prescribed by the standard protocol with a fixed dosage, the response rate and toxicity range vary widely among patients.[16, 17] The disease-free survival time, even after adjusting for the well-established predictors (disease stage, histology, performance status, and treatment) also varies significantly. Published literature suggests that a majority of NSCLC recurrences occur within two years after surgery and late recurrence may occur up to 10 years later. Recurrent disease has been reported in 4 to 5 percent of five-year NSCLC survivors.[15, 18] Up to 10 percent of recurrences may be discovered beyond five years after initial curative therapy. Differentiation between lung cancer recurrence in the lung and occurrence of a new primary lung cancer often is difficult. Molecular characterization of recurrent tumors is ongoing and, if successful, potentially could aid in providing second-line treatment on a timely basis.[19]

The current standard of care for patients with stage IA NSCLC lung cancer is surgical resection alone, without adjuvant chemotherapy. Approximately 30 to 35 percent of these patients will relapse after initial surgery, however, and thus have a much worse prognosis. Studies have documented the benefits of chemotherapy administered after resection, but a number of questions remain regarding how overall outcomes can be further improved. To provide the oncology community with clear direction, guidelines developed through consensus statements are available.[8, 9] Neoadjuvant systemic therapy generally is not recommended but can be considered to downstage an unresectable patient. Currently, there is no role for preoperative radiation or chemoradiation. Adjuvant systemic therapy is not recommended for stage IA and IB patients; however, adverse prognostic factors are acceptable reasons to consider adjuvant systemic therapy in the latter. Adjuvant systemic therapy is recommended for stage IIA, IIB, and IIIA patients. A cisplatin-based regimen should be started within 60 days after surgery, but if relatively contraindicated, carboplatin is an acceptable alternative. Adjuvant radiation therapy is not recommended for N0 and N1 patients, but is used in N2 patients to decrease local

recurrence.[20] A comprehensive list of treatment options and registry of all clinical trials with experimental therapies are available on the National Cancer Institute (NCI) Web site (http://www.clinicaltrials.gov/).

ADVANCES IN SMOKING PREVENTION AND CESSATION EFFORTS

Since the landmark report on smoking and health by the Surgeon General in 1964,[21] substantial progress has been made in reducing smoking rates in the United States. In 2003, about 21.6 percent of adults smoked, far less than that in earlier decades.[22] The American Cancer Society (ACS), the Centers for Disease Control and Prevention (CDC), NCI, and the North American Association of Central Cancer Registries (NAACCR) collaborate annually to provide updated information on cancer occurrence and trends in the United States. Recent data show a prominent variation by geography and gender. Lung cancer incidence and death rates among women, for example, increased in 18 states, 16 of them in the South or Midwest, where, on average, the prevalence of smoking was higher and the annual percentage decrease in current smoking among adult women was lower than in the West and Northeast. California was the only state with decreasing lung cancer incidence and death rates in women. Large state and regional differences in lung cancer trends underscore the need to maintain and strengthen many state tobacco control programs through stricter public policy.[5]

The Healthy People 2010 goal is to reduce smoking prevalence to 12 percent by the year 2010.[23] Much of the decline in smoking rates is attributable to effective tobacco control policies. Smoking prevention policies primarily include tax, bar, and worksite restrictions; youth access policies; media policies; and advertising bans.[23–25] Although some states (for example, California and Utah) have made impressive gains in reducing the number of individuals who smoke, other states still have a long way to go. Smoking rates as high as 30 percent still exist in some states.[26] Continued reductions in smoking rates are difficult to achieve and will require more targeted policies with information on how individual policies can best be implemented, a better capability to sort out the independent effects that different policies have on different populations, and an ability to determine how the effects of multiple policies interact with each other. Except for a small number of empirical studies that simultaneously considered the effect of two tobacco control policies,[27–29] most consider the effect of only one policy, thus making it difficult to learn about the interaction of multiple policies.

Researchers have resorted to simulation modeling to fill in the gaps left by empirical research. Simulation models that replicate past successes and evaluate performance of planned or current policies have been found to be useful in this setting. The Pacific Institute for Research and Evaluation lung model (SimSmoke), for example, has been applied to estimate the impact of tobacco control policies in three states, Arizona, California, and Kentucky.[30–32] Arizona and California have comprehensive tobacco control programs, and Kentucky has the dubious honor of being the state with the highest adult smoking prevalence. Evidence consistently shows that price increases and media policies have the largest impact on smoking rates, with substantially smaller components attributable to telephone quit lines, youth access policies, and clear air laws.

Many smoking cessation approaches, including counseling by phone or the Web and various pharmacotherapies, have been found to be promising. Comparative studies are being done to further understand differences in median survival time among never, former, and current smokers with NSCLC.[33] Proactive telephone counseling has helped smokers who are interested in quitting. Based on a meta-analysis of 48 trials from the Cochrane Tobacco Addiction Group trials register, among smokers who contacted help-lines, the quit rates were higher for groups randomized to receive multiple sessions of callback counseling (eight studies, more than 18,000 participants, odds ratio [OR] for long-term cessation was 1.41). Evidence of a dose response was found. One or two brief calls, however, were less likely to provide a measurable benefit. Three or more calls increased the odds of quitting compared with a minimal intervention, such as providing standard self-help materials, or brief advice, or compared with pharmacotherapy alone.

Telephone quit lines provided an important route of access to support for smokers, and callback counseling enhanced their usefulness.[34] The real-world utilization patterns and outcomes for 11,000 tobacco users enrolled in a comprehensive phone and Web-based tobacco cessation program, called "Free & Clear Quit For Life Program," found that Web utilization was significantly associated with increased call completion and tobacco abstinence rates at the six-month follow-up evaluation.[35] In another study (a 13-month follow-up of a six-arm randomized controlled trial of comparing Internet assistance for smoking cessation), 6,451 participants were followed. Findings showed that Internet assistance was an attractive and potentially cost-effective means to help individuals stop smoking. The findings also suggest that tailored, interactive Web sites may help cigarette smokers who do not report an indicator of depression at baseline to quit and maintain cessation.[36]

As empirical evidence on smoking cessation accumulates, simulation modeling is being developed to fill in the gaps and aid in projection. A Markov model for evaluation of various strategies and prediction, the Benefits of Smoking Cessation on Outcomes (BENESCO) model, was developed to simulate the lifetime direct costs and consequences of a hypothetical cohort of U.S. adult smokers who make a one-time attempt to quit smoking. The smoking cessation strategies compared were pharmacotherapy such as Varenicline and Bupropion, nicotine replacement therapy, and unaided quitting. The model used the hazard ratios from the Cancer Prevention Study (CPS-II) for the mortality of smoking-related diseases as a proxy to calculate the relative risks of the incidence and prevalence of these diseases. Varenicline was found to be superior to all other smoking cessation strategies that were investigated for both the 20-year and lifetime timeframe. It was predicted that if 25 percent of the current population of U.S. smokers made a one-time attempt to quit using Varenicline compared with unaided cessation, almost 144,000 smoking-related deaths and more than 261,000 cases of asthma exacerbations, chronic obstructive pulmonary disease, coronary heart disease, stroke, and lung cancer could be avoided compared with an unaided smoking cessation strategy.[37]

ADVANCES IN LUNG CANCER SCREENING

A screening test is a procedure that is performed to detect the presence of a specific disease in an asymptomatic population. There have been exhaustive reviews of

lung cancer screening techniques,[38–42] and all of these reports are in almost complete agreement that screening for lung cancer with either chest x-ray (CXR) or sputum cytology *is not* appropriate. Low-dose Computer Tomography (LDCT) scanning remains the most promising of lung cancer screening techniques, but these randomized trials are ongoing as of this writing and results are not expected for at least couple of years. Based on the available evidence, all national guideline committees have concluded that the evidence is insufficient at this time to recommend for or against screening asymptomatic persons for lung cancer using LDCT, CXR, sputum cytology, or a combination of these tests.

With an approved lung cancer screening program, it would be most appropriate not to screen all individuals from the general population, but rather to focus on current or former cigarette smokers. The issue of long-term survival, however, continues to bedevil researchers. Recent single-arm studies of lung cancer screening with computed tomography (CT) have shown that whereas screening detects more than twice as many early-stage lung cancers than would be expected to be detected without screening, improvement in mortality remains elusive. Single-arm study designs that compare a screened population versus an external nonscreened population do not definitively demonstrate reductions in either lung cancer–specific or all-cause mortality rates. The possibility of observing higher interval (for example, five-year) survival after diagnosis in the absence of a mortality reduction is explained by several well-known biases present in screening trial data: lead-time, length-time, and overdiagnosis biases. Because all three biases can contribute to longer survival of patients with screening-detected cancers, a control arm is critical for parsing out any true effect of screening on mortality.

The problem is not so much in detecting lung cancer, but in increasing survival time for those diagnosed with the disease. While awaiting the results of the randomized trials with appropriate control arms, some researchers are using a simulation modeling approach to integrate available data to evaluate screening programs and to inform those making screening decisions. That is, researchers who have evaluated the impact of CT screening on lung cancer outcomes through simulation modeling and have found that, although early stage lung cancer is possible to detect, improvement in the mortality rate is not significantly improved, perhaps due to other competing causes of death.[43–45]

The Lung Cancer Policy Model (LCPM) is a comprehensive microsimulation model of lung cancer development, disease progression, lung cancer detection, treatment results, and survival. Using LCPM modeling, findings showed that at six-year follow-up, the screening arm had an estimated 37 percent relative increase in lung cancer detection, compared with the control arm (9 percent at 15 years follow-up). The relative reduction in cumulative lung cancer–specific mortality from five annual screening examinations was 28 percent at six-year follow-up (15 percent at 15 years). The relative reduction in cumulative all-cause mortality from five annual screening examinations was 4 percent at six-year follow-up (2 percent at 15 years). Based on these findings, screening seems to reduce lung cancer–specific mortality slightly, but it may offer only a smaller reduction in overall mortality because of the increased competing mortality risks associated with smoking.

In the meantime, exciting alternatives to imaging technologies such as biomolecular marker screening and proteomics also are being developed for screening

purposes.[46–50] Furthermore, genomewide and candidate gene association studies of cigarette smoking behaviors are underway for finding genes that are associated with measures of smoking behavior, which could help in the development of targeted therapies for smoking cessation.[51]

In summary, findings have shown that patients whose early stage lung cancers were detected by CT screening and whose cancers were surgically resected have a 10-year survival in excess of 80 percent.[52] In the absence of screening, however, most lung cancers are diagnosed when distant spread already has occurred. Only 15 percent of patients with newly diagnosed lung cancer survive five years, primarily because the cancer is at a more advanced stage.[53] Long-term survival continues to be elusive despite gains made in detecting this form of cancer.

INCREASED UTILIZATION OF PATIENT-REPORTED OUTCOMES

Balancing the risks and benefits of potentially efficacious treatment with quality of life (QOL), including physical and psychosocial functioning, is an important component of cancer care. Many efforts have been made to assess the multidimensional entity of patient QOL as well as to help in the clinical decision-making process about treatment and in quantifying success of treatment regimens. A comprehensive review of tools for capturing QOL in lung cancer patients identified 50 instruments and recommended the best tool to be the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Lung Cancer Questionnaire (EORTC-LC13) in conjunction with the EORTC QLQ-C30 questionnaire developed to assess quality of life.[54] The Lung Cancer Symptom Scale (LCSS) [55] and the Functional Assessment of Cancer Therapy-Lung (FACT-L) [56] are two additional instruments with good reliability and validity. These quantitative measures of QOL increasingly are being used as a formal endpoint in clinical trials for deciding treatment efficacy and are playing a role in treatment decision-making.[57, 58]

In a multicenter randomized phase III trial, supportive care plus a combination chemotherapy was shown to be superior to supportive care only for both endpoints of QOL and survival.[59] Patients weigh the potential of survival benefit and chemotherapy-induced side effects in deciding their treatment. For example, in a study of 81 previously treated patients with platinum drugs for stage III/IV NSCLC, more than 50 percent said that they would not choose chemotherapy for an estimated additional survival of three months because the benefit was achieved at the expense of high-grade toxicity.[60] Patients often chose against a particular type of surgery or treatment after being presented with information about survival, side effects, and the expected QOL.[61]

EFFORTS TOWARD DEVELOPMENT OF PERSONALIZED MEDICINE: TARGETING THERAPY

In the field of lung cancer, emphasis has been placed not only on studying the "average" patient, but also studying those who experience long-term survival or who attain improved QOL. Using an understanding of genetics and genomic technologies, an area of research has emerged in which a subgroup of lung cancer patients with a particular pattern of biomarker expression or other clinical characteristics are

identified and then matched to a particular therapy, one that is expected to yield better outcomes in this particular subgroup. This personalized medicine approach is an exciting and promising avenue for lung cancer treatment.

An important observation was made by Potti et al. (2006) when they identified a gene expression profile that predicted the risk of recurrence in 89 patients with stage IA NSCLC.[11] This gene expression profile subsequently was combined with clinical variables in a classification-tree analysis to create an overall risk predictor, the lung metagene model. In two independent validation cohorts, the lung metagene model had overall predictive accuracy of 72 percent and 79 percent and performed better than clinical variables alone for clinical decision making. A large NCI-funded prospective phase III trial based on the lung metagene model is under way. In this study, patients who are predicted to be at high risk of recurrence will undergo resection of early stage disease and will receive further adjuvant chemotherapy. The hope is that this tailored approach will result in better overall survival and QOL compared with current standard practice among these early stage patients.

In addition to gene expression data, single gene markers such as ribonucleotide reductase M1 gene (RRM1) and excision repair cross-complementation 1 gene (ERCC1) also have been used to predict prognosis in NSCLC patients.[62] Gene expression signatures that distinguish primary from metastatic tumors and predict metastatic potential are being developed. Epidermal growth factor receptor (EGFR) inhibitors, such as Tarceva (erlotinib), have only a modest effect on overall survival of the "average" lung cancer patient but dramatically benefit 10 percent of those treated.[63] This observation has resulted in a successful patient-specific targeted therapy.

Better understanding of biologic mechanism for some being more susceptible to lung cancer than others and the knowledge of who will respond to certain therapy is helping those in the medical field make a strong stride toward the "personalized medicine" era. The price tag of these developments has been extremely high, however. Cost-effectiveness analysis needs to accompany these developments to ensure that society can pay for these developments. Efforts from national and international lung cancer consortiums with the aim of sharing comparable clinical and patient-reported data from ongoing lung cancer randomized clinical trial, case-control and cohort studies, as well as biological samples from different geographic areas and ethnicities are needed to achieve greater power (especially for subgroup analyses), to reduce duplication of research effort, to replicate novel findings, and to afford substantial cost savings through large collaborative efforts.

DEALING WITH DISPARITIES

Studies focusing on gender, race, and access to care differences have shown wide disparities in treatment received and survival.[64] Findings from a nested, matched case-control study of patients with NSCLC showed that African Americans were more likely to be smokers, have a lower per capita annual income, have a greater delay to treatment, and were less likely to agree to neoadjuvant therapy (assuming that it was offered). Whites had better five-year overall survival rates than African Americans for stage I (84 percent versus 78 percent, $p = 0.037$), stage II (52 percent versus 44 percent, $p = 0.041$), and stage III (32 percent versus 20 percent,

$p = 0.008$) NSCLC. This survival advantage disappeared for earlier stages of NSCLC (I and II) when adjusted for socioeconomic status and smoking status. The survival advantage for stage IIIA was lost when adjusted for neoadjuvant chemoradiotherapy.[65]

In another study evaluating the relationship of gender and race with the receipt of timely and clinically appropriate NSCLC treatment for each stage of diagnosis, substantial disparity was observed. Using Surveillance Epidemiology and End Result (SEER) data linked to Medicare claims for beneficiaries diagnosed with NSCLC between 1995 and 1999, findings showed that among stage I or II patients, women were 25 percent less likely to receive timely surgical resection relative to men, and African Americans were 66 percent less likely to receive timely and appropriate treatment compared with whites. African American men were least likely to receive resection (22.2 percent compared with 43.7 percent for white men); were 34 percent less likely to receive timely surgery, chemotherapy, or radiation for stage III disease; and were 51 percent less likely to receive chemotherapy in a timely fashion for stage IV disease relative to white males.[66] Many other studies have confirmed these findings and have concluded that the efforts in the last decade to mitigate cancer therapy disparities appear to have been unsuccessful.[67]

A limitation of using large administrative databases is that they do not contain information about the process by which the patient and physicians make decisions about the treatment and the patient's care experience. An important and potentially helpful initiative to address this gap is the $34 million NCI-funded Cancer Care Outcomes Research and Surveillance Consortium (CanCORS). CanCORS is a population-based cohort study of patients newly diagnosed with lung and colorectal cancer. CanCORS investigators are collecting data from patients, caregivers, physicians, and patient medical records from multiple regions across the United States and also are including data from different health care systems. Findings will supplement data from randomized clinical trials and hopefully will fill gaps in knowledge. Undoubtedly, the data will generate additional research questions toward measurement of effectiveness.

A CALL FOR COMPARATIVE EFFECTIVENESS RESEARCH

The ever-increasing cost of health care has provided the impetus to quantify the effectiveness and efficiency of treatment regimens. The United States spends more per citizen on health than any other country in the world. In 2007, total U.S. health expenditures reached $2.2 trillion ($7,421 per person), which translates to 16.2 percent of the nation's gross domestic product (GDP). At current growth rates, total health expenditures in the United States are estimated to account for 25 percent of the GDP by 2014.[68] The cost of treating cancer is particularly staggering. In 2004, the direct economic costs of cancer treatment in the United States, including inpatient and outpatient care, drugs, and medical devices, were estimated by the NCI to be $72.1 billion—representing just under 5 percent of U.S. spending for all medical treatment.[69] Research funding in 2008 totaled more than $247 million for lung cancer. In comparison, $274 million was allocated for colon and rectal cancer, $285.5 million for prostate cancer, and $572.5 million for breast cancer out of the total NCI budget of $4.83 billion. Given the amount of money spent on cancer

research, it is not surprising that there has *not* been a call for a quantitative accounting of the effectiveness of clinical regimens. Certainly, the recent financial crisis has provided the impetus for a more quantitative accounting of the efficacy and cost-effectiveness of treatment modalities.

While all of the new avenues of treatment for lung cancer are exciting and potentially useful, there is an urgent need to assess not only the effectiveness of these modalities, but also their cost effectiveness. CER is a means of evaluation to guide clinicians and policy makers in recommending the most efficient and effective treatments. CER can be defined as a comparison of the effectiveness of two or more interventions that are administered to a population of patients.[70] The objective is that decision making could be improved based on data obtained from comparative effectiveness studies.[71, 72] CER encompasses three concepts: (1) efficacy, (2) effectiveness, and (3) efficiency.[73]

- *Efficac*y refers to the extent to which an intervention does more good than harm, taking into account confounding factors. To what extent does the intervention "work"?
- *Effectiveness* refers to the extent to which an intervention does more good than harm when provided by physicians practicing in a clinical setting. That is, does it work in practice?
- *Efficienc*y measures the effect of an intervention in relation to the resources it consumes. That is, is it worth it? Most of the research efforts in lung cancer are concentrated on the endpoint of efficacy. Recently the Institute of Medicine (IOM) and others have raised the need for moving toward the endpoints of effectiveness and efficacy and their incorporation into clinical decision making and reimbursement decisions. Researchers in the lung cancer field have responded positively to this call.

The rationale for any cost-effectiveness analysis is that it offers an explicit and transparent approach to quantify the costs and benefits of a prevention or treatment strategy by using a common denominator (that is, years of life saved, Quality-Adjusted Life Years [QALYs] saved). The resulting incremental cost-effectiveness ratios (ICERs) then can be compared across conditions with each other, or with a threshold value, to identify the most efficient ways of maximizing health at the population level.[74] This approach has the potential advantage of facilitating a deliberative, systematic, and data-driven decision-making process for the allocation of public resources.

A recent review paper showed that most NSCLC therapies are cost-effective when the patient has an incremental cost-effectiveness ratio under $50,000 per life-year gained. Systematic analysis of the cost-effectiveness of treatments for the different stages of NSCLC, with particular emphasis on more recently approved agents, has suggested that, in patients with localized disease, adjuvant chemotherapy appears to have greater cost-effectiveness than observation (following patients after surgery without any adjuvant therapy). In locally advanced disease, combined modalities (chemotherapy, surgery, and radiotherapy) are probably cost-effective, but higher quality economic analyses are needed to demonstrate this. In advanced NSCLC, traditionally used doublets (cisplatin with docetaxel or gemcitabine; carboplatin with paclitaxel) and new agents (docetaxel, pemetrexed, and erlotinib) have acceptable cost-effectiveness.[75] Given the paucity of cost-effectiveness analyses among elderly lung cancer patients and patients with a poor prognosis, no

conclusions at this time can be drawn. In all likelihood, given the poor odds of survival among these individuals, no treatment would be considered cost-effective. Cost-effectiveness analysis in the context of SCLC is also not available. Moreover, it is not known whether physicians, patients, and payers utilize these data in treatment decision-making.

Recently, the IOM, Congressional Budget Office, the Blue Cross and Blue Shield Association, the Medicare Payment Advisory Commission, the Health Industry Forum, and other entities have called for a large expansion of CER.[76–78] A report from Friends of Cancer Research [79] was issued in 2009. The report was prepared to assess the extent to which evidence-based information on the effectiveness and efficiency of medical care included new diagnostic and therapeutic interventions. The report recommended that CER should be structured to ensure not only continuous learning, but also the rapid translation of the best available evidence into clinical practice. Recommendations were made for finding ways to accumulate evidence about all health care options for a given condition and included issues of access by patients not limited to geographic region, gender, race or ethnicity, and socioeconomic status. Additionally, the report recommended that data from public and private entities be linked to build on existing data collection and research capabilities, and most important, that there be a development of processes to allow information obtained through CER to be incorporated into clinical practice for use in clinical decision-making. Databases routinely established, maintained, and audited for clinical research contain detailed information about individual patients and their health outcomes; however, such data sets frequently are not configured to be readily combined with other data sets, or are proprietary to manufacturers.

To address the challenges of linking and sharing information from clinical databases, biospecimen repositories, and clinical researchers in the field of oncology, the NCI developed a biomedical informatics infrastructure to enable cancer researchers, physicians, and patients to share data and knowledge. The cancer Biomedical Informatics Grid™ (caBIG™) was established by NCI and its cancer centers as a pilot project in 2003.[80] The caBIG™ participants have developed readily disseminated standards, tools, and information systems for the management of clinical and research activities in oncology.[81] These tools should be disseminated more aggressively, and use of these tools should be mandatory for all stakeholders to allow for data standardization. Assessments of disparities in outcome by geographic region, gender, race or ethnicity, and socioeconomic status would be much easier to perform if the data from all regions are standardized. The goal is to reduce health disparities and to close the gap between the care that we know works well and the care patients actually receive.

CONCLUSION

Of all the cancers, lung cancer is probably one of the most difficult to diagnose at an early, treatable stage. Unlike cervical cancer, breast cancer, or even colon cancer, no cost-effective screening tool is available for population-based lung cancer screening. Yet, we have known for decades that tobacco smoking is the leading cause of this cancer, and evidence is quite clear that if individuals stopped smoking (or never started), the incidence of lung cancer would be greatly reduced. Lung

cancer can be prevented. Efforts to encourage people not to smoke, and if they do smoke, to quit, have accelerated over the past decade. Smoking bans in public and private buildings are helpful, but more work needs to be done to encourage people from starting to smoke in the first place.

Most recently, U.S. Congress passed the Family Smoking Prevention and Tobacco Control Act of 2009, which granted authority over tobacco products to the U.S. Food and Drug Administration (FDA). Some of the key elements of this legislation include a ban on candy-flavored cigarettes, a mandatory full disclosure of ingredients and additives in cigarettes, a stoppage on youth-focused marketing, a prohibition on marketing ads that mislead consumers, and a call for new warning labels. The act provides the FDA with authority to regulate the marketing and promotion of tobacco products and sets performance standards for tobacco products to protect the health of the public.

Screening utilizing imaging modalities have been found to be useful in detecting disease at early stage for breast and colon cancer, but the U.S. Preventive Services Task Force has concluded that the evidence is insufficient to recommend for or against screening asymptomatic persons for lung cancer with either LDCT, CXR, sputum cytology, or a combination of these tests. Although it has been demonstrated convincingly that low-dose, multi-detector-row CT is highly sensitive for the detection of SCLCs, uncertainty remains about whether screening with this modality will decrease lung cancer mortality and do so sufficiently to offset the harms and costs of screening. The results of two large, ongoing randomized clinical trials combined with simulation modeling should resolve this uncertainty in a few years and be useful in guiding the implementation of CT screening should it be proven to be effective.

REFERENCES

1. Horner M, et al., eds. *SEER Cancer Statistics Review, 1975–2006*. Bethesda, MD: National Cancer Institute; 2008. http://seer.cancer.gov/csr/1975_2006/. Accessed 2009.
2. Doll R. A study of the aetiology of carcinoma of the lung. *Br Med J*. 1952;2: 1271–1286.
3. Doll R. Mortality in relation to smoking: 20 years' observations on male British doctors. *Br Med J*. 1976;2:1525–1536.
4. Alberg A, Ford J, Samet J, American College of Chest Physicians. Epidemiology of lung cancer: ACCP evidence-based clinical practice guidelines. *Chest*. 2007; 132(3 Suppl):29S–55S.
5. Jemal A, et al. Annual report to the nation on the status of cancer, 1975–2005, featuring trends in lung cancer, tobacco use, and tobacco control. *J Natl Cancer Inst*. 2008;100(23):1672–1694.
6. Schachter E, Neuman T. Targeted therapies for the prevention of lung cancer. *Drugs Today*. 2007;43(12):897–936.
7. Massion P, et al. Smoking-related genomic signatures in non-small cell lung cancer. *Am J Respir Crit Care Med*. 2008;178(11):1164–1172.
8. Pfister D, et al. American Society of Clinical Oncology treatment of unresectable non-small-cell lung cancer guideline: update 2003. *J Clin Oncol*. 2004;22(2):330–353.
9. Pisters KM, et al. Cancer Care Ontario and American Society of Clinical Oncology adjuvant chemotherapy and adjuvant radiation therapy for stages I–IIIA resectable non small-cell lung cancer guideline. *J Clin Oncol*. 2007;25(34):5506–5518.

10. Kratz JR, et al. Genomic prognostic models in early-stage lung cancer. *Clin Lung Cancer*. 2009;10(3):151–157.
11. Potti A, et al. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med*. 2006;355(6):570–580.
12. Wang S. Utility of adaptive strategy and adaptive design for biomarker-facilitated patient selection in pharmacogenomic or pharmacogenetic clinical development program. *J Formos Med Assoc*. 2008;107(12 Suppl):19–27.
13. Wang SJ, O'Neill R, Hung H. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharm Stat*. 2007;6(3):227–244.
14. Yang P. Epidemiology of lung cancer prognosis: quantity and quality of life. *Methods Mol Biol*. 2009;471:469–486.
15. Martini N, et al. Factors influencing ten-year survival in resected stages I to IIIA non-small lung cancer. *J Thorac Cardiovasc Surg*. 1999;117:32–38.
16. Yang P, et al. Clinical features of 5,628 primary lung cancer patients: experience at Mayo Clinic from 1997–2003. *Chest*. 2005;128:452–62.
17. Nesbitt J, et al., eds. Cancer of the lung. In: Holland J, et al., eds. *Cancer Medicine*. Baltimore, MD: Williams and Wilkins; 2008:1723–1803.
18. Immerman S, et al. Site of recurrence in patients with stages I and II carcinoma of the lung resected for cure. *Ann Thorac Surg*. 1981;32:23–27.
19. Graham A, Potti A. Translating genomics into clinical practice: applications in lung cancer. *Curr Oncol Rep*. 2009;11(4):263–268.
20. Bordoni R. Consensus conference: multimodality management of early- and intermediate-stage non-small cell lung cancer. *Oncologist*. 2008;13(9):945–953.
21. US Department of Health and Human Services. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health*. Atlanta, GA: US Department of Health, Education, and Welfare; 1964.
22. Centers for Disease Control and Prevention. Tobacco Information and Prevention Source. http://www.cdc.gov/tobacco/research_data/adults_prev/adstat3.htm and http://www.cdc.gov/tobacco/research_data/adults_prev/tab_3.htm. Accessed October 10, 2004.
23. US Department of Health and Human Services. *Healthy People 2010*. Atlanta, GA: Centers for Disease Control and Prevention, Office of Disease Prevention and Health Promotion; 2000.
24. Levy D, Cummings K, Hyland A. Increasing taxes as a strategy to reduce cigarette use and deaths: results of a simulation model. *Prev Med*. 2000;31:279–286.
25. Centers for Disease Control and Prevention. State smoking restrictions for private-sector worksites, restaurants, and bars: United States, 2004 and 2007. *MMWR*. 2008;57(20):549–552.
26. Centers for Disease Control and Prevention. State-specific prevalence of current cigarette smoking among adults: United States, 2003. *MMWR*. 2004;53:1035–1037.
27. Chaloupka F, Wechsler H. Tobacco control policies and smoking among young adults. *J Health Econ*. 1997;16:359–373.
28. Farrelly M, Pechacek T, Chaloupka F. The impact of tobacco control program expenditures on aggregate cigarette sales: 1981–2000. *J Health Econ*. 2003;22:843–859.
29. Hu T, Sung H, Keeler T. Reducing cigarette consumption in California: tobacco taxes vs an antismoking media campaign. *Am J Public Health*. 1995;85:1218–1222.
30. Levy D, et al. The role of public policies in reducing smoking prevalence in California: results from the California tobacco policy simulation model. *Health Policy*. 2007;82(2):167–185.
31. Levy D, Mumford E, Gerlowski D. Examining trends in quantity smoked. *Nicotine Tob Res*. 2007;9(12):1287–1296.

32. Levy D, et al. The Kentucky SimSmoke tobacco policy simulation model: reaching Healthy People 2010 goals through policy change. *South Med J*. 2008;101(5): 503–507.
33. Ebbert J, et al. Lung cancer risk reduction after smoking cessation: observations from a prospective cohort of women. *J Clin Oncol*. 2003;21(5):921–926.
34. Stead L, Perera R, Lancaster T. Telephone counseling for smoking cessation. *Cochrane Database Syst Rev*. 2006;3:CD002850. doi: 10.1002/14651858. CD002850.pub2.
35. Zbikowski S, et al. Phone and web-based tobacco cessation treatment: real-world utilization patterns and outcomes for 11,000 tobacco users. *J Med Internet Res*. 2008;10(5):e41.
36. Rabius V, et al. Comparing internet assistance for smoking cessation: 13-month follow-up of a six-arm randomized controlled trial. *J Med Internet Res*. 2008;10(5):e45.
37. Caporaso N, et al. Genome-wide and candidate gene association study of cigarette smoking behaviors. *PLoS ONE*. 2009;4(2):e4653.
38. Black W. Computed tomography screening for lung cancer: review of screening principles and update on current status. *Cancer*. 2007;110(11):2370–2384.
39. Bach P, et al. American College of Chest Physicians screening for lung cancer: ACCP evidence-based clinical practice guidelines, 2nd edition. *Chest*. 2007; 132(3 Suppl):69S–77S.
40. Midthun D, Jett J. Update on screening for lung cancer. *Semin Respir Crit Care Med*. 2008;29(3):233–240.
41. Mazzone P, Mekhail T. Lung cancer screening. *Curr Oncol Rep*. 2007;9(4):265–274.
42. Field J, Duffy S. Lung cancer screening: the way forward. *Br J Cancer*. 2008;99(4): 557–562.
43. McMahon P, et al. Estimating long-term effectiveness of lung cancer screening in the Mayo CT screening study. *Radiology*. 2008;248(1):278–287.
44. Bach P. Lung cancer screening. *J Natl Compr Cancer Netw*. 2008;6(3):271–275.
45. Cancer Intervention and Surveillance Modeling Network. Washington, DC: National Institutes of Health; 2008. NIH Publication No.09-7354.
46. Anglim P, Alonzo T, Laird-Offringa I. DNA methylation-based biomarkers for early detection of non-small cell lung cancer: an update. *Mol Cancer*. 2008;7:81.
47. Brower V. Biomarker studies abound for early detection of lung cancer. *J Natl Cancer Inst*. 2009;101(1):11–13.
48. Church TR, et al. A prospectively measured serum biomarker for a tobacco-specific carcinogen and lung cancer in smokers. *Cancer Epidemiol Biomark Prev*. 2009;18(1):260–266.
49. Paci M, et al. Circulating plasma DNA as diagnostic biomarker in non-small cell lung cancer. *Lung Cancer*. 2009;64(1):92–97.
50. Sozzi G, et al. Plasma DNA quantification in lung cancer computed tomography screening: five-year results of a prospective study. *Am J Respir Crit Care Med*. 2009;179(1):69–74.
51. Yee J, et al. Connective tissue-activating peptide III: a novel blood biomarker for early lung cancer detection. *J Clin Oncol*. 2009;27(17):2787–2792.
52. Henschke C, Yankelevitz D. International Early Lung Cancer Action Program investigators: Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med*. 2006;355:1763–1771.
53. American Cancer Society. Cancer facts and figures 2007. http://www.cancer.org/downloads/STT/CAFF2007PWSecured.pdf. Accessed August 31, 2009.
54. Montazeri A, Gillis CR, McEwen J. Quality of life in patients with lung cancer: a review of literature from 1970 to 1995. *Chest*. 1998;113(2):467–481.

55. Hollen P, et al. Quality of life during clinical trials: conceptual model for the Lung Cancer Symptom Scale. *Support Care Cancer*. 1994;2:213–222.
56. Cella D, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Lung quality of life instrument. *Lung Cancer*. 1995;12:199–200.
57. Lee SM, et al. Comparison of gemcitabine and carboplatin versus cisplatin and etoposide for patients with poor-prognosis small cell lung cancer. *Thorax*. 2009; 64(1):75–80.
58. Ozturk A, et al. Evaluating quality of life and pulmonary function of long-term survivors of non-small-cell lung cancer treated with radical or postoperative radiotherapy. *Am J Clin Oncol*. 2009;32(1):65–72.
59. Helsing M, Bergman B, Thaning L, Hero U. Quality of life and survival in patients with advanced non-small-cell lung cancer receiving supportive care plus chemotherapy with carboplatin and etoposide or supportive care only: a multicentre randomized phase III trial. *Eur J Cancer*. 1998;34(7):1036–1044.
60. Silvestri G, Pritchard R, Weich HG. Preferences for chemotherapy in patients with advanced non-small-cell lung cancer: descriptive study based on scripted interviews. *Br Med J*. 1998;317:771–775.
61. Li W, Lee T, Yim A. Quality of life after lung cancer resection. *Thorac Surg Clin*. 2004;14:353–365.
62. Zheng Z, et al. DNA synthesis and repair genes RRM1 and ERCC1 in lung cancer. *N Engl J Med*. 2007;356(8):800–808.
63. Eberhard DA, et al. Biomarkers of response to epidermal growth factor receptor inhibitors in Non-Small-Cell Lung Cancer Working Group: standardization for use in the clinical trial setting. *J Clin Oncol*. 2008;26(6):983–994.
64. Leitner DW, et al. An epidemiologic approach to assessing primary care needs in rural Illinois. *J Rural Health*. 1996;12(2):110–119.
65. Bryant A, Cerfolio R. Impact of race on outcomes of patients with non-small-cell lung cancer. *J Thorac Oncol*. 2008;3(7):711–715.
66. Shugarman L, et al. Race and sex differences in the receipt of timely and appropriate lung cancer treatment. *Med Care*. 2009;47(7):774–781.
67. Gross C, et al. Racial disparities in cancer therapy: did the gap narrow between 1992 and 2002? *Cancer*. 2008;112(4):900–908.
68. Centers for Medicare and Medicaid Services. National health expenditure data. http://www.cms.hhs.gov/NationalHealthExpendData/02_NationalHealthAccountsHistorical.asp. Published January 6, 2009. Accessed August 31, 2009.
69. National Cancer Institute. Costs of cancer care: cancer trends progress report, 2007. http://progressreport.cancer.gov/doc_detail.asp?pid=1&did=2007&chid=75&coid=726&mid. Published November 16, 2007. Accessed August 31, 2009.
70. US Congressional Budget Office. *Research on the Comparative Effectiveness of Medical Treatments: Issues and Options for an Expanded Federal Role*. Washington, DC: US Congressional Budget Office; 2007. http://www.cbo.gov/ftpdocs/88xx/doc8891/12-18-ComparativeEffectiveness.pdf. CBO Paper Series.
71. Institute of Medicine. *Learning What Works Best: The Nation's Need for Evidence on Comparative Effectiveness in Health Care*. Washington, DC: National Academies of Science; 2007. http://www.iom.edu/~/media/Files/Activity%20Files/Quality/VSRT/ComparativeEffectivenessWhitePaperESF.pdf. Accessed August 31, 2009.
72. Institute of Medicine. Engineering a learning healthcare system: a look at the future: Institute of Medicine Workshop; April 29–30, 2008; Washington, DC. http://www.iom.edu/CMS/28312/RT-EBM/52747.aspx. Accessed August 31, 2009.
73. Cochrane A. *Effectiveness and Efficiency: Random Reflection on Health Services*. London: Nuffield Provincial Hospitals Trust; 1972.

74. Gold M, et al. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.
75. Chouaid C, et al. Economics of treatments for non-small cell lung cancer. *Pharmacoeconomics*. 2009;27(2):113–125.
76. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies of Science; 2001. http://www.iom.edu/?id=12736. Accessed August 31, 2009.
77. Institute of Medicine. *Knowing What Works in Health Care: A Roadmap for the Nation*. Washington, DC: National Academies Press; 2008.
78. Congressional Budget Office. *Research on the Comparative Effectiveness of Medical Treatments: Options for an Expanded Federal Role*. Washington, DC: Congressional Budget Office; 2007. http://www.cbo.gov/doc.cfm?index=8891. Accessed August 31, 2009.
79. Benson A, Lyerly K. Friends of cancer research committee: improving medical decisions through comparative effectiveness research: cancer as a case study. http://www.focr.org/comparative-effectiveness/. Published 2009.
80. National Cancer Institute. caBIG™ community website homepage. https://cabig.nci.nih.gov/. Accessed February 20, 2009.
81. National Cancer Institute. Cancer biomedical informatics grid: caBIG™ 2006 strategic plan. https://cabig.nci.nih.gov/working_groups/SP_SLWG/Documents/2006_caBIG_Strategic_Plan_Draftpercent20_post(2).pdf. Published April 6, 2006. Accessed March 15, 2009.

CHAPTER 10

Prostate Cancer Screening: Friend or Foe?

Carvell T. Nguyen, MD, and Michael W. Kattan, PhD, MBA

INTRODUCTION

The practice of medicine traditionally has been based on a strategy of enacting the most efficacious treatment available for a disease after it has been diagnosed. While the efficacy of modern medical treatment has improved greatly, some diseases, such as prostate cancer, still are associated with significant morbidity and mortality if found at an advanced stage. The contemporary approach to medicine now emphasizes early detection of disease, allowing capture of disease while it is still in a state of favorable biology (that is, low grade and stage), thus improving the chances of a cure. The purpose of screening for malignancies in the general population is to aid in the detection of disease at earlier stages, thereby improving the chances of cure and survival. As such, the search for effective screening assays for prostate cancer and other malignancies has become a critical part of advancing public health.

Advances in prostate cancer screening made over the last 20 years, including the discovery of biomarkers like prostate-specific antigen (PSA), have led to a dramatic shift in the diagnosis and management of prostate cancer. A stage migration effect has been observed, with the vast majority of patients now diagnosed with organ-confined disease.[1] The likelihood of cure with definitive treatment has increased, along with a concomitant reduction in the mortality rate.[2] A PSA-based prostate cancer screening, however, does not predict disease with perfect accuracy, and it is not able to predict the natural history of the tumors it detects. This is reflected by a lack of definitive data demonstrating any mortality benefit among men with screen-detected cancer. There is also concern that PSA screening has resulted in the overdiagnosis and overtreatment of many men with biologically indolent prostate cancer.

Despite these favorable outcomes, the adequacy and even the necessity of current screening strategies have been called into question. Although current screening tools have improved detection of prostate cancer, neither PSA testing nor any of the other screening tool predicts disease with perfect accuracy. PSA correlates only with the risk of cancer and does not detect the actual presence of malignancy. Moreover,

definitive evidence is lacking to demonstrate any survival or mortality benefit in screened men diagnosed and treated for prostate cancer. Indeed, given that the majority of men diagnosed with prostate cancer will not die of their disease, there is concern that screening has led to overdetection and overtreatment.[3] Consequently, these issues have stimulated much debate regarding the balance of benefit and harm associated with prostate cancer screening.

This chapter reviews the current status of prostate cancer screening and assesses its benefits and potential deleterious effects, with the aim of determining ways to improve its predictive accuracy and efficacy. To better understand the controversy surrounding prostate cancer screening and perhaps offer a solution, the chapter provides a review of current screening modalities, assesses their accuracy and utility in contemporary medical practice, and suggests future directions for the improvement of prostate cancer screening.

THE EPIDEMIOLOGY OF PROSTATE CANCER: AN OVERVIEW

Although prostate cancer is a major health burden throughout the world, there is a wide variation in its incidence. The highest rates are in the United States, Canada, Sweden, Australia, and France (48.1 to 137 cases per 100,000 person-years during 1988–1992); European countries have intermediate rates (23.9 to 31 cases per 100,000 person-years), and Asian countries the lowest rates (2.3 to 9.8 cases per 100,000 person-years).[4] Prostate cancer is the most common type of cancer found in U.S. men, other than skin cancer, and is second only to lung cancer as the most common cause of cancer death. Evidence is consistent across different racial and ethnic groups that a family history of prostate cancer increases the risk that a man will develop this cancer. This type of cancer is about twice as common among African American men as it is among white American men. In fact, African American men have the highest rates of prostate cancer in the world and have a 9.8 percent lifetime risk of developing this cancer. This is slightly higher than the 8 percent lifetime risk for American Caucasian men.[5]

Although men of any age can develop prostate cancer, it is found most often in men over the age of 50. In fact, more than 80 percent of men with prostate cancer are over the age of 65. Epidemiological studies have provided some insight about the cause of prostate cancer in terms of diet and genetic factors, but these risk factors need to be investigated further as the link to prostate cancer risk is not yet clear. Compared with other common cancers, such as breast and lung cancer, the causes of prostate cancer remain poorly understood.[6]

DIAGNOSING AND STAGING PROSTATE CANCER

Tumor grading of any cancer is a fundamental determinant of disease biology and prognosis. The Gleason score, the most widespread method of prostate cancer tissue grading used in the 21st century, is the single most important prognostic factor in prostate cancer. The Gleason staging system is used to evaluate the prognosis of men with prostate cancer. Together with other parameters, it is incorporated into a strategy of prostate cancer staging, which helps guide therapy. The Gleason grading system assigns a grade to each of the two largest areas of cancer in the tissue

samples. Grades range from one to five, with one being the least aggressive and five being the most aggressive.

Certain factors affect prognosis and treatment options, including the stage of the cancer (whether it affects part of the prostate, involves the whole prostate, or has spread to other places in the body); the patient's age and health status; and whether the cancer has just been diagnosed or has recurred.

TREATING PROSTATE CANCER: WHAT ARE THE OPTIONS?

Different treatment options are available and certainly are dependent on many factors, including stage of disease and patient preferences. Each has benefits as well as risks and none is considered to be superior to the other. Standard therapies include the following:

- **Watchful waiting:** a patient is closely monitored in absence of receiving any treatment until symptoms appear or change. This usually is an option used in older men with other medical problems and early stage disease.
- **Surgery:** several surgical options are available, with some being less invasive than others.
- **Radiation therapy:** treatment that uses high-energy x-rays or other types of radiation to kill cancer cells or keep them from growing. There are two types of radiation therapy: external radiation therapy that uses a machine outside the body to send radiation toward the cancer, and internal radiation therapy that uses a radioactive substance sealed in needles, seeds, wires, or catheters that are placed directly into or near the cancer. The way the radiation therapy is given depends on the type and stage of the cancer being treated.

New and exciting therapies are being tested to treat prostate cancer, including the following:

- **Hormone therapy:** treatment that removes hormones or blocks their action and stops cancer cells from growing. In prostate cancer, male sex hormones can cause prostate cancer to grow. Drugs, surgery, or other hormones are used to reduce the production of male hormones or block them from working.
- **Biologic therapy:** a treatment that uses the patient's immune system to fight cancer, biologic therapy represents a new avenue of treatment. Substances made by the body or made in a laboratory are used to boost, direct, or restore the body's natural defenses against cancer. This type of cancer treatment is also called biotherapy or immunotherapy.
- **High-intensity focused ultrasound:** this treatment uses ultrasound to destroy cancer cells. To treat prostate cancer, an endorectal probe is used to make the sound waves.

Whichever treatment choice, the risks (such as impotence and urinary problems including incontinence) as well as the potential benefits must be discussed with the patient. Each therapy has trade-offs.

NON-PSA SCREENING TOOLS

Before the advent of molecular biomarkers, digital rectal examination (DRE) was the traditional screening method for prostate cancer. DRE alone performs poorly in the early detection of prostate cancer with studies demonstrating that the

positive predictive value (PPV) of a palpable nodule on rectal examination ranges from 4 percent to 29 percent.[7–9] The cancer detection rate for DRE alone is approximately 2 percent, while multimodal screening with DRE, PSA, and transrectal ultrasound (TRUS) achieves rates from 5 percent to 15 percent.[10] A substantial proportion of DRE-detected prostate cancers (ranging from 30 percent to 80 percent) is not clinically localized at diagnosis,[11] a finding that likely contributed to the dismal cure rate for prostate cancer before PSA testing. With a lack of evidence showing any reduction in either mortality or metastatic disease with DRE-screening, and given that DRE findings are variable and operator-dependent, this method is now rarely used as a unimodal-screening tool.

Prostatic imaging has been evaluated as a potential screening strategy, but high cost and poor detection rates (for example, with computed tomography [CT] or magnetic resonance imaging [MRI]) generally have precluded their use in early detection of prostate cancer. TRUS has demonstrated the most potential of any imaging modality given its low cost, ease of use, and lack of radiation. TRUS, which has superior sensitivity and PPV compared with DRE, uses differences in echogenicity to differentiate neoplastic from normal prostate.[12, 13] Such ultrasonographic findings have low specificity for cancer as lesions can appear hypoechoic, hyperechoic, or isoechoic. Attempts to overcome the poor diagnostic performance of standard TRUS with more advanced modalities, including color-Doppler and contrast ultrasound, have improved both sensitivity and specificity for cancer detection, but not to levels that would justify TRUS as a primary screening tool.[14]

IMPACT OF PSA ON PROSTATE CANCER EPIDEMIOLOGY

The introduction of PSA testing over two decades ago revolutionized the approaches to the diagnosis and management of prostate cancer. Most dramatic has been the downward stage migration in prostate cancer that was observed since PSA became widely used as a screening tool. Whereas before PSA testing nearly half of patients were diagnosed with incurable disease, in the 21st century, greater than 90 percent of men who are diagnosed have organ-confined cancer that is presumably amenable to definitive treatment.[15] Incidence rates of prostate cancer also increased substantially early in the PSA era, a trend that was believed to be the result of detection of previously undiagnosed cases.[2]

Oncologic outcomes apparently have improved during the PSA era. The rate of cure following radical prostatectomy has increased,[16, 17] while the mortality rate associated with prostate cancer has decreased by almost 40 percent.[18] Whether these improved outcomes are a consequence of PSA screening or due to other factors, such as improved treatment efficacy or more aggressive patient selection for active therapy, remains unclear.

PERFORMANCE OF PSA SCREENING

The predictive accuracy of PSA testing is greater than any other currently available modality and truly has facilitated early detection of prostate cancer. Yet, PSA still does not predict cancer with perfect accuracy. Several studies have reported predictive accuracies for PSA testing alone in the range of 53 percent to 68 percent.[19–22] The sensitivity and specificity of PSA testing has been reported

to be around 70 percent and 90 percent, respectively, if the traditional cutoff of 4.0 ng/mL is utilized.[23] Although superior to DRE and TRUS, such performance characteristics are not ideal for a screening assay, and the majority of men with an "elevated" PSA, in fact, do not have cancer.

The suboptimal accuracy of PSA testing is likely related to the dichotomous cutoff system that has been used to define "normal" levels of PSA and the threshold at which a prostate biopsy should be performed. Such a system fails to account for the fact that PSA is a continuous variable rather than a categorical one. Indeed, data from the Prostate Cancer Prevention Trial (PCPT) showed that a significant number of men with PSA levels below the traditional threshold of 4.0 ng/mL can have prostate cancer—for example, nearly 7 percent of men with a PSA less than 0.5 ng/mL were diagnosed with cancer.[24] Based on such data, the authors of the PCPT concluded that PSA levels demonstrate a spectrum of prostate cancer risk for which there is no lower limit. In other words, there is essentially no PSA level below which a man's risk of prostate cancer is zero.

Several confounding factors also can affect PSA levels independently of a neoplastic process, thereby reducing its specificity as a screening tool. These include genetic and racial variations in PSA production, drug-induced changes (for example, with finasteride), and patient-specific factors such as body mass index.[25, 26]

VARIATIONS ON PSA-BASED SCREENING

Investigators have evaluated the performance of other tests that incorporate more sophisticated PSA-related parameters, hoping to improve on the diagnostic accuracy of standard PSA testing and decrease the number of patients who undergo unnecessary biopsies. Simple PSA cutoffs have demonstrated low specificity for prostate cancer screening. One reason may be the normal increase in PSA levels that occurs with age such that many older men can present with elevated PSA levels without harboring prostate cancer. Evidence indicates that using age-adjusted PSA levels reduces overdetection and false positives, but it does so at the potential cost of inducing unnecessary biopsies in younger men and missing cancers in older men.[27]

PSA velocity, which is the change in PSA level over the span of a year, also has been extensively studied, but the literature is equivocal regarding its predictive capability. Various methods for determining a significant PSA velocity have been evaluated, including percentage increases as well as actual rates. Smith and Catalona reported that a PSA velocity cutoff of 0.75 ng/ml per year maximized the sensitivity and specificity for predicting cancer in men with normal PSA levels.[28] Other investigators have reported that increasing PSA velocity is associated with increased PPV and specificity for cancer on biopsy.[29, 30] This appears to come at the cost of reduced overall sensitivity, however. Furthermore, multivariate analyses including PSA velocity with other prebiopsy variables, such as age, PSA, prostate volume, and DRE findings, have not found it to be an independent predictor of disease.[30]

Traditional PSA testing measures total PSA levels in the blood, which consists of both free and complexed forms of the protein. Assays are available to detect the free form of PSA, while complexed PSA is calculated by subtracting free PSA from total PSA. Fractionated PSA (complexed or free) has demonstrated a low specificity for prostate cancer, making it unsuitable as a standalone-screening assay.[31]

Investigators also have evaluated the utility of combining fractionated PSA with total PSA. Hoffman et al. performed a meta-analysis assessing the accuracy of free-to-total PSA ratio in predicting prostate cancer, finding that there was no percent-free PSA level that optimized both sensitivity and specificity of PSA screening.[32] Lower percent–free PSA has been linked with more aggressive prostate cancer (for example, associated with a larger tumor volume and a higher risk of extraprostatic spread) and may help identify patients who are more likely to progress and, therefore, benefit from intervention.[33]

PSA density, which is calculated by dividing serum PSA by prostate volume (derived from a TRUS examination), has been compared to standard PSA testing, but it has not been shown to be independently predictive of prostate cancer.[34, 35] PSA density of the transition zone alone also has been studied and may have greater utility than total PSA density. One study showed that it had the highest predictive accuracy for prostate cancer of any PSA variant,[36] while other data suggested that it may be useful in adjusting for benign causes of elevated PSA (for example, benign prostatic hyperplasia [BPH], or "enlarged prostate," a common condition that can mimic the symptoms of prostate cancer).[37]

PREDICTION MODELS AS SCREENING TOOLS

Predicting any clinical endpoint using only clinical judgment or single clinical variables is flawed because such simple methods of risk estimation are unable to fully account for the complex tumor biology and behavior of prostate cancer. This invariably leads to inaccurate predictions and inappropriate treatment assignment. To address this inadequacy, investigators have developed more complex and powerful multivariable models in an effort to improve the accuracy of cancer detection. Such prediction tools, which include risk groupings, probability tables, and nomograms, have seen widespread use in other areas of prostate cancer management (for example, predicting oncological outcomes after treatment).

A prediction tool that has been widely used in screening regimens is the PCPT risk calculator, which considers PSA in addition to age, ethnicity, family history, and DRE findings. The calculator, which has been validated on several external patient cohorts, is designed to provide individualized estimates of prostate cancer risk and determine the need for a biopsy. The calculator has enjoyed widespread clinical use despite demonstrating a predictive accuracy (57 to 70 percent) that is marginally greater than PSA testing alone (52 to 60 percent).[19–21, 24] A recent external validation of the risk calculator demonstrated reduced accuracy for predicting prostate cancer when it was applied to a more contemporary cohort of screened men.[20] Based on these findings, the authors emphasized the importance of continuously validating and refining prediction models as clinical practice evolves.

Nomograms currently represent the most accurate means of risk estimation and now are being used as screening tools.[38] Using data from the European Randomized Study of Screening for Prostate Cancer (ERSSPC), Roobol et al. developed a nomogram that predicts the chances of a positive biopsy by incorporating ultrasound volume, DRE, TRUS findings, and PSA level.[39] Compared with PSA alone, their nomogram increased the PPV for cancer detection and reduced the number of unnecessary biopsies by one-third. Although prediction modeling in the field of

prostate cancer screening is still in its nascent stages, these data show that it offers a potential solution to the suboptimal predictive accuracy of PSA testing alone.

CONTROVERSIES REGARDING PROSTATE CANCER SCREENING

Does prostate cancer screening save lives? This question is critical to the continued utilization of screening regimens for prostate cancer because the rationale for any screening effort is to improve survival and reduce the risk of death. Despite a reduction in prostate cancer mortality during the PSA era, until recently, there has been little definitive evidence suggesting that PSA screening was directly responsible for a survival or mortality benefit in patients with screen-detected prostate cancer. Interim data from two long-term screening studies addressing this very issue recently have been published, but their conflicting findings have done little to settle the debate. A report from the Prostate, Lung, Colorectal, and Ovarian (PLCO) screening trial found no significant difference in prostate cancer death rates between men who were screened and those who were not.[40] In contrast, the ERSSPC reported a 20 percent reduction in the mortality rate of men who were screened with PSA compared with controls.[41]

The apparent contradiction between the findings of the PLCO and ERSSPC trials may be explained by a substantial degree of prescreening found in the PLCO population. Nearly half of the men enrolled in the PLCO trial had either a PSA test (45 percent) or a DRE (55 percent) before study entry, and half of the men in the control arm underwent some PSA testing during the actual trial itself. These confounders could have diluted the effect of any screening that occurred as part of the trial. In any case, definitive conclusions regarding the mortality benefit of PSA screening will have to await the final analyses of these trials.

The uncertainty surrounding the ability of screening to affect oncological outcomes reflects what is perhaps the greatest deficiency of current PSA-based screening: the inability to characterize the tumor biology and natural history of the cancers that are detected by PSA. Even among patients who initially are diagnosed with organ-confined disease, we cannot reliably distinguish those whose disease will remain localized from those whose cancer may progress or cause mortality. This variable history of prostate cancer ultimately determines the need for and efficacy of any therapeutic intervention. As such, the ability to accurately predict which cancers will progress can facilitate rational and appropriate application of definitive therapy and limit unnecessary exposure to treatment-related side effects, many of which are not trivial.

It is this potential for inflicting harm without providing a concomitant benefit in a substantial percentage of men with screen-detected cancer that has caused a critical reevaluation of the utility of PSA screening. Indiscriminate use of PSA screening may have led to the detection of a disproportionate number of clinically indolent prostate cancers that are unlikely to adversely affect a man's health or life span. Indeed, a recent analysis of Surveillance, Epidemiology and End Results (SEER) Program data regarding prostate cancer incidence between 1986 and 2005 found that PSA screening resulted in the additional detection and treatment of more than a million men.[42] Assuming that the observed decline in prostate cancer mortality during this same period was due to screening, the authors found that nearly 20 men had to be diagnosed and treated for each death that was prevented. This has

been further corroborated by the PLCO and ERSSPC studies, both of which reported very low prostate cancer death rates, suggesting that the majority of prostate cancer is unlikely to progress. The conclusion, then, is that the majority of men who are diagnosed with prostate cancer are unlikely to benefit from treatment.

The ethical and economic implications of this overdiganosis and overstatement are profound and may have grave consequences for the viability of PSA-based screening. Men who are treated unnecessarily for indolent cancers may suffer significant harm in the form of treatment-related complications, such as impotence and incontinence. And even if a man with screen-detected cancer eschews intervention, he still must deal with the psychological burden of a cancer diagnosis. The cost-effectiveness of prostate cancer screening also is quite suspect. Based on data from the SEER, PLCO, and ERSSPC studies, the number of men who must be screened and treated to prevent just one cancer-related death is significant. The cost associated with their work-up, treatment, hospitalization, and follow-up is likely to be substantial, adding yet more burden to an already overstretched health care system.

FUTURE DIRECTIONS

PSA-based screening has had a profound impact on the epidemiology and management of prostate cancer, but it still does not predict disease with perfect accuracy. Moreover, it has resulted in the overdiagnosis and overtreatment of clinically insignificant prostate cancers. Accurate identification of such indolent cancers is important for counseling and rational application of definitive treatment. Because PSA alone likely has reached the limits of its diagnostic potential, new screening modalities based on novel biomarkers or more sophisticated prediction models represent the best hope for improving the accuracy and utility of prostate cancer screening.

Novel urinary and serum markers, including PCA3 and EPCA-2, have been found to correlate with the presence of cancer; in fact, both of these markers demonstrated superior accuracy in the prediction of prostate cancer when compared with PSA.[43, 44] There is also emerging data suggesting that certain genetic variants (for example, single nucleotide polymorphisms) may predict cancer in some subpopulations of men (for example, those with a family history).[45, 46] However, it is not enough to simply detect cancer in its earliest stages. Additional markers and prognostic factors also are needed to determine which cancers will progress to become lethal and help target therapy to those patients who truly need it.

In 2006, a novel gammaretrovirus, called xenotropic murine leukemia-related virus (XMRV), was isolated from prostatic tissue of patients with prostate cancer who harbored a certain mutation in RNase L, a protein with antiviral activity.[47] Although no etiologic link has yet been established, a recent study confirmed that XMRV infection is indeed associated with prostate cancer and also demonstrated a correlation between XMRV positivity and higher grade tumors.[48] Moreover, the investigators found that XMRV infection was not limited to men with RNase L mutations. Taken together, these data suggest that testing for XMRV may detect more aggressive cancers that pose significant harm to a man during his lifetime, and that such a test can be applied to a more generalized population (that is, not just men with a specific mutation). Obviously, further study is warranted to determine whether any of these tests can replace or augment current PSA-based screening.

As discussed, prediction models offer a promising alternative to standard PSA testing in the early detection of prostate cancer and also may have potential utility in the prediction of indolent disease. Currently available models however, do not predict perfectly (for example, the PCPT risk calculator), require biopsy pathological data,[49–51] or were based on clinical patient series from single institutions, thus limiting their generalized application. To be useful as a screening tool, new prediction models should consider only prebiopsy data, incorporate novel biomarkers that are specific for higher grade cancer, and be based on large, multi-institutional datasets (thus increasing their generalizability).

CONCLUSION

Current evidence suggests that PSA-based screening for prostate cancer is a double-edged sword. PSA testing undoubtedly has improved the early detection of prostate cancer, leading to more men being diagnosed and treated for prostate cancer. This, however, has not translated into a definitive survival or mortality benefit for men with screen-detected cancer and actually may cause harm. The increased incidence of prostate cancer includes a large number of indolent tumors that do not pose a significant health threat and do not require treatment. According to recent estimates in the literature, approximately 1 million men have been subject to over-diagnosis and overtreatment during the PSA era, unnecessarily exposing them to treatment-related side effects, psychological distress, and financial costs.

Unfortunately, current PSA-based screening modalities cannot distinguish the minority of aggressive tumors from the majority of cancers that will remain localized and asymptomatic. Novel biomarkers and prognostic factors that can predict the natural history of a given patient's cancer are required to selectively and rationally apply definitive treatment to those patients who actually need it. The future of prostate cancer screening will be based on individual risk estimation afforded by more powerful multivariable prediction models (for example, nomograms) that incorporate these predictive factors. Until then, we must continue to improve the accuracy of PSA-based screening, appropriately counsel prostate cancer patients regarding the risks of disease progression versus treatment, and be judicious in the application of definitive therapy.

REFERENCES

1. Catalona WJ, et al. Detection of organ-confined prostate cancer is increased through prostate-specific antigen-based screening. *JAMA*. 1993;270:948.
2. Horner MJ, et al., eds. *SEER Cancer Statistics Review: 1975–2006*. Bethesda, MD: National Cancer Institute; 2009.
3. Bill-Axelson A, et al. Radical prostatectomy versus watchful waiting in early prostate cancer. *N Engl J Med*. 2005;352:1977.
4. Hsing A, Tsao L, Devesa S. International trends and patterns of prostate cancer incidence and mortality. *Int J Cancer*. 2000;85:60–67.
5. American Cancer Society. http://www.cancer.org.
6. Gronberg H. Prostate cancer epidemiology. *Lancet*. 2003;361(9360):859–864.
7. Chodak GW, Keller P, Schoenberg HW. Assessment of screening for prostate cancer using the digital rectal examination. *J Urol*. 1989;141:1136.

8. Chodak GW, Schoenberg HW. Early detection of prostate cancer by routine screening. *JAMA*. 1984;252:3261.
9. Schroder FH, et al. Evaluation of the digital rectal examination as a screening test for prostate cancer: Rotterdam section of the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst*. 1998;90:1817.
10. Cooner WH, et al. Prostate cancer detection in a clinical urological practice by ultrasonography, digital rectal examination and prostate specific antigen. *J Urol*. 1990; 143:1146.
11. Wajsman Z, Chu TM. Detection and diagnosis of prostate cancer. In: Murphy GP, ed. *Prostatic Cancer*. Littleton, MA: PSG Publishing Company; 1987:91–99.
12. Lee F, et al. Transrectal ultrasound in the diagnosis of prostate cancer: location, echogenicity, histopathology, and staging. *Prostate*. 1985;7:117.
13. Lee F, Torp-Pedersen ST, Siders DB. The role of transrectal ultrasound in the early detection of prostate cancer. *CA Cancer J Clin*. 1989;39:337.
14. Sedelaar JP, et al. Transrectal ultrasound in the diagnosis of prostate cancer: state of the art and perspectives. *Eur Urol*. 2001;40:275.
15. Murphy GP, et al. The national survey of prostate cancer in the United States by the American College of Surgeons. *J Urol*. 1982;167:928.
16. Han M, Partin AW, Piantadosi S, Epstein JI, Walsh PC. Era specific biochemical recurrence-free survival following radical prostatectomy for clinically localized prostate cancer. *J Urol*. 2001;166:416.
17. Han M, et al. Biochemical (prostate specific antigen) recurrence probability following radical prostatectomy for clinically localized prostate cancer. *J Urol*. 2003; 169:517.
18. Lu-Yao G, Stukel TA, Yao SL. Changing patterns in competing causes of death in men with prostate cancer: a population based study. *J Urol*. 2004;171:2285.
19. Hernandez DJ, et al. Predicting the outcome of prostate biopsy: comparison of a novel logistic regression-based model, the prostate cancer risk calculator, and prostate-specific antigen level alone. *Br J Urol Int*. 2009;103:609.
20. Nguyen CT, et al. Performance of the Prostate Cancer Prevention Trial risk calculator in a contemporary cohort of men screened for prostate cancer and diagnosed with extended prostate biopsy schemes. *J Urol*. 2010;183(2):529–533.
21. Parekh DJ, et al. External validation of the Prostate Cancer Prevention Trial risk calculator in a screened population. *Urology*. 2006;68:1152.
22. Thompson IM, et al. Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower. *JAMA*. 2005;294:66.
23. Partin AW, Oesterling JE. The clinical usefulness of prostate specific antigen: update 1994. *J Urol*. 1994;152:1358.
24. Thompson IM, et al. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. *J Natl Cancer Inst*. 2006;98:529.
25. Ahn JO, Ku JH. Relationship between serum prostate-specific antigen levels and body mass index in healthy younger men. *Urology*. 2006;68:570.
26. Chia SE, et al. Effect of ageing and body mass index on prostate-specific antigen levels among Chinese men in Singapore from a community-based study. *Brit J Urol Int*. 2009;103:1487.
27. Etzioni R, Cha R, Cowen ME. Serial prostate specific antigen screening for prostate cancer: a computer model evaluates competing strategies. *J Urol*. 1999;162: 741–748.
28. Smith DS, Catalona WJ. Rate of change in serum prostate specific antigen levels as a method for prostate cancer detection. *J Urol*. 1994;152:1163.

29. Berger AP, et al. Large-scale study of clinical impact of PSA velocity: long-term PSA kinetics as method of differentiating men with from those without prostate cancer. *Urology*. 2007;69:134.
30. Wolters T, Roobol MJ, Bangma CH, Schroder FH. Is prostate-specific antigen velocity selective for clinically significant prostate cancer in screening? European Randomized Study of Screening for Prostate Cancer. *Eur Urol*. 2009;55:385.
31. Brawer MK, et al. Measurement of complexed PSA improves specificity for early detection of prostate cancer. *Urology*. 1998;52:372.
32. Hoffman RM, et al. Using the free-to-total prostate-specific antigen ratio to detect prostate cancer in men with nonspecific elevations of prostate-specific antigen levels. *J Gen Intern Med*. 2000;15:739.
33. Arcangeli CG, et al. Percentage of free serum prostate-specific antigen as a predictor of pathologic features of prostate cancer in a screening population. *Urology*. 1998;51:558.
34. Babaian RJ, Kojima M, Ramirez EI, Johnston D. Comparative analysis of prostate specific antigen and its indexes in the detection of prostate cancer. *J Urol*. 1996;156:432.
35. Bangma CH, Grobbee DE, Schroder FH. Volume adjustment for intermediate prostate-specific antigen values in a screening population. *Eur J Cancer*. 1995; 31A:12.
36. Djavan B, et al. Combination and multivariate analysis of PSA-based parameters for prostate cancer prediction. *Tech Urol*. 1999;5:71.
37. Horninger W, et al. Improvement of specificity in PSA-based screening by using PSA-transition zone density and percent free PSA in addition to total PSA levels. *Prostate*. 1998;37:133.
38. Ross PL, et al. Comparisons of nomograms and urologists' predictions in prostate cancer. *Semin Urol Oncol*. 2002;20:82.
39. Roobol MJ, et al. A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer. *Eur Urol*. 2009;57(1):79–85.
40. Andriole GL, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med*. 2009;360:1310.
41. Schroder FH, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med*. 2009;360:1320.
42. Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986–2005. *J Natl Cancer Inst*. 2009;101(19):1325–1329.
43. Marks LS, et al. PCA3 molecular urine assay for prostate cancer in men undergoing repeat biopsy. *Urology*. 2007;69:532.
44. Leman ES, et al. EPCA-2: a highly specific serum marker for prostate cancer. *Urology*. 2007;69:714.
45. Witte J S. Prostate cancer genomics: towards a new understanding. *Nat Rev Genet*. 2009;10:77.
46. Zheng SL, et al. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med*. 2008;358:910.
47. Urisman A, et al. Identification of a novel gammaretrovirus in prostate tumors of patients homozygous for R462Q RNASEL variant. *PLoS Pathog*. 2006;2:e25.
48. Schlaberg R, Chloe DJ, Brown KR, Thaker HM, Singh IR. XMRV is present in malignant prostatic epithelium and is associated with prostate cancer, especially high-grade tumors. *Proc Natl Acad Sci U S A*. 2009;106(38):16351–16356.

49. Epstein JI, et al. Nonpalpable stage T1c prostate cancer: prediction of insignificant disease using free/total prostate specific antigen levels and needle biopsy findings. *J Urol*. 1998;160:2407.
50. Steyerberg EW, et al. Prediction of indolent prostate cancer: validation and updating of a prognostic nomogram. *J Urol*. 2007;177:107.
51. Roemeling S, et al. Nomogram use for the prediction of indolent prostate cancer: impact on screen-detected populations. *Cancer*. 2007;110:2218.

CHAPTER 11

Advances in Skin Cancer Prevention, Screening, and Treatment

Marianne Berwick, PhD, MPH, Esther Erdei, PhD, Melissa Gonzales, PhD, Salina Torres, PhD, and Kristina Flores, PhD

INTRODUCTION

The skin is the largest organ in the human body composed of three layers; the epidermis (outer layer that serves as a barrier to the elements and to infection, as well as making new skin cells to replace the dead skin cells on the surface of the skin), the dermis (consists of connective tissue and cushions the body from stress and strain), and the hypodermis (not technically part of the skin, the hypodermis' purpose is to attach the skin to underlying bone and muscle as well as to supply it with blood vessels and nerves). The skin plays an important role in protecting the body against infection, injury, and excessive water loss (it is a water-resistant barrier), helps to regulate temperature, acts as a storage center for lipids, serves as a synthesis of vitamin D, and also supports its own ecosystem of microorganisms, including yeasts and bacteria. Skin pigmentation is determined primarily by the amount and type of melanin (which determines the skin tone) that can vary from very dark brown-black to light-skinned white. Melanin also offers protection against ultraviolet (UV) rays; fair-skinned individuals are much less protected and much more susceptible to sunburn and to skin cancer.

Over the past decades, the incidence of skin cancer has increased substantially. This chapter focuses on the epidemiology of skin cancer and illustrates the growing public health burden of this particular form of cancer. Advances in screening and treatment are discussed and preventive measures are explored.

EPIDEMIOLOGY

Skin cancer is the most common cancer in the United States, estimated at more than 1 million newly diagnosed cases each year, close to the total new cases of all other cancer combined.[1] Worldwide, it has been estimated that every year 2.75 million new cases of nonmelanoma skin cancer (NMSC) will be diagnosed.[2] The incidence of cutaneous malignant melanoma also is increasing worldwide.[3–5] Current estimates place the number of new cases of melanoma at 160,177 worldwide

and deaths from melanoma at 40,781.[6] In Australia, where skin cancer is quite prevalent, estimates for melanoma incidence range from 0.59 per 100,000 in dark-skinned populations to 40.5 per 100,000 in light-skinned populations. Although most skin cancers generally are considered to be benign, melanoma is considered the most deadly of the skin cancers. In the United States, mortality has been steady at 3.1 percent per year between 1992 and 2004, the last year for which accurate statistics are available.[7]

TYPES OF SKIN CANCER

Most skin cancer has a higher incidence among light-skinned individuals and is less common among dark-skinned individuals. Individuals with darker skin color produce more melanin, which is effective at blocking solar rays that can cause skin damage and skin cancer. Light-skinned individuals produce melanin only in the presence of sunlight, and after the UV rays have penetrated the lower portion of the outer skin layer effectively causing skin damage.

The three most common types of skin cancer are basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and cutaneous malignant melanoma (CMM). BCC and SCC often are referred to as NMSC. Each type of skin cancer seems to develop from different patterns of sun exposure, with SCC developing after a high level of continuous sun exposure, CMM after intermittent sun exposure, and BCC somewhere in between.

Basal Cell Carcinoma

BCC develops from epithelial keratinocytes in the basal layer of the skin. It is the most common form of skin cancer and accounts for more than 90 percent of all skin cancer in the United States. These cancers almost never metastasize to other parts of the body and generally appear on the head, neck, arms, or back. There is no known "precursor" lesion for BCC, so new molecular discoveries are important to clarify its etiology and thus the ability to prevent this form of NMSC.

Squamous Cell Carcinoma

SCC also develops from keratinocytes and certain pathological alterations, such as actinic or solar keratoses, which are considered "precursor" lesions. These lesions frequently occur on the face, hands, and forearms, and commonly are detected among individuals older than 40 years of age. It has been estimated that much of the time actinic keratoses spontaneously regress, and only a few develop further to SCC.

The incidence of BCC and SCC is difficult to estimate accurately because of the difficulty of obtaining consistent and reliable data from outpatient clinics and physician offices where these cancers usually are diagnosed. Moreover, NMSCs often are not counted by tumor registries. That being said, in 2006 in the United States, more than 1 million cases of BCC and SCC cancer were diagnosed.[1] Among dark-skinned individuals, the incidence rates of BCC and SCC are much lower than among light-skinned Caucasians. Further, evidence shows that NMSC is being diagnosed more frequently among younger individuals, which had not been the norm in the past.[8, 9]

The U.S. National Cancer Institute sponsored a population-based skin cancer survey in 1977–1978 that found noticeable geographic variability in the NMSC incidence rates within the United States.[10] In 1998–1999, a follow-up NMSC survey in New Mexico, one of the original sites, showed that the incidence rate of BCC had increased by 50 percent in males and 20 percent in females and the incidence rate of SCC roughly doubled for both males and females.[11] Trends in Canada, on the other hand, have shown more modest annual percent changes in the rates of BCC and SCC: an increase of 2.4 percent from the early 1970s to 2000.[12] Although most analyses report an increase in incidence, some have reported a decline in SCC rates.[13] Explanations for this are not clear. In Australia, however, high incidence rates of BCC and SCC have been reported (1,170 per 100,000 population), and the increases have been the greatest for people age 60 years and older.[14] The medical, public health, and economic implications of the increase are significant.[15]

The risk of developing subsequent tumors appears to be increased among those diagnosed with NMSC. An increased risk of second primary cancer after a diagnosis of NMSC has been observed in the first four years after diagnosis; elevated risk remained higher in all age-groups up to 75 years of age.[16] In addition, individuals with a diagnosis of NMSC are reported to have a twofold increased risk of dying from other solid tumors.[17] So even though BCC and SCC are not considered melanoma skin cancer, the more deadly type of skin cancer, individuals with these cancers are at elevated risk of developing other tumors, which in some cases can lead to death.

Cutaneous Malignant Melanoma

Melanoma occurs when melanocytes (pigment cells) become malignant; approximately 40 percent of these are associated with common nevi (moles). Although melanoma can occur in other parts of the body, it is most commonly found on the skin. In men, melanoma is often found on the trunk (the area between the shoulders and the hips) or the head and neck. In women, this cancer often develops on the lower legs. Melanoma is rare in dark-skinned individuals, but quite common among light-skinned individuals. Whereas melanoma is more often diagnosed among the better educated and economically well off, disadvantaged individuals—those with less education and who are less well off—are more likely to have advanced disease once the diagnosis is made.[3] This difference might be due to the fact that the better educated are more likely to have health insurance and a greater access to physicians and tend to be screened for skin cancer compared with their less-well-off counterparts.

A large proportion of melanomas appear to evolve through a slow-growing, radial growth phase that can develop a vertical growth phase leading to metastases. Other melanoma lesions, most commonly nodular melanomas as well as some superficial spreading melanomas, can arise rapidly and without nevus involvement. When melanoma spreads, cancer cells may show up in nearby lymph nodes, an indication that cancer cells probably have spread to other parts of the body, such as the liver, lungs, or brain. In such cases, the cancer cells in the new tumors are still considered melanoma cells, and the disease is called metastatic melanoma. The prognosis is uniformly poor once the cancer has spread to distant organs.

Alexander Breslow, in 1975, observed that as the thickness of the tumor increases, the chance of survival decreases. The Breslow thickness is a common

measure used around the world to characterize and classify this type of skin cancer. The depth of a melanoma lesion is measured from the basement membrane of the epidermis to the deepest identified melanoma tumor cells. This thickness is the most important prognostic factor for melanoma, as deeper melanomas metastasize more easily through the blood and lymph systems. The thickness of a melanoma is related to the five-year survival rate after surgical removal of the tumor. For example, a thickness of the melanoma of less than 0.76 millimeters is associated with a five-year survival for 97 percent of patients, whereas a tumor thickness of more than 8.0 millimeters is associated with a five-year survival of 32 percent. By far the majority of melanomas (70 percent) are diagnosed at very thin (less than 1 mm Breslow thickness) and highly curable stages. The prognosis for those with a Breslow thickness greater than 1 millimeter is uniformly poor. To date, no satisfactory cures have been found for advanced melanoma.

Epidemiological studies repeatedly have shown that exposure to sunlight is the major risk factor for the development of melanoma, especially in individuals with a fair-skinned complexion.[18–21] The association between sun exposure and melanoma is complex, however; there is not a clear dose-response association of sun exposure with melanoma. The effects of sun exposure and its relationship with other factors, such as latency period, body site exposure, distribution of cutaneous melanoma, histogenetic melanoma types, and other factors, are not well understood.[22–24] In addition to sun exposure, pigmentary traits (hair color, eye color, and sun sensitivity), the presence and number of melanocytic and atypical nevi, as well as variants in DNA (deoxyribonucleic acid) repair genes and tumor suppressor genes, also have been shown to increase melanoma risk in susceptible populations.[25–31]

The burden of this disease in the United States is growing. More than 62,000 new cases of melanoma were diagnosed in 2008.[5] Although the number of new cases appears to be increasing, there is good news insofar as the mortality rates appear to be decreasing, especially among younger cohorts of melanoma patients.[6] It is fair to say that the increase in the incidence could be attributed to better screening—for example, the increase could be artifactual as early screening and detection has led to an increase in the number of biopsies performed. In the 21st century, doctors are finding lesions that 20 years ago would never have been called melanoma.[32] The problem is that scientists are not able to differentiate between melanomas that will progress from melanomas that are indolent or from those that are more deadly. A similar issue applies to breast and prostate cancer screening and detection. With melanoma, however, the causes for increased risk for this cancer are generally understood; therefore, it is imperative that research continues to investigate the best means for prevention.

RISK FACTORS FOR CUTANEOUS MALIGNANT MELANOMA

Melanocytic nevi are the strongest risk factor for the development of melanoma, although they appear to have little relationship to the development of NMSC. Studies have shown that individuals with a higher than average number of moles have the highest risk for melanoma. The way in which nevi might be involved has been debated. One theory is that nevi are actually on the causal pathway and that some develop into melanoma. Supporting this theory is the finding that

approximately 40 percent of melanomas have an adjacent nevus or nevus remnant still observable by pathology.[33] Whiteman suggested that individuals with fewer nevi require repeated exposure to sunlight to drive carcinogenesis because they are more likely to develop melanomas on high-sun-exposure areas, such as the head and neck, while those with many nevi, who are more likely to develop melanoma of the trunk, may possess host factors that drive carcinogenesis after minimal sunlight exposure.[23] Therefore, it is plausible that the oncogenic pathway characterized by pigment cell instability is associated with a higher number of nevi and may lead to more aggressive melanoma and a worse prognosis. In any case, the available data indicate that having more nevi than average, or having dysplastic nevi, substantially increases the risk for melanoma.[28, 34]

Risk factors for skin cancer generally can be divided into environmental and lifestyle factors as well as genetic factors. The search to understand the interaction of the two has been accelerated with advances in genomics. The most predominant *environmental risk factor* for melanoma is ultraviolet radiation (UVR). The UVR wavelengths primarily responsible for the development of skin cancers are in the ultraviolet B (UVB) 280–320 nm and ultraviolet A (UVA) 320–400 nm range. Exposure to UVR is the main cause of both melanoma cancers and NMSCs with UVR from sunlight implicated as the main environmental agent responsible for the initial transformation of benign melanocytes into melanoma.[35]

UVR induces skin cancers by three mechanisms: direct DNA damage leading to mutations; production of activated oxygen molecules that in turn damage DNA and other cellular structures; and localized immunosuppression blocking the body's natural anticancer defenses. Early research focused on UVB, in the belief that this component of natural light was more important in carcinogenesis.[35] Recent work recognizes the role of UVA as well.[36] Far more UVA reaches melanocytes than UVB. On average, the epidermal layers overlaying the basal layer in Caucasian skin absorb 56 percent of the UVB and only 27 percent of UVA.[37] As UVB is absorbed in the epidermis by various molecules such as the keratins and DNA, it can suppress immune reactions, induce tolerance to antigens, upregulate gene expression, and induce mutations.[38] UVB directly mutates DNA [39–41] and is demonstrated to initiate CMM in genetically engineered mice.[42]

UVB also plays an important role in stimulating photoprotective adaptation of the skin. UVB-induced mutations (thymidine dinucleotides) in the epidermis are believed to stimulate a photoprotective response (PER), which includes the synthesis and release of melanosomes by melanocytes.[43–45] This in turn reduces the penetration of UV radiation to the basal epidermis and melanocytes. PER includes the proliferation of keratinocytes, leading to a thickening of the stratum corneum, improved scattering of UV radiation, and reduced UV penetration of the skin. Under certain conditions (for example, high latitudes where UVB flux is low or when UVB blocking sunscreens are used), the natural protective epidermal response from UVB exposure is reduced and the basal epithelium, including the melanocytes, is exposed to a relatively large flux of UVA photons. These UVA photons can cause oxidative damage to the guanine bases of DNA, which ultimately may result in mutation and melanoma promotion.[46–47]

Patterns of UVR exposure are important components of skin cancer. Although UVR is a major etiologic agent known to be associated with skin cancer, it is surprising

that most people still do not know that different patterns of sun exposure have different effects in the development of the different types of skin cancer. For example, chronic sun exposure that one receives during outdoor work on a daily basis does not increase risk for melanoma and is even associated with inhibition of melanoma.[19, 48] Chronic exposure, however, is directly related to SCC. The pattern of UV exposure that appears to be responsible for BCC is intermediate between that responsible for SCC and melanoma. Intermittent sun exposure, large blasts of UVR often received on weekends or holidays, seems to be the major form of UVR that is causal for the development of melanoma. Furthermore, excessive sunlight exposure in youth, particularly among individuals with multiple nevi (more than 100 over the entire body) and dysplastic nevi, seems to increase the risk of melanoma, particularly for melanoma associated with BRAF mutations, while continuous, intermittent sun exposure over a lifetime seems to be associated with NRAS mutations among older individuals.[49]

Increasing evidence shows a negative (harmful) effect of tanning beds.[50] Lights used in tanning beds and sun lamps give off mainly UVA radiation. The International Agency for Research on Cancer (IARC) convened an expert panel of epidemiologists in 2006 to evaluate the risks for melanoma and other skin cancers from the use of tanning beds.[51] A meta-analysis of 19 studies was undertaken to evaluate the association between tanning bed exposure and an increased risk of skin cancer. Findings showed that early life exposure to tanning beds is the most damaging. The overall relative risk for melanoma was 1.75 (95 percent confidence interval [CI], 1.35–2.26) for "first exposure under the age of 35." The relative risk for melanoma for "ever use" was less robust but still statistically significant (relative risk of 1.15, 95 percent CI, 1.00–1.31 for "ever use"). The analysis calculated a summary relative risk of 1.49 (95 percent CI, 0.93–2.38) for exposure distant in time, and a summary relative risk of 1.10 (95 percent CI, 0.76–1.60) for recent exposure. Based on the evidence, in 2009, another panel of IARC experts reclassified tanning bed UV from a "probable carcinogen" to a Class I carcinogen.[52] Clearly, exposure to tanning beds, especially at young ages, contributes to the higher risk of developing skin cancer.

Measurement of intermittent sun exposure or recreational sun exposure represents an important research challenge because measures of past sun exposure necessarily depend on subject recall of exposure, which is not always a reliable measure.[53] Epidemiologic studies invariably have shown weak associations between episodes of sunburn and melanoma incidence. Estimates of the effect of intermittent exposure have ranged from a protective effect reported by MacKie of 0.44 (CI 0.21–0.91),[54] to an adverse risk of 8.41 (CI 3.63–19.6) estimated by Grob.[55] It is likely that the imprecise measurement associated with self-reports is an important rate-limiting factor in determining the relationship between intermittent sun exposure and melanoma incidence. Also it is important to differentiate risk by skin type, as those with lighter skin color are at greatest risk of developing skin cancer. Other methods of assessment of recent sun exposure have been developed, which are designed to increase reliability and enhance participant recall of past sun exposure.[56]

Genetic Risk Factors

Genetic risk factors for melanoma focus on phenotypic characteristics. Phenotype is independently important as risk factors for the development of all skin

cancers: light hair color, light eye color, and light skin color, including skin that freckles easily. These phenotypic characteristics often are subsumed under the heading of "skin type." Fitzpatrick [57] classified skin type into six groups:

- Skin type 1: skin that always burns and never tans
- Skin type 2: skin that usually burns and tans with difficulty
- Skin type 3: tans gradually, sometimes has mild burns
- Skin type 4: rarely burns, tans with ease
- Skin type 5: Very rarely burns and tans very easily
- Skin type 6: Never burns and tans easily

Family history of melanoma appears to place an individual at higher risk for melanoma cancer. Familial melanoma accounts for an estimated 5 to 10 percent of all cases of melanoma.[58] First-degree relatives of melanoma patients have a higher risk of the disease than individuals without positive family history, suggesting that a distinct hereditary component is possible. Characteristics that distinguish the familial from the nonfamilial form of the disease include younger age at first diagnosis, better survival, thinner lesions, multiple primary lesions, and increased occurrence of NMSCs.[59] To put this in perspective, a study pooling data from eight case-control studies found that an individual's risk of melanoma increases twofold if the individual has an affected first-degree relative.[60] This effect was independent of host factors, such as age, nevus count, hair and eye color, and freckling. Little research, however, has been carried out in relationship to family history and NMSC. Nevertheless, it is likely that any association depends on an individual's skin type.

Major advances in genetics have made it possible to identify genetic factors that are critical to susceptibility to melanoma. Genome Wide Association Studies (GWAS) have identified new variants in genes that may play an important role in the development of melanoma.

Pigmentation Genes

Foremost among the pigmentary genes known to be associated with melanoma is the melanocortin 1 receptor (MC1R). Data show that some MC1R variants are associated both with melanoma and phenotype, particularly red hair, while others are associated only with melanoma development, suggesting that MC1R variants could play a role in melanoma development both via pigmentary and nonpigmentary pathways. Pigmentary and other genes recently have been identified by GWAS and their effects remain to be evaluated.[61, 62]

DNA Repair Genes

Genetic variants in DNA repair genes are obvious candidates for melanoma susceptibility based on the Xeroderma pigmentosum (XP) paradigm. XP patients have an approximately 1,000-fold increased risk for developing melanoma and variants of XP and other repair genes recently have been associated with risk.[30, 31]

Cell Cycle Genes

To date, few mutations have been found in cell cycle genes in melanoma etiology. The familial melanoma gene, CDKN2A, is rarely mutated among those with

sporadic melanoma, although it accounts for approximately 30 percent of mutations in those with a hereditary form of the disease.[63]

GENE-ENVIRONMENT INTERACTION

The pattern of sun exposure that appears to induce skin cancer, in particular melanoma, is complex and clearly is different by skin type (that is, propensity to burn, ability to tan). Armstrong and Kricker [18] have proposed a model consistent with data from other epidemiologic studies in which the risk for melanoma increases with increasing sun exposure among those who tan easily, but only by a small amount, after which risk decreases with increasing exposure. In other words, a nonlinear relationship exists between sun exposure and risk for melanoma. Among subjects who are intermediate in their ability to tan, risk continues to increase slowly and then at some point declines with increasing exposure. On the other hand, those subjects who have great difficulty tanning (those with light to fair skin) have an almost linear increase in risk with increasing sun exposure. This model recognizes that individuals are differentially susceptible to sun exposure and have different levels of risk based on skin type. Moreover, it suggests that different *types* or *patterns* of sun exposure are associated with different levels of risk for melanoma.

CHANGING BEHAVIOR BY EDUCATION AND PREVENTION

As the incidence and mortality of melanoma increase, there is an urgent need for effective educational programs. Skin cancer prevention is fairly straightforward because it is well documented that sun exposure is the major cause of skin cancer. Additionally, skin cancer is easy to detect as it is on the body's surface. Efforts to protect oneself from UV radiation exposure—staying in the shade, staying out of the sun during the peak hours of UV radiation, and wearing protective clothing—is simple enough to do. Evidence is weak, however, that "just staying out of the sun" will prevent skin cancer. Public health messages tell people to stay out of the sun, to wear protective clothing, and to use sunscreen are important but often unheeded messages. Unfortunately, complex behavioral change is difficult to accomplish. This is a critical juncture for the development of education programs. There is somewhat confusing recommendations promulgated by different factions—some saying to avoid the sun and others saying to enjoy a moderate amount of sun exposure. The tanning industry is even promoting the use of tanning parlors to develop a "safe" tan.

The U.S. Preventive Services Task Force reported that education and policy approaches to increasing sun-protective behaviors were effective when implemented in primary schools and in recreational and tourism settings. The task force found insufficient evidence to determine effectiveness when implemented in other settings, such as childcare centers, secondary schools and colleges, and occupational settings, and insufficient evidence to determine the effectiveness of interventions targeting health care settings and providers. Media campaigns alone, interventions oriented to parents or caregivers of children, and communitywide multicomponent interventions have not been shown to be especially effective.[64] Designing an educational or prevention message is not as easy as it appears.

The point of this discussion is not to suggest that individuals should increase sun exposure per se, but that they should take care in the sun and at the same time realize that a "small amount" (that is, 15 minutes at noon) of sun exposure may be beneficial. A number of groups—for example, the Australian and New Zealand Bone and Mineral Society, Osteoporosis Australia, Australasian College of Dermatologists, and the Cancer Council Australia—recently modified a long-standing message to the public from one of staying out of the sun to one of short periods of exposure for health benefits.[65] A study conducted in Queensland, Australia, however, found that even though most individuals tried to protect themselves when in the sun, those who did not stated that they did not think that they were out long enough to get sunburned.[66] Basically, the most consistent and robust protection from sunburn and UVR exposure is to "cover up." Wearing a hat, long sleeves, and long pants during the midday when the sun it at its hottest is advocated as there is strong evidence that shows a protective effect of clothing.[67, 68] Yet, it also must be acknowledged that some sun does penetrate cotton clothing.

DO SUNSCREENS WORK?

Numerous programs have been developed to promulgate awareness of the UV Index, which indicates the intensity of UV, and the SPF (sun protection factor), a term that refers to the increase in the amount of sun one can get without getting burned. SPF only applies to UVB, not UVA. Sunscreen companies have attempted to develop sunscreens that protect against UVA and UVB (broad-spectrum sunscreens). The SPF number listed on the product is an important indicator of sunscreen protection. A sunscreen with an SPF of 15, for example, filters 92 percent of the UVB. Put another way, a sunscreen with an SPF of 15 will delay the onset of sunburn in a person who otherwise would burn in 10 minutes to burn in 150 minutes. The SPF 15 sunscreen allows a person to stay out in the sun 15 times longer.

Armed with this information, individuals who use sunscreens may be under a false sense of security. For some, reliance on sunscreens can extend the duration of intentional sun exposure, such as sunbathing, which may increase the risk for melanoma. Often sunscreens are used as "tanning aides to avoid sunburn." Although sunscreens have been shown to reduce the risk of sunburn and probably prevent SCC of the skin, there is no evidence to support the belief that sunscreens will prevent the development of melanoma. Suggestions have been made that sunscreens actually increase risk for skin cancer, although the most likely way that this association might occur is when individuals use sunscreens to prolong their stay in the sun.[69]

A randomized trial was conducted to evaluate the effects of sunscreens on the putative precursor lesion for melanoma, the development of nevi. Gallagher et al. [70] found a small effect of sunscreens in reducing nevus formation—only among those who freckled—and the effect was extremely small although statistically significant. Conversely, Whiteman et al.,[71] in an observational study in Queensland, found that sun protection, including frequent sunscreen reapplication, was associated with a reduced number of nevi in children ages 1 to 3 years. A German study, however, found no evidence that education and sunscreen use affected the development of nevi in 1,232 children ages 2 to 7 years old.[72] Thus, the debate about sunscreen effectiveness continues. Perhaps the message one can take away from the

conflicting study results is that it is still not clear that sunscreens of any sort provide protection from developing melanoma, although use clearly prevents sunburn—a sign that sensitive skin has had too much UV.

SUN-PROTECTION STRATEGIES

Sun protection is clearly an important public health strategy. The American Cancer Society recommends sun avoidance between the hours of 10 A.M. and 4 P.M., when the sun is at its strongest. Total avoidance of artificial UV sources such as tanning beds is also recommended. Unfortunately, less than half (47 percent) of the U.S. population engages in any sun protection.[73] Coups et al. [74] examined data from the 2005 National Health Interview Survey and found that less than half (43 to 51 percent across age-groups) reported frequent (sometimes/most of the time/always) use of sunscreen; 65 to 80 percent did not usually stay in the shade when outside on a sunny day; and 15 to 51 percent used sun-protection clothing. More risky behavior was found among younger individuals. General population approaches to improve overall sun protection have been developed in Australia. For example, the Australian state of Victoria adopted the SunSmart Program integrating both environmental change and mass media public education; this work documents that a mix of strategies is highly effective but expensive in improving sun protection behaviors.[75]

Given the prevailing belief that a suntan is viewed by society as being "attractive" and "healthy," concerns for appearance often present a challenge for sun protection campaigns. Accordingly, an important approach to promoting sun protection involves the use of appearance appeals, which are designed to emphasize the harm to physical appearance associated with sun exposure, or to increase the perceived attractiveness of untanned skin. Such strategies, such as using photoaging to accentuate the appearance of sun damage, have been used to target teenagers and young adults.[76]

SCREENING FOR MELANOMA: WOULD THIS MAKE A DIFFERENCE?

Screening is a potentially useful tool to reduce morbidity and mortality from skin cancer. Although there are no randomized trials to support widespread population-based screening for melanoma, the idea of skin screening remains appealing for melanoma given that if detected early, the likelihood of survival is quite high. Fair-skinned individuals, persons with a family history of melanoma or with multiple or atypical moles, as well as those with a previous diagnosis of melanoma are all excellent candidates for periodic skin examination. But, problems persist. Evidence is lacking that screening for skin cancer prevents mortality.[77] In fact, there is controversy as to whether the increased incidence of melanoma is an artifact of more intensive screening.[78] Until the issue of screening efficacy is clarified, prudence dictates that individuals should conduct a skin self-examination (or have a partner help examine skin that cannot be viewed by oneself) and physicians should be encouraged to conduct a full body skin exam periodically, especially on those patients who are at high risk of developing skin cancer. For individuals at high risk, such as those with multiple nevi or atypical nevi, fair skin, or family history, physicians may use a variety of types of photography to evaluate change in the skin or particular nevus over time.

VITAMIN D SYNTHESIS AND SKIN CANCER

Unfortunately, the public has received confusing messages about sun exposure and vitamin D. Vitamin D is fundamental to bone development and a healthy immune system. Humans obtain vitamin D from UVB exposure, diet, and supplements. UVB rays convert 7-dehydrocholesterol (7-DHC) in the skin to previtamin D3, which is then converted to vitamin D3. Given the worry about sun exposure and skin cancer, the question remains: How does one achieve favorable vitamin D levels yet also practice skin cancer prevention?

A significant number of American (adults and children) have low blood levels of vitamin D.[79, 80] Whether the cause is a vitamin D–poor diet or the anti–skin cancer campaign that has encouraged people to keep out of the sun, the fact remains that a large number of Americans are vitamin D deficient. Hence, the question as to whether appropriate vitamin D levels can be maintained without risking sun damage is pertinent. Some data suggest that moderate sun exposure is likely to be beneficial for melanoma progression.[81] Other data suggest that the population attains adequate vitamin D through outdoor recreational activities and that increasing sun exposure would lead to an increase in skin cancer.[82] Despite the lack of evidence, the tanning industry has capitalized on the perceived health benefits of enhanced vitamin D levels to promote its product. The indoor tanning industry is a $5 billion per year business in America with customers composed primarily of Caucasian women, mostly female teenagers, who are at greatest lifetime risk for developing skin cancers and perhaps at risk for vitamin D deficiency. As the dilemma is played out, the benefits of exposure to UVB radiation cannot be separated from the harmful effects. Future studies will determine optimal vitamin D intake and whether sun-produced vitamin D confers the same health benefits as enhanced doses of oral supplements. For the moment, public health messages should encourage both of these approaches.

CONCLUSION

The increasing incidence of skin cancers and mortality from melanoma worldwide is a growing public health concern. Because sun exposure is the easiest and most modifiable risk factor for skin cancer, strategies to increase knowledge about skin cancer need to be widespread. Messages to encourage responsible sun protection and avoidance of excess sun exposure need to be disseminated more effectively. From a basic science perspective, more research across disciplines is needed to understand the interplay of genetic, environmental, and behavioral factors in the etiology of skin cancer. Although sun exposure is the major risk factor for skin cancer, it also is necessary for synthesis of vitamin D, necessary for bone and muscle health, and a possible protective factor for many diseases, including colon cancer. At this time, public health educators need to design more effective education programs to alert the population to the potentially deadly dangers of excessive sun exposure and to design programs to encourage safe sun behavior. In particular, given the mounting evidence that tanning beds increase the risk of melanoma and other skin cancers, such information needs to be disseminated more effectively and more widely than has been the case in the past. In the meantime, all individuals, regardless of skin color, should be advised to be cautious in the sun: love the sun but respect it and practice "safe sun."

REFERENCES

1. Jemal A, et al. Cancer statistics: 2006. *CA Cancer J Clin*. 2006;56:106–130.
2. Armstrong BK, Kricker A. Skin cancer. *Dermatoepidemiology*. 1995;13:583.
3. Bevona C, Sober AJ. Melanoma incidence trends. *Dermatol Clin*. 2002;20: 589–595.
4. Lends MB, Dawes M. Global perspectives of contemporary epidemiological trends of cutaneous malignant melanoma. *Br J Dermatol*. 2004;150:179–185.
5. International Agency for Research on Cancer. *GLOBOCAN 2000: Cancer Incidence, Mortality and Prevalence Worldwide, Version 1.0. IARC Cancer Base No. 5*. Lyon, France: International Agency for Research on Cancer Press; 2001.
6. Parkin DM, et al. Global cancer statistics: 2002. *CA Cancer J Clin*. 2005;55:74–108.
7. Linos E, et al. Increasing burden of melanoma in the United States. *J Invest Dermatol*. 2009;129(7):1666–1674.
8. Collins GL, Nickoonahand N, Morgan MB. Changing demographics and pathology of nonmelanoma skin cancer in the last 30 years. *Semin Cutan Med Surg*. 2004; 23(1):80–83.
9. Christenson LJ, et al. Incidence of basal cell and squamous cell carcinomas in a population younger than 40 years. *JAMA*. 2005;294:681–690.
10. Scotto J, Fears TR, Fraumeni JR. *Incidence of Nonmelanoma Skin Cancer in the United States*. Washington, DC: United States Department of Health and Human Services; 1981. Publication No. NIH 82–2433.
11. Athas WF, Hunt WC, Key CR. Changes in nonmelanoma skin cancer incidence between 1977–1978 and 1998–1999 in Northcentral New Mexico. *Cancer Epid Biomarker Prev*. 2003;2:1105–1108.
12. Demers A, et al. Trends of nonmelanoma skin cancer from 1960 through 2000 in Canadian population. *J Am Acad Dermatol*. 2005;53:320–328.
13. Stern RS. The mysteries of geographic variability in nonmelanoma skin cancer incidence. *Arch Dermatol*. 1999;135:843–844.
14. Staples MP, et al. Non-melanoma skin cancer in Australia: the 2002 national survey and trends since 1985. *Med J Aust*. 2006;184:6–10.
15. Chen JG, et al. Cost of nonmelanoma skin cancer treatment in the United States. *Dermatol Surg*. 2001;27:1035–1038.
16. Nugent Z, et al. Risk of secondary primary cancer and death following a diagnosis of nonmelanoma skin cancer. *Cancer Epidemiol Biomarkers Prev*. 2005;14: 2584–2590.
17. Kahn HS, et al. Increased cancer mortality following a history of nonmelanoma skin cancer. *J Natl Cancer Inst*. 2008;100(17):1215–1222.
18. Armstrong BK, Kricker A, English DR. Sun exposure and skin cancer. *Australas J Dermatol*. 1997;38(Suppl 1):S1–S6.
19. Gandini S, et al. Meta-analysis of risk factors for cutaneous melanoma: II. Sun exposure. *Eur J Cancer*. 2005;41(1):45–60.
20. Elwood JM, Jopson J. Melanoma and sun exposure: an overview of published studies. *Int J Cancer*. 1997;73:198–203.
21. Nelemans PJ, et al. An addition to the controversy on sunlight exposure and melanoma risk: a meta-analytical approach. *J Clin Epidemiol*. 1995;48(11):1331–1334.
22. Armstrong BK. Epidemiology of malignant melanoma: intermittent or total accumulated exposure to the sun? *J Dermatol Surg Oncol.* 1988;14(8):835–849.
23. Whiteman DC, et al. Melanocytic nevi, solar keratoses, and divergent pathways to cutaneous melanoma. *J Natl Cancer Inst*. 2003;95(11):806–812.

24. Boscoe FP, Schymura MJ. Solar ultraviolet-B exposure and cancer incidence and mortality in the United States: 1993–2002. *BMC Cancer*. 2006;6:264.
25. Veierød M, et al. A prospective study of pigmentation, sun exposure, and risk of cutaneous malignant melanoma in women. *J Natl Cancer Inst*. 2003;95(20):1530–1538.
26. Skender-Kalnenas TM, English DR, Heenan PJ. Benign melanocytic lesions: risk markers or precursors of cutaneous melanoma? *J Am Acad Dermatol*. 1995;33(6):1000–1007.
27. Olsen CM, et al. Nevus density and melanoma risk in women: a pooled analysis to test the divergent pathway hypothesis. *Int J Cancer*. 2009;124(4):937–944.
28. Chang YM, et al. A pooled analysis of melanocytic nevus phenotype and the risk of cutaneous melanoma at different latitudes. *Int J Cancer*. 2008;124(2):420–428.
29. Raimondi S, et al. MC1R variants, melanoma and red hair color phenotype: a meta-analysis. *Int J Cancer*. 2008;122(12):2753–2760.
30. Millikan RC, et al. Polymorphisms in nucleotide excision repair genes and risk of multiple primary melanoma: the Genes Environment and Melanoma study. *Carcinogenesis*. 2006;27(3):610–618.
31. Kerat K, et al. The Gln/Gln genotype of XPD codon 751 as a genetic marker for melanoma risk and Lys/Gln as an important predictor for melanoma progression: a case-control study in the Swedish population. *Oncol Report*. 2008;20(1):179–183.
32. Qin J, et al. Quantifying the change in melanoma incidence by Breslow thickness. *Biometrics*. 2002;58(3):665–670.
33. Massi D, et al. Naevus-associated melanomas: cause or chance? *Melanoma Res*. 1999;9(1):85–91.
34. Gandini S, et al. Meta-analysis of risk factors for cutaneous melanoma:1. Common and atypical naevi. *Eur J Cancer*. 2005;41(1):28–44.
35. Gilchrest BA, et al. The pathogenesis of melanoma induced by ultraviolet radiation. *N Engl J Med*. 1999;340(17)1341–1348.
36. Wang SQ, et al. Ultraviolet A and melanoma: a review. *J Am Acad Dermatol*. 2001;44:837–846.
37. Kaidbey K, et al. Photoprotection by melanin–a comparison of black and Caucasian skin. *J Am Acad Dermatol*. 1979;1:249–260.
38. Ichihashi M, et al. UV-induced skin damage. *Toxicology*. 2003;189:21–39.
39. Wikonkal NM, Brash DE. Ultraviolet radiation induced signature mutations in photocarcinogenesis. *J Investig Dermatol Symp Proc*. 1999;4:6–10.
40. Clingen PH, et al. Induction of cyclobutane pyrimidine dimers, pyrimidine (6–4) pyrimidone photoproducts, and Dewar valence isomers by natural sunlight in normal mononuclear cells. *Cancer Res*. 1995;55:2245–2248.
41. Young AR, et al. Human melanocytes and keratinocytes exposed to UVA or UVB in vivo show comparable levels of thymine dimers. *J Invest Dermatol*. 1998;111:936–940.
42. De Fabo EC, et al. Ultraviolet B but not ultraviolet A radiation initiates melanoma. *Cancer Res*. 2004;64:6372–6376.
43. Eller M, et al. Enhancement of DNA repair in human skin cells by thymidine dinucleotides: evidence for a p53-mediated mammalian SOS response. *Proc Natl Acad Sci U S A*. 1997;94:12627–12632.
44. Hadshiew I, et al. Photoprotective mechanisms of human skin: modulation by oligonucleotides. *Hautarzt*. 2002;52:167–173.
45. Agar N, Young AR. Melanogenesis: a photoprotective response to DNA damage? *Mutation Research*. 2005;571:121–132.

46. Lim HW, et al. American Academy of Dermatology Consensus Conference on UVA protection of sunscreens: summary and recommendations. *J Am Acad Dermatol*. 2001;44:505–508.
47. Wood SR, et al. UV causation of melanoma in Xiphophorus is dominated by melanin photosensitized oxidant production. *Proc Natl Acad Sci*. 2006;103:4111–4115.
48. Gallagher RP, Elwood JM, Yang CP. Is chronic sunlight exposure important in accounting for increases in melanoma incidence? *Int J Cancer*. 1989;44(5):813–815.
49. Thomas NE, Edmiston SN, Alexander A. Number of nevi and early-life ambient UV exposure are associated with BRAF-mutant melanoma. *Cancer Epidemiol Biomarkers Prev*. 2007;16(5):991–997.
50. Berwick M. Are tanning beds "safe"? human studies of melanoma. *Pigment Cell Melanoma Res*. 2008;21(5):517–519.
51. The International Agency for Research on Cancer Working Group on artificial ultraviolet light and skin cancer. *Int J Cancer*. 2006;12:1116–1122.
52. El Ghissassi F, et al. for International Agency for Research on Cancer. Special report: policy a review of human carcinogens, part d: radiation. *Lancet Oncol*. 2009;19:751–752.
53. Berwick M, Chen YT. Reliability of reported sunburn history in a case-control study of cutaneous malignant melanoma. *Am J Epidemiol*. 1995;141(11):1033–1037.
54. MacKie RM, Aitchison T. Severe sunburn and subsequent risk of primary cutaneous malignant melanoma in Scotland. *Br J Cancer*. 1982;46(6):955–960.
55. Grob JJ, et al. Study of sunbathing habits in children and adolescents: application to the prevention of melanoma. *Dermatology*. 1993;186(2):94–98.
56. Glanz K, Mayer JA. Reducing ultraviolet radiation exposure to prevent skin cancer methodology and measurement. *Am J Prev Med*. 2005;29(2):131–142.
57. Fitzpatrick TB. The validity and practicality of sun-reactive skin types 1 through 6. *Arch Dermatol*. 1988;124:869–871.
58. Greene MH, Fraumeni JJ. The hereditary variant of malignant melanoma. In: Clark WH, Goldman LI, Mastrangelo MJ, eds. *Human Malignant Melanoma*. New York: Grune and Stratton; 1979.
59. Kopf AW, et al. Familial malignant melanoma. *JAMA*. 1986;256:1915–1919.
60. Ford D, et al. Risk of cutaneous melanoma associated with a family history of the disease. *Int J Cancer*. 1995;62:377–381.
61. Bishop DT, et al. Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet*. 2009;41(8):920–925.
62. Falchi M, Bataille V, Hayward NK, et al. Genome-wide association study identifies variants at 9p21 and 22q13 associated with the development of cutaneous nevi. *Nat Genet*. 2009;41(8):915–919.
63. Hansson J. Familial melanoma. *Surg Clin North Am*. 2008;88(4):897–916.
64. Saraiya M, Glanz K, Briss P. Preventing Skin cancer: findings of the Task Force on Community Preventive Services on Reducing Exposure to Ultraviolet Light. *MMWR*. 2003;52:1–12.
65. Working Group of the Australian, New Zealand Bone and Mineral Society, Endocrine Society of Australia, Osteoporosis Australia. Vitamin D and adult bone health in Australia and New Zealand: a position statement. *Med J Aus*. 2005;182(6):281–285.
66. Stanton WR, Moffatt J, Clavarino A. Community perceptions of adequate levels and reasons for skin protection. *Behav Med*. 2005;31(1):5–15.
67. English DR, Milne E, Simpson JA. Sun protection and the development of melanocytic nevi in children. *Cancer Epidemiol Biomarkers Prev*. 2005;14(12):2873–2876.

68. Severi G, et al. European Organization for Research Treatment of Cancer (EORTC) Melanoma Cooperative Group: sun exposure and sun protection in young European children, an EORTC multicentric study. *Eur J Cancer.* 2002;38(6): 820–826.
69. Autier P, Boniol M, Dore JF. Sunscreen use and increased duration of intentional sun exposure: Still a burning issue. *Int J Cancer*. 2007;121(1):2755–2759.
70. Gallagher RP, et al. Broad-spectrum sunscreen use and the development of new nevi in white children: A randomized controlled trial. *JAMA*. 2000;283(22):2955–2960.
71. Whiteman DC, Brown RM, Purdie DM, Hughes MC. Melanocytic nevi in very young children: the role of phenotype, sun exposure, and sun protection. *J Am Acad Dermatol*. 2005;52(1):40–47.
72. Bauer J, et al. Interventional study in 1,232 young German children to prevent the development of melanocytic nevi failed to change sun exposure and sun protective behavior. *Int J Cancer*. 2005;116(5):755–761.
73. Hall HI, et al. Sun protection behaviors of the US white population. *Prev Med*. 1997;26(4):401–407.
74. Coups EJ, Manne SL, Heckman CJ. Multiple skin cancer risk behaviors in the US population. *Am J Prev Med*. 2008;34(2):87–93.
75. Montague M, Borland R, Sinclair C. Slip! slop! slap! and SunSmart, 1980–2000: skin cancer control and 20 years of population-based campaigning. *Health Educ Behav*. 2001;28(3): 290–305
76. Broadstock M, Borland R, Gason R. Effects of suntan on judgements of healthiness and attractiveness by adolescents. *J Appl Soc Psychol*. 1992;22:157–172.
77. US Preventive Services Task Force. Recommendations and rationale for screening for skin cancer. *Am J Prev Med*. 2001;203(Suppl):44–46.
78. Levell NJ, et al. Melanoma epidemic: a midsummer night's dream? *Br J Dermatol*. 2009;161:630–634.
79. Ginde AA, Liu MC, Camargo CA Jr. Demographic differences and trends of vitamin D insufficiency in the US population: 1988–2004. *Arch Intern Med*. 2009;169(6):626–632.
80. Kumar J, Muntner P, Kaskel FJ, Hailpern SM, Melamed ML. Prevalence and associations of 25-hydroxyvitamin D deficiency in US children: NHANES 2001–2004. *Pediatrics.* 2009; Epub ahead of print.
81. Berwick M, et al. Sun exposure and mortality from melanoma. *J Natl Cancer Inst*. 2005;97(3):195–199.
82. Diffey B. Do white British children and adolescents get enough sunlight? *BMJ*. 2005;331:3–4.

This page intentionally left blank

CHAPTER 12

Cervical Cancer Screening in the Developing World

Rengaswamy Sankaranarayanan, MD, Somanathan Thara, MD, Twalib A. Ngoma, MD, Paulo Naud, MD, PhD, and Namory Keita, MD

Cervical cancer is caused by persistent infection with one or more of the high-risk oncogenic types of human papillomaviruses (HPV).[1] The peak risk of acquiring HPV infection occurs soon after the onset of sexual activity. HPV infection resolves spontaneously in about 80 to 85 percent of infected women, but it may persist, leading to the occurrence and progression of precancerous lesions such as cervical intraepithelial neoplasia (CIN), particularly grade three CIN, and adenocarcinoma in situ in some women. If untreated, these precursor lesions may progress to invasive cervical cancer over a period of 5 to 15 years. Although early detection and treatment of asymptomatic precancerous lesions by screening leads to the prevention of invasive cervical cancer, preventing oncogenic HPV infection by vaccination provides a new option for the prevention of cervical cancer. Early clinical diagnosis constitutes a third approach to disease control in which the emphasis is on rapid detection of early stages of invasive cancer and offering effective treatment. Lack of effective screening programs and the high prevalence of oncogenic HPV infection (more than 10 percent in women age 30 years or more) are primarily responsible for the high burden of cervical cancer in many developing countries.[2–4] This chapter reviews the current status and future prospects for controlling cervical cancer in developing countries.

BURDEN OF CERVICAL CANCER

An estimated 493,000 new cervical cancer cases, 1.4 million prevalent cases, and 273,000 cervical cancer deaths occurred worldwide in 2002, four-fifths of which were reportedly experienced in developing countries.[5] The estimated new and prevalent cervical cancer cases and deaths in different world regions are given in table 12.1. It is quite likely that the burden of disease in developing countries is underestimated, given the grossly inadequate diagnostic and treatment services and cancer information systems. More than an eightfold difference between the highest and lowest incidence rates of cervical cancer has been observed worldwide.[6, 7] In many Sub-Saharan African, Central and South American, and South and Southeast

Table 12.1
Cancer of the Uterine Cervix, New Cases, Deaths, and Five-year Prevalence in 18 World Regions, 2002

	Cervix cancer burden		
Region	Cases	Deaths 5-year	Prevalence
World	492,800	273,200	1,409,200
More developed countries	83,400	39,500	309,900
Less developed countries	409,400	233,700	1,099,300
Eastern Africa	33,900	27,100	57,200
Middle Africa	8,200	6,600	13,900
Northern Africa	8,100	6,500	14,000
Southern Africa	7,600	4,400	13,100
Western Africa	20,900	16,700	35,700
Caribbean	6,300	3,100	18,400
Central America	17,100	8,100	49,300
South America	48,300	21,400	139,200
Northern America	14,600	5,700	58,200
Eastern Asia	61,100	31,300	191,900
Southeastern Asia	42,500	22,500	132,500
South Central Asia	157,700	86,700	446,100
Western Asia	4,400	2,100	13,700
Eastern Europe	30,800	17,100	107,700
Northern Europe	5,600	2,800	21,100
Southern Europe	10,600	4,100	40,900
Western Europe	12,700	5,600	49,200
Oceania	2,000	800	6,500

Source: Ferlay J, Bray F, Pisani P, Parkin DM. *GLOBOCAN 2002. Cancer Incidence, Mortality and Prevalence Worldwide. IARC Cancer Base No. 5 version 2.0*. Lyon, France: International Agency for Research on Cancer Press; 2004.

Asian countries, annual age-standardized incidence rates of cervical cancer exceed 25 cases per 100,000 women; the rates are lower than 7 per 100,000 women in Arab countries and lower than 10 per 100,000 women in most developed countries.[6, 7] Age-adjusted cervical cancer mortality rates exceed 10 per 100,000 women in most developing countries, with rates exceeding 25 per 100,000 in East Africa as opposed to less than 5 per 100,000 in developed countries.[5] The high mortality is due to advanced clinical stage at presentation and due to deficiencies in treatment availability, accessibility, and affordability in many developing countries.[3] Despite the existing know-how and experience to organize, expand, or reorganize screening programs to reduce the burden of cervical cancer, in most developing countries, such initiatives continue to be neglected.

CERVICAL SCREENING TESTS

Conventional cytology, the main workhorse in screening programs throughout the world, is the most widely used and time-tested cervical screening test. It has led to substantial reductions in cervical cancer burden in developed countries since the

Table 12.2
Test Characteristics of Cervical Screening Tests to Detect CIN 2 and 3 Lesions in Cross-sectional Studies in Developing Countries

Test	Sensitivity	Specificity
Cytology	31–78%	91–96%
HPV testing	61–100%	62–96%
VIA	37–95%	49–97%
VILI	44–92%	75–85%

Source: Sankaranarayanan R, Gaffikin L, Jacob M, Sellors J, Robles S. A critical assessment of screening methods for cervical neoplasia. *Int J Gynaecol Obstet.* 2005;89(Suppl 2):S4-S12.
Note: HPV = human papillomavirus; VIA = visual inspection with acetic acid; VILI = visual inspection with Lugol's iodine.

1950s. The strengths and challenges of cytology have become evident through the long experience with this remarkable screening test. In routine screening settings, the sensitivity of cytology varies widely. In reviews of several studies, the sensitivity of a single cytology smear to detect CIN 2 and 3 lesions ranged from 47 to 62 percent and the specificity ranged from 60 to 95 percent.[2, 8–10] The sensitivity varied from 31 to 78 percent and the specificity ranged from 91 to 96 percent for detecting CIN 2–3 lesions in cross-sectional studies in developing countries, when cytology positivity was defined at the ASCUS (atypical squamous cells of undetermined significance) cut-off level (see table 12.2).[11]

Liquid-based cytology (LBC) is a more expensive test than conventional cytology and requires additional instrumentation to prepare the smears; it is not feasible to implement LBC in many low-resource settings due to the costs and instrumentation required. The challenges and resources required to organize cytology screening in many parts of the developing world, and the apparent limited impact of such programs, especially in several Latin American countries, have prompted the search for and evaluation of alternative screening tests and paradigms.[12] Ensuring consistently high performance in detecting high-grade precursor lesions can be challenging under field conditions and doing so requires constant monitoring and frequent retraining of test providers.

Visual inspection with 3 to 5 percent dilute acetic acid (VIA), the most widely studied visual screening test, involves naked-eye inspection of the cervix under bright light one minute after the application of acetic acid using a cotton swab or a spray.[13] A positive test is characterized by well-defined acetowhite areas close to the squamocolumnar junction (SCJ), or to the external or the entire cervix, or a cervical growth turning acetowhite.[13] VIA can be taught easily and quickly to primary health workers, nurses, and doctors, and gives immediate results to permit a single-visit approach for diagnosis and treatment.[14] The sensitivity of VIA to detect CIN 2 and 3 lesions and invasive cervical cancer varied from 37 percent to 95 percent and the specificity varied from 49 percent to 97 percent in cross-sectional studies in developing countries (see table 12.2).[11] The wide-ranging accuracy data for VIA underscore the subjective nature of the test as well as the range in the validity, completeness, and accuracy of the reference standard used in different studies.

Visual inspection with Lugol's iodine (VILI) also involves naked-eye examination of the cervix, and results are apparent immediately after application of the iodine. A positive result is based on the appearance of definite mustard-yellow area on the cervix close to the SCJ or on a cervical growth.[13] The sensitivity of VILI varied between 44 and 92 percent and specificity ranged between 75 and 85 percent based on data from cross-sectional studies in India, Africa, and Latin America (see table 12.2).[11, 15–17]

Another relatively new cervical cancer screening test relies on HPV testing by hybrid capture II (HCII). This test is the most objective and reproducible of all cervical screening tests. The sensitivity of HPV testing by HCII test in detecting CIN 2 and 3 lesions varied from 66 to 100 percent and the specificity varied from 62 to 96 percent in cross-sectional studies in developing countries (see table 12.2).[2, 11, 18] Studies have shown that this test has a higher sensitivity but lower specificity than cytology in detecting high-grade lesions. In recently reported randomized trials from Sweden, the Netherlands, and Canada, for example, HPV testing has greater sensitivity for the detection of CIN when compared with Pap testing.[19–21] Although self-sampling for HPV DNA (deoxyribonucleic acid) testing seems to be potentially promising for use in underresourced areas as well as for women who might be reluctant to participate in screening programs, further definitive research is needed to validate its clinical utility.[22]

Compared with VIA or VILI testing, HPV testing is currently more expensive (US$20 to US$40) than other cervical screening tests and requires sophisticated laboratory infrastructure for processing the cervical cells, which are not available in most developing countries. Further developments in terms of more rapid, affordable, and simple tests are essential to make HPV testing feasible in low-resource settings. The *care* HPV test, a simple, user-friendly, affordable, fast (results within three hours), and accurate HPV test suitable for use in low-resource settings, has been evaluated in China and was found to have similar accuracy as that of HCII, significantly higher sensitivity than VIA (90.2 percent versus 41.4 percent), but lower specificity (84.2 percent versus 94.5 percent).[23] It is expected to be commercially available for developing countries in the next few years.

SCREENING PROGRAMS FOR CERVICAL NEOPLASIA

The aim of cervical screening is to prevent death from invasive cervical cancer by detecting and treating women with high-grade CIN 2 and 3 lesions. Screening tests such as conventional cytology, LBC, VIA, VILI, and HPV testing have a high accuracy of identifying women having CIN as well as early, asymptomatic, preclinical invasive cancer if conducted by well-trained providers. Effective treatment methods for women diagnosed with CIN include cryotherapy, cold coagulation, laser ablation, loop electrosurgical excision procedure (LEEP), laser conization, and cold knife conization. Colposcopy is a useful diagnostic tool to triage women with positive tests to assess the nature and extent of lesions and to direct biopsies and treatment.

The critical components of successful cervical screening programs are high coverage of the target population of women by means of accurate, quality-assured screening tests as well as the ability to refer screen-positive women for further diagnostic workup. Women with confirmed cervical neoplasia must be referred for

treatment and follow-up care. Organized cervical cancer screening programs are more likely to have a systematic call-recall-follow-up system in place as well as an information system useful for monitoring and evaluating the efficacy of the program. In contrast, unorganized programs generally are characterized by a more informal protocol whereby screening tests are provided when clients themselves request testing or the tests are prescribed coincidentally during routine health exam. Organized screening with a callback protocol have been shown to have the greatest effect and use fewer resources than unorganized programs.

Population-based screening programs are yet to be implemented in many developing countries where the prevalence of cervical cancer is quite high, such as in Africa, and some Southern and East Asian countries. Taiwan, Hong Kong, the Republic of Korea, Singapore, and Thailand, in contrast, have introduced a much wider scale of cytology screening than other economies in Asia, primarily because they have greater resources to do so. Colposcopy and treatment (LEEP, laser, cryotherapy) services for CIN are more extensively available in these wealthier countries than elsewhere in Asia. A 48 percent reduction in cervical cancer incidence during 1995–2006 was reported in Taiwan, following the improvements in coverage of screening due to reimbursement of Pap smear costs by the Taiwan National Health Insurance Scheme.[24] Nevertheless, large-scale VIA-based screening currently is being implemented in Thailand and Bangladesh.

Mass cytology screening programs do not exist anywhere in Africa, although South Africa has plans to introduce a cytology screening program, but it is yet to be implemented on a large scale. Although cytology screening is widespread in Mauritius, screen-positive women seldom are diagnosed and treated. Not surprisingly, there has been a net effect of no appreciable impact on disease burden over the last three decades. VIA-based screen-and-treat services currently are being implemented in limited regions of Angola, Burkina Faso, Cameroon, Congo Brazzaville, Ghana, Guinea, Kenya, Mali, Malawi, Madagascar, Tanzania, Uganda, and Zimbabwe.

Large-scale cytology screening has been introduced in many Caribbean and Latin American countries—for example, nationally in Cuba, Costa Rica, Chile, and Uruguay, and regionally in Brazil, Argentina, Mexico, Peru, Colombia, Ecuador, Panama, Venezuela, and Bolivia. Yet, despite the fact that an estimated 20 million cytology smears currently are taken annually in the entire Latin American region, there has been very little impact on disease burden for several years after their introduction due to a combination of factors, such as suboptimal cytology testing, lack of quality assurance, poor coverage of women at risk, and inadequate follow-up of screen-positive women with diagnosis and treatment.[2, 3, 25] For example, two decades of opportunistic annual screening for cervical cancer from the mid-1960s to the mid-1980s did not reduce cervical cancer mortality in Chile. Perhaps in response to this reality, in 1987, the screening program was reorganized to target women age 25–64 years every three years, thus optimizing existing resources, screening promotion strategies, quality of Pap smear, and timeliness of diagnosis and treatment. This led to more than 80 percent coverage of target women.[25] As a result, the age-adjusted cervical cancer mortality declined from 12.8 in 1980 to 6.8 per 100,000 women in 2001.[25] Efforts are ongoing to reorganize programs in Argentina, Brazil, Costa Rica, Mexico, Uruguay, and Panama.

NEW PARADIGMS OF SCREENING

Screening interventions available in the developed world, where women beginning in their 20s are screened at one-to-five-year intervals, are not feasible in developing countries. Multiple visits will result in substantial nonparticipation or loss of follow-up among women in low- and medium-resource countries. As such, in recent years, new paradigms have been introduced or proposed to maximize participation of women for screening and treatment in a cost-effective way. Programs such as single lifetime screening targeting women ages 30–59 or 30–49 years[12, 26, 27]; a single-visit approach, in which a positive screen is followed by colposcopy, directed biopsies, and treatment with cryotherapy or LEEP in the same sitting[14, 26, 27]; and a single-visit screen-and-treat approach, in which screen-positive women, without evidence of invasive cancer, are treated with cryotherapy in the same session, without diagnostic procedures such as colposcopy and biopsy having been proposed.[14, 28–30] As discussed earlier, low-tech visual screening tests such as VIA and VILI provide immediate results making the single-visit approach quite feasible and cost-effective.

The cost-effectiveness of different cervical cancer screening strategies in India, Kenya, Peru, South Africa, and Thailand, were assessed using computer-based models.[31] Outcomes included the lifetime risk of cancer, years of life saved, lifetime costs, and cost-effectiveness ratios (cost per year of life saved). The most cost-effective strategies were those with the fewest visits, resulting in improved follow-up testing and treatment. Screening women once in their lifetime, at the age of 35 years, with a one-visit or two-visit screening strategy involving VIA or HPV testing in cervical cell samples, reduced the lifetime risk of cancer by approximately 25 to 36 percent, and cost less than US$500 per year of life saved. Relative risk of cancer declined by an additional 40 percent with two screenings (at 35 and 40 years of age), resulting in a cost per year of life saved that was less than each country's per capita gross domestic product, a very cost-effective result, according to the Commission on Macroeconomics and Health. These results indicate that cervical cancer screening strategies incorporating VIA or HPV testing in one or two clinical visits are cost-effective alternatives to cytology screening programs in resource-poor settings. Cost-effectiveness data will further contribute to public health decisions on implementing alternative tests to cytology.

Although a large amount of accuracy and reproducibility data are available for visual and HPV tests, these alone do not make a sufficient case for their introduction in routine population-based screening programs. High accuracy of a given test to detect CIN does not automatically indicate a reduction in incidence and mortality as the tests may miss real precursors progressing to invasive cancer. A high accuracy is not a surrogate for efficacy for reducing disease burden. The following discusses the evidence for visual and HPV tests from randomized controlled trials in developing countries.

REDUCTION IN DISEASE BURDEN FOLLOWING VISUAL AND HPV SCREENING IN DEVELOPING COUNTRIES

The safety and efficacy of VIA or HPV testing, followed by cryotherapy, in reducing the prevalence of CIN 2 and 3 lesions as compared with a delayed evaluation group (the control group) were evaluated in a randomized controlled trial in

South Africa involving approximately 6,000 women.[29] After six months from treatment, CIN 2 and advanced lesions were diagnosed in 0.8 percent of the women in the HPV testing group and 2.2 percent in the VIA group compared with 3.6 percent in the delayed evaluation group ($p < 0.001$ and $p = 0.02$ for the HPV and VIA groups, respectively). The respective cumulative prevalence rates at 12 months were 1.2 percent in the HPV testing group, 2.9 percent in the VIA group, and 5.4 percent in the control group. It was concluded that both screen-and-treat approaches are safe and result in a lower prevalence of high-grade cervical cancer precursor lesions compared with delayed evaluation at both 6 and 12 months.

The efficacy and effectiveness of VIA screening in reducing cervical cancer incidence and mortality were addressed in randomized controlled trials in India.[26, 27, 32] In a randomized trial in Dindigul district, South India, 57 clusters with 49,311 eligible women ages 30 to 59 years were randomized to a single round of VIA by trained nurses, and 57 clusters (30,958) to a control group to receive routine care.[27] Of the 3,088 (9.9 percent) women who tested positive, 3,052 had a colposcopy and 2,539 had a biopsy. Of the 1,874 women with precancerous lesions in the intervention group, 72 percent received treatment. During 2000–2006, there were 167 cervical cancer cases and 83 cervical cancer deaths in the intervention group, compared with 158 cases and 92 deaths in the control group, indicating a 25 percent reduction in cancer incidence and a 35 percent reduction in cancer mortality among the targeted women in the intervention group (see table 12.3).[27] The greatest reduction in incidence and mortality rates were observed for the 30- to 39-year-old age-group.

The impact of screening by a single round of VIA, cytology, or HPV testing on cervical cancer incidence and mortality was investigated in a cluster randomized controlled trial in Osmanabad District, India.[32] Eligible women ($N = 131,746$) ages 30 to 59 years living in 497 villages in Osmanabad District were grouped into 52 clusters, which were randomized to receive screening with either VIA (13 clusters, 34,074 women), cytology (13 clusters, 32,058 women), or HPV testing (13 clusters,

Table 12.3
Effectiveness of VIA in Reducing Cervical Cancer Incidence and Mortality in a Cluster Randomized Controlled Trial, Dindigul District, India, 2000–2006

End point	VIA group	Control group	Adjusted hazard ratio[a] (95% CI)
Total number of women	49,311	30,958	—
Number screened	31,343	951	—
Number screen positive	3,088	—	—
Cervical cancer cases	167	158	—
Age-standardized incidence rate	75.2/100,000	99.1/100,000	0.75 (0.59–0.95)
Cervical cancer deaths	83	92	—
Age-standardized mortality rate	38.3/100,000	54.9/100,000	0.65 (0.47–0.89)

Source: Sankaranarayanan R, Esmy PO, Rajkumar R, et al. Effect of visual screening on cervical cancer incidence and mortality in Tamil Nadu, India: a cluster-randomised trial. *Lancet.* 2007;370:398-406.

Note: CI = confidence interval; VIA = visual inspection with acetic acid.

a. Adjusted for cluster design, age, education, marital status, and parity.

34,126 women). The remaining women constituted the control group (13 clusters, 31,488 women) for which no screening was offered, but these women received education about existing facilities in routine health services for early detection and prevention of cervical cancer.[32] More than 79 percent of women participated in screening. Screen-positive women had colposcopy or directed biopsies; those with cervical precancerous lesions received treatment with cryotherapy, loop excision, or conization; and those with invasive cancers received anticancer treatment.

Table 12.4 shows the comparative efficacy of three different cervical screening tests in reducing cervical cancer mortality in a randomized controlled trial in Osmanabad District, India. Overall, in view of the active detection of cases, the intervention groups had a higher cervical cancer incidence. The HPV group had a significant 53 percent reduction in the incidence rate of stage II or worse stages of invasive cervical cancer (hazard ratio 0.47 [95 percent confidence interval [CI], 0.32–0.69]) and had a significant 48 percent reduction in cervical cancer mortality (hazard ratio 0.52 [95 percent CI, 0.33–0.83]) compared with the control group.[32] The reduction in cervical cancer mortality observed in the cytology (hazard ratio 0.89 [95 percent CI, 0.62–1.27]) and in the VIA groups (hazard ratio: 0.86 percent [95 percent CI, 0.60–1.25]) as compared with the control group did not reach statistical significance. During the eight-year follow-up period, 8 of the 24,380 HPV-negative women developed cervical cancer as compared to 22 of 23,762 Pap smear–negative women and 25 of 23,032 VIA-negative women, thus indicating that HPV testing more accurately identified women at risk for developing cervical cancer.[32]

The lack of significant cervical cancer mortality reductions following VIA or Pap smear screening in this study have caused concern in light of some countries beginning to implement VIA-based screen-and-treat programs. Whereas HPV testing seems to be more sensitive to detect potentially progressive precursor lesions than cytology or VIA, long-term follow-up of the study cohorts may clarify the current lack of effect following VIA or Pap smear screening. An underestimation of the number of cervical cancer cases and deaths in the control group cannot be ruled out because cervical cancers were diagnosed in the control group in women with symptoms seeking diagnostic services and health services. Some women with disease might not have interacted with routine health services and such cases and deaths might have been missed due to the lack of hospital records for them. The above possibilities may have contributed to an apparent lack of the effectiveness of VIA and cytology screening in this study.

SAFETY AND EFFECTIVENESS OF FIELD-BASED TREATMENTS FOR CIN IN DEVELOPING COUNTRIES

Cryotherapy provided by nurses and loop excision by doctors in field conditions in India, Thailand, and South Africa has been found to be safe, acceptable, and effective.[28, 30, 33, 34] Of 1,879 women who were treated for cervical cancer, by adapting see-and-treat procedures involving VIA screening, colposcopy, directed biopsy, and double-freeze cryotherapy by trained nurses in a single-visit in field clinics supervised by a doctor, four-fifths of the women with CIN were cured. Cure rates were 81.4 percent (752/924) for women with CIN 1; 71.4 percent (55/77) for CIN 2; and 68 percent (17/25) for CIN 3 cases.[33] Minor side effects and complications

Table 12.4
Comparative Efficacy of Three Different Cervical Screening Tests in Reducing Cervical Cancer Mortality in a Randomized Controlled Trial in Osmanabad District, India

Variable	HPV Group (N = 34,126)	Pap smear group (N = 32,058)	VIA group (N = 34,074)	Control (routine care) group (N = 31,488)
Women screened	27,192	25,549	26,765	—
Test-positive women	2,812	1,787	3,733	—
Women with high-grade precancerous changes	245	262	195	—
Women diagnosed with cervical cancer during 2000–2007 (rate per 100,000 person-years)	127 (47.4)	152 (60.7)	157 (58.7)	118 (47.6)
Hazard ratio (95% CI)	*1.05 (0.77–1.43)*	*1.34 (0.99–1.82)*	*1.30 (0.95–1.78)*	*1.00*
Women with stage II or more advanced cancer (rate per 100,000 person-years)	39 (14.5)	58 (23.2)	86 (32.2)	82 (33.1)
Hazard ratio (95% CI)	*0.47 (0.32–0.69)*	*0.75 (0.51–1.10)*	*1.04 (0.72–1.49)*	*1.00*
Women dying from cervical cancer during 2000–2007 (rate per 100,000 person-years)	34 (12.7)	54 (21.5)	56 (20.9)	64 (25.8)
Hazard ratio (95% CI)	*0.52 (0.33–0.83)*	*0.89 (0.62–1.27)*	*0.86 (0.60–1.25)*	*1.00*

Source: Sankaranarayanan R, Nene BM, Shastri SS, et al. HPV screening for cervical cancer in rural India. *N Engl J Med*. 2009;360:1385–1394.
Note: Rates and hazard ratios for the comparison between each intervention group and control group have been adjusted for age. CI = confidence interval; HPV = human papillomavirus; VIA = visual inspection with acetic acid.

were documented in less than 3 percent of women. In a screening trial in Osmanabad District, 94 percent of 574 women with CIN treated with cryotherapy by nurses were cured; cure rates were 96.4 percent (451/468) for women with CIN 1; 82.1 percent for CIN 2 and 3 lesions (55/67 and 32/39, respectively); and minor side effects and complications were documented in 5.2 percent of women.[34]

Of the 488 women with CIN treated with LEEP in the screening trial in Osmanabad District, cure rates were 98.1 percent (253/258) for women with CIN 1; 93.6 percent (103/110) for CIN 2; and 85 percent (102/120) for CIN 3.[35] Minor side effects and complications were documented in 12.1 percent of women. In Kerala, India, of 283 women with CIN treated with LEEP, 248 (87.6 percent) had no evidence of CIN at follow-up; cure rates were 93.0 percent (145/156) for women with CIN 1; 85.5 percent (71/83) for CIN 2; and 72.7 percent (32/44) for CIN 3. Minor side effects were documented in 34 women (12 percent) and complications in 5 women (1.8 percent).[36] These experiences are similar to the results reported from developed countries where these procedures were carried out by experienced doctors.[37, 38]

CLINICAL EARLY DIAGNOSIS

Clinical early diagnosis aims to diagnose and treat invasive cancer in early stages among symptomatic women. Education and awareness among women, effective referral, and sufficient diagnostic and treatment facilities in the health care system are critical for the success of this approach. Programs focused on achieving earlier diagnosis of symptomatic disease offer great potential to reduce mortality with a minor investment in resources (manpower and financial) in settings where mass screening is not feasible. Detecting cervical cancers in earlier clinical stages and treating them effectively resulted in the reduction in mortality from this cancer in developed countries before cervical screening programs were introduced.[39, 40] The proportion of stage III and IV cervical cancers was reportedly reduced from 60 percent in 1993 to 26 percent in 1998 in Sarawak province of Malaysia after the introduction of a clinical early diagnosis program and improvements in health care infrastructure.[41] Considerable investments in cancer health care infrastructure and human resources must be made in several developing countries, particularly those in Sub-Saharan Africa, to diagnose cervical cancers early and treat them effectively.

PREVENTION BY VACCINATION

The potential is huge to reduce the cervical cancer burden by means of HPV vaccination. Globally, 70 percent of cervical cancers are caused by HPV 16 and 18 infections, which equates to 350,000 potentially preventable cervical cancers annually by vaccination against HPV 16 and 18. Bivalent (HPV 16, 18) and quadrivalent (HPV 6, 11, 16, 18) HPV L1 virus-like particle (VLP) vaccines have been evaluated and the results indicate, with remarkable consistency, that a regimen of three intramuscular injections of HPV vaccine offers HPV-naive adolescent girls a very high level of protection (99 percent) from infections and CIN associated with the HPV types included in the vaccine.[42–48] The vaccines were found to be safe and well tolerated with relatively few side effects.

A recommendation for HPV vaccination for adolescent women for a disease that occurs during adulthood is a major paradigm shift in cervical cancer control.

HPV vaccination holds great promise and has been licensed for use in more than 100 countries, but several challenges remain for its widespread implementation through national immunization programs in high-risk developing countries.[49] The current cost of these vaccines precludes sustained global delivery, and they target only 2 of the approximately 15 known oncogenic HPV types, although approximately 70 percent of cervical cancer cases are attributed to these two types and evidence points to some degree of cross-protection against other closely related types. The other concerns include logistics of vaccine delivery (in view of the need for three doses spread over six months), improved strategies and vaccine platforms to reach out to pre- or early adolescent girls, long-term immunogenicity and efficacy in preventing cervical neoplasia, cross-protection against HPV types not targeted by the vaccine antigens, and the efficacy of different, more logistically feasible dose regimes in inducing and maintaining immunogenicity and long-term protection against cervical neoplasia. It has been shown in a Canadian randomized trial that antibody responses to HPV 16, 18, 6, and 11 following two doses of quadrivalent HPV vaccine at months zero and six in 259 girls ages 9 to 13 years was not inferior when compared with that of three doses at zero, two, and six months in 261 girls.[50]

A possible approach to broader immunity at lower cost is L2 vaccines that can be produced inexpensively and have the promise of conferring much broader cross-type protective immunity than that observed with L1 VLP vaccines.[51] However, L2 vaccine development lags behind L1 VLP vaccines and several technical hurdles remain. While prophylactic vaccination will provide important future health protection, if vaccination is offered to girls before the onset of sexual activity, cervical screening should be continued for older women as there still is a risk of already being infected with the oncogenic HPV types or of acquiring CIN due to an HPV type other than 16 or 18. Although the HPV vaccination provides hope for the future, screening provides the means for the present.

CONCLUSION

Cervical cancer reflects striking global health inequity, resulting in the deaths of women in their most productive years in developing countries, with a devastating effect on the society at large. It remains as the largest single cause of years of life lost to cancer in the developing world. The large evidence and experience base for different screening approaches should drive the implementation of appropriate screening policies in different settings. It is time to focus attention on putting in place the important programmatic components related to education, participation, good quality testing, diagnosis, treatment, follow-up care, and evaluation. Be it for cervical cancer or for other diseases, the need is urgent to improve health services infrastructure (delivery and finance) for effective disease control. Delaying investments for cervical cancer screening in resource-poor countries in particular means that many women will continue to die unnecessarily from this highly preventable and treatable disease. Clinically effective, safe, and cost-effective options are available. Implementation of any one of the options would do much to afford women in the developing world the same opportunities that women in the developed world already receive. Being a female in a developing country should no longer be a reason for developing and even dying of cervical cancer.

REFERENCES

1. Munoz N, Castellsague X, de Gonzalez AB, Gissmann L. HPV in the etiology of human cancer. *Vaccine*. 2006;24(Suppl 3):S1–S10.
2. International Agency for Research on Cancer. *IARC Handbooks on Cancer Prevention*. Vol. 10: *Cervix Cancer Screening*. Lyon, France: International Agency for Research on Cancer Press; 2004.
3. Sankaranarayanan R, Budukh AM, Rajkumar R. Effective screening programmes for cervical cancer in low- and middle-income developing countries. *Bull World Health Organ*. 2001;79:954–962.
4. International Agency for Research on Cancer HPV Prevalence Surveys Study Group. Worldwide distribution of human papillomavirus types in cytologically normal women in the International Agency for Research on Cancer HPV prevalence surveys: a pooled analysis. *Lancet*. 2005;366:991–998.
5. Ferlay J, Bray F, Pisani P, Parkin DM. *GLOBOCAN 2002. Cancer Incidence, Mortality and Prevalence Worldwide. IARC Cancer Base No. 5 version 2.0*. Lyon, France: International Agency for Research on Cancer Press; 2004.
6. Parkin DM, Whelan SW, Ferlay J, Teppo L, Thomas DB. *Cancer Incidence in Five Continents*. Vol. 8. Lyon, France: International Agency for Research on Cancer Press; 2002. IARC Scientific Publications No.115.
7. Curado MP, Edwards B, Shin HR, et al. *Cancer Incidence in Five Continents*. Vol. 9. Lyon, France: International Agency for Research on Cancer Press; 2007. IARC Scientific Publications No.160.
8. Nanda K, McCrory DC, Myers ER, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med*. 2000;132:810–819.
9. Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. *Am J Epidemiol*. 1995;141:680–689.
10. Mitchell MF, Schottenfeld D, Tortolero-Luna G, Cantor SB, Richards-Kortum R. Colposcopy for the diagnosis of squamous intraepithelial lesions: a meta-analysis. *Obstet Gynecol*. 1998;91:626–631.
11. Sankaranarayanan R, Gaffikin L, Jacob M, Sellors J, Robles S. A critical assessment of screening methods for cervical neoplasia. *Int J Gynaecol Obstet*. 2005;89(Suppl 2):S4–S12.
12. Alliance for Cervical Cancer Prevention. *Planning and Implementing Cervical Cancer Prevention and Control Programs: A Manual for Managers*. Seattle, WA: Alliance for Cervical Cancer Prevention; 2004.
13. Sankaranarayanan R, Wesley R. *A Practical Manual on Visual Screening for Cervical Neoplasia*. Lyon, France: International Agency for Research on Cancer Press; 2003. IARC Technical Publication No.41.
14. World Health Organization. *Comprehensive Cervical Cancer Control: A Guide to Essential Practice*. Geneva, Switzerland: World Health Organization; 2006.
15. International Agency for Research on Cancer Multicentre Study Group on Cervical Cancer Early Detection. Accuracy of visual screening for cervical neoplasia: results from an IARC multicentre study in India and Africa. *Int J Cancer*. 2004;110:907–913.
16. Sarian LO, Derchain SF, Naud P, et al. Evaluation of visual inspection with acetic acid, Lugol's iodine, cervical cytology and HPV testing as cervical screening tools in Latin America. This report refers to partial results from the Latin American Screening study. *J Med Screen*. 2005;12:142–149.

17. Sangwa-Lugoma G, Mahmud S, Nasr SH, et al. Visual inspection as a cervical cancer screening method in a primary health care setting in Africa. *Int J Cancer*. 2006;119:1389–13895.
18. Arbyn M, Sasieni P, Meijer CJ, Clavel C, Koliopoulos G, Dillner J. Clinical applications of HPV testing: a summary of meta-analyses. *Vaccine*. 2006;24(Suppl 3): S78–S89.
19. Naucler P, Ryd W, Törnberg S, et al. Human papillomavirus and Papanicolaou tests to screen for cervical cancer. *N Engl J Med*. 2007;357:1589–15897.
20. Mayrand MH, Duarte-Franco E, Rodrigues I, et al. Canadian Cervical Cancer Screening Trial Study Group: human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. *N Engl J Med*. 2007;357:1579–1588.
21. Bulkmans NW, Berkhof J, Rozendaal L, et al. Human papillomavirus DNA testing for the detection of cervical intraepithelial neoplasia grade 3 and cancer: 5-year follow-up of a randomised controlled implementation trial. *Lancet*. 2007;370:1764–1772.
22. Stewart DE, Gagliardi A, Johnston M, et al. HPV Self-collection Guidelines Panel: self-collected samples for testing of oncogenic human papillomavirus: a systematic review. *J Obstet Gynecol Can*. 2007;29:817–828.
23. Pretorius RG, Zhang X, Belinson JL, et al. Distribution of cervical intraepithelial neoplasia 2, 3 and cancer on the uterine cervix. *J Low Genit Tract Dis*. 2006;10:45–50.
24. Taiwan Cervical Cancer Screening Task Force. Effectiveness of national cervical cancer screening programme in Taiwan: 12-year experiences. *Br J Cancer*. 2009;101:174–177.
25. Sepulveda C, Prado R. Effective cervical cytology screening programmes in middle-income countries: the Chilean experience. *Cancer Detect Prev*. 2005;29:405–411.
26. Sankaranarayanan R, Nene BM, Dinshaw KA, et al. Osmanabad District Cervical Screening Study Group: a cluster randomized controlled trial of visual, cytology and human papillomavirus screening for cancer of the cervix in rural India. *Int J Cancer*. 2005;116:617–623.
27. Sankaranarayanan R, Esmy PO, Rajkumar R, et al. Effect of visual screening on cervical cancer incidence and mortality in Tamil Nadu, India: a cluster-randomised trial. *Lancet*. 2007;370:398–406.
28. Blumenthal PD, Gaffikin L, Deganus S, Lewis R, Emerson M, Adadevoh S. Cervical cancer prevention: safety, acceptability, and feasibility of a single-visit approach in Accra, Ghana. *Am J Obstet Gynecol*. 2007;196:407–408.
29. Denny L, Kuhn L, De SM, Pollack AE, Dupree W, Wright TC Jr. Screen-and-treat approaches for cervical cancer prevention in low-resource settings: a randomized controlled trial. *JAMA*. 2005;294:2173–2181.
30. Gaffikin L, Blumenthal PD, Emerson M, Limpaphayom K. Safety, acceptability, and feasibility of a single-visit approach to cervical-cancer prevention in rural Thailand: a demonstration project. *Lancet*. 2003;361:814–820.
31. Goldie SJ, Gaffikin L, Goldhaber-Fiebert JD, et al. Alliance for Cervical Cancer Prevention Cost Working Group: cost-effectiveness of cervical-cancer screening in five developing countries. *N Engl J Med*. 2005;353(20):2158–2168.
32. Sankaranarayanan R, Nene BM, Shastri SS, et al. HPV screening for cervical cancer in rural India. *N Engl J Med*. 2009;360:1385–1394.
33. Sankaranarayanan R, Rajkumar R, Esmy PO, et al. Effectiveness, safety and acceptability of 'see and treat' with cryotherapy by nurses in a cervical screening study in India. *Br J Cancer*. 2007;96:738–743.
34. Nene BM, Hiremath PS, Kane S, Fayette JM, Shastri SS, Sankaranarayanan R. Effectiveness, safety, and acceptability of cryotherapy by midwives for cervical

intraepithelial neoplasia in Maharashtra, India. *Int J Gynaecol Obstet*. 2008;103:232–236.
35. Sankaranarayanan R, Keshkar V, Kothari A, Kane S, Fayette JM, Shastri S. Effectiveness and safety of loop electrosurgical excision procedure for cervical neoplasia in rural India. *Int J Gynaecol Obstet*. 2009;104:95–99.
36. Rema P, Suchetha S, Thara S, Fayette JM, Wesley R, Sankaranarayanan R. Effectiveness and safety of loop electrosurgical excision procedure in a low-resource setting. *Int J Gynaecol Obstet*. 2008;103:105–110.
37. Martin-Hirsch PL, Paraskevaidis E, Kitchener H. Surgery for cervical intraepithelial neoplasia. *Cochrane Database Syst Rev*. 2000;2:CD001318.
38. Alliance for Cervical Cancer Prevention. *Effectiveness, Safety, and Acceptability of Cryotherapy: A Systematic Literature Review*. Seattle, WA: Alliance for Cervical Cancer Prevention; 2003. Cervical Cancer Prevention Issues in Depth No.1.
39. Pontén J, Adami HO, Bergström R, et al. Strategies for global control of cervical cancer. *Int J Cancer*. 1995;60:1–26.
40. Sparen P, Gustafsson L, Friberg LG, Ponten J, Bergstrom R, Adami HO. Improved control of invasive cervical cancer in Sweden over six decades by earlier clinical detection and better treatment. *J Clin Oncol* 1995;13:715–725.
41. Devi BC, Tang TS, Corbex M. Reducing by half the percentage of late-stage presentation for breast and cervix cancer over 4 years: a pilot study of clinical downstaging in Sarawak, Malaysia. *Ann Oncol*. 2007;18:1172–1176.
42. Harper DM, Franco EL, Wheeler C, et al. GlaxoSmithKline HPV Vaccine Study Group: efficacy of a bivalent L1 virus-like particle vaccine in prevention of infection with human papillomavirus types 16 and 18 in young women: a randomised controlled trial. *Lancet*. 2004;364:1757–1765.
43. Harper DM, Franco EL, Wheeler CM, et al. HPV Vaccine Study group: sustained efficacy up to 4.5 years of a bivalent L1 virus-like particle vaccine against human papillomavirus types 16 and 18: follow-up from a randomised control trial. *Lancet*. 2006;367:1247–1255.
44. Mao C, Koutsky LA, Ault KA, et al. Efficacy of human papillomavirus-16 vaccine to prevent cervical intraepithelial neoplasia: a randomized controlled trial. *Obstet Gynecol*. 2006;107:18–27.
45. Villa LL, Costa RL, Petta CA, et al. Prophylactic quadrivalent human papillomavirus (types 6, 11, 16, and 18) L1 virus-like particle vaccine in young women: a randomised double-blind placebo-controlled multicentre phase II efficacy trial. *Lancet Oncol*. 2005;6:271–278.
46. Villa LL, Ault KA, Giuliano AR, et al. Immunologic responses following administration of a vaccine targeting human papillomavirus Types 6, 11, 16, and 18. *Vaccine*. 2006;24:5571–5583.
47. Koutsky LA, Harper DM. Current findings from prophylactic HPV vaccine trials. *Vaccine*. 2006;24(Suppl 3):S114–S121.
48. The Future II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *N Engl J Med*. 2007;356:1915–1927.
49. Agosti J, Goldie S. Introducing HPV vaccine in developing countries-key challenges and issues. *N Engl J Med*. 2007;356:1908–1910.
50. Dobson S, Dawar M, Scheifele D, et al. Are two doses of HPV vaccine adequate in girls? 25th International Papillomavirus Conference; May 8–14, 2009; Malmö, Sweden.
51. Karanam B, Jagu S, Huh WK, Roden RB. Developing vaccines against minor capsid antigen L2 to prevent papillomavirus infection. *Immunol Cell Biol*. 2009;87:287–299.

SECTION 3

TREATMENT AND PREVENTION OF DISEASE: INFECTIOUS DISEASES

This page intentionally left blank

CHAPTER 13

The Global AIDS Epidemic: Could It Have Been Prevented?

Sandra Demars, MD

In the period October 1980–May 1981, 5 young men, all active homosexuals, were treated for biopsy-confirmed *Pneumocystis carinii* pneumonia at 3 different hospitals in Los Angeles, California. Two of the patients died. . . . Pneumocystis pneumonia in the United States is almost exclusively limited to severely immunosuppressed patients. The occurrence of pneumocystis in these 5 previously healthy individuals without a clinically apparent underlying immunodeficiency is unusual. The fact that these patients were all homosexuals suggests an association between some aspects of a homosexual lifestyle or disease acquired through sexual contact and pneumocystis pneumonia in this population.[1]

The above quote, taken from a report published in 1981 by the Centers for Disease Control and Prevention (CDC), focused on an unusual presentation of illness in a small group of individuals. The presentation of illness was something new and the medical community was indeed perplexed. With hindsight, we now know that this was the beginning of something that would dramatically change things both in the United States as well as the rest of the world. In fact, the world was introduced to the reality of a new and deadly infectious disease that attacked and rapidly killed previously healthy individuals. Twenty-five years later, an estimated 40 million people have been infected with the virus that we now call human immunodeficiency virus (HIV), which we also now know leads to acquired immune deficiency syndrome (AIDS). Tragically, it is estimated that at least 25 million people worldwide have lost their lives to this horrific disease, and millions more probably will do so as well unless immediate and dramatic treatment and prevention measures are instituted. How did such a disease spread so rapidly and in such a deadly way? More important, to what extent have national and global politics played a

This chapter was originally printed as Demars, S. "The Global AIDS Epidemic: Could It Have Been Prevented?" in Finkel, M.L. *Truth, Lies, and Public Health.* Westport, CT: Praeger Publishers, 2007. Reprinted with permission.

role in the spread of this disease? The chapter serves to provide an historical overview of the identification of the HIV/AIDS virus and how the world addressed the challenges that this disease presented.

WHAT IS HIV? WHAT IS AIDS?

Dr. Luc Montagnier in France and Dr. Robert Gallo in the United States first independently isolated a new retrovirus, to be known as HIV, in 1984. To put it simply, HIV is the virus that causes AIDS. It destroys a certain kind of blood cells (CD4+ T cells), which are crucial to the normal function of the human immune system. HIV disease becomes AIDS when a person's immune system is seriously damaged, which can be detected by a test to count CD4+ T cells. Tests can show a strong connection between the amount of HIV in the blood, the decline in CD4+ T cells, and the development of AIDS. A positive test means that an individual is infected by the virus, which can be transmitted to others but *not* through casual contact. Healthy individuals have between 500 and 1,500 CD4+T cells in a milliliter of blood; those with less than 200 CD4+ T cells are considered to have AIDS.

While the virus is not a disease *per se,* it progressively damages the body's immune system. There is no cure as of this writing, and once infected, one is infected for life. But, being HIV positive, or having HIV disease, is not the same thing as having AIDS, although infected individuals can pass on the virus to others through blood, semen, vaginal fluid, or breast milk. In fact, mother-to-child transmission is a significant mode of HIV transmission, especially in the developing world. Estimates are that 90 percent of the HIV-positive children worldwide were infected from their mothers, either in utero, during birth, or from breastfeeding.[2] During pregnancy, maternal blood typically does not mix with fetal blood, but once in a while, a slight hemorrhage can occur at which time HIV can enter fetal circulation. It is more common for HIV transmission to occur during childbirth, where mixing of maternal and fetal blood is much more likely. Other means of HIV transmission include sharing a needle with someone who is infected; an accidental needle stick; and in the early days of the disease, getting a transfusion of infected blood.

Research shows that most people infected with HIV carry the virus for years before AIDS develops. Before 1985, there was no reliable way to test for HIV, but in 1985, the U.S. Food and Drug Administration (FDA) approved the first enzyme-linked immunosorbent assay (ELISA) test kit to screen for antibodies to HIV. The American Association of Blood Banks and the Red Cross began screening the country's blood supply for HIV antibodies to prevent the spread of the virus through blood. In the 21st century, the risk of HIV transmission from blood is rare.

WHERE DID HIV/AIDS ORIGINATE?

Although medical professionals in the United States first became aware of HIV/AIDS in 1981, it is impossible to know when, where, and how the disease first appeared in this country. HIV was moving within and between countries apparently for years before its detection, resulting in many deaths attributed to infections of unknown causes. In 1981, there were 339 cases of AIDS in the United States, and retrospective analyses of medical records show that at least 100 cases of AIDS went

unnoticed before 1981.[3] From our current understanding of the course of the disease, however, we now know that the opportunistic infections that brought these individuals to medical attention possibly presented years after the initial HIV infection occurred.

In the early 1980s, the "African hypothesis for AIDS" began to percolate within the scientific community. The consensus among scientists is that the first AIDS case probably originated in Africa.[4] Although cases were identified as originating from as many as eight West Central African countries, 80 percent of these could be traced back to Zaire (present-day Republic of Congo).[5] Residents of Zaire had known since 1975 that something unusual and deadly was spreading through local communities.[6] When CDC researchers went to Zaire to investigate these strange cases, it was almost immediately apparent to them that AIDS was indeed in Africa, and this was the cause of the high fatality seen in individuals infected with the virus.

Since 1982, scientists have been trying to understand how and why Zaire came to be the primary source of the HIV/AIDS epidemic. In May 2006, a collaborative group of researchers from Cameroon, France, the United Kingdom, and the United States isolated the HIV virus in feces collected from a subspecies of chimpanzees native to west equatorial Africa. This discovery finally provided the evidence to support the long-standing hypothesis that HIV actually is a mutated form of the primate virus simian immunodeficiency virus that gained the ability to infect humans. It is believed that the virus was introduced into the human population either when hunters became exposed to infected blood or when the virus was transmitted locally via Cameroon's Sangha River south to the Congo River and then into Kinshasa, Congo, the geographic episource of the pandemic.[7] The earliest known documented case of AIDS dates from 1959 based on analysis of an unidentified Kinshasa man who apparently was infected and died of the disease.[8]

IN THE BEGINNING: AIDS IN AMERICA

Years before the official discovery of the HIV virus, hospitals on both coasts of the United States began witnessing a frightening and unexplainable increase of rare diseases in otherwise healthy, young homosexual men. The New York University Medical Center in New York City treated two young patients in 1979 for an extremely rare form of skin cancer that was hardly ever seen in people under the age of seventy.[9] By March 1981, at least eight cases of this rare malignancy, Kaposi's Sarcoma (KS), had been documented in young gay men of New York City. Meanwhile, in California, physicians were noting an increase in the number of cases of the rare lung infection *Pneumocystis carinii* (PCP). At the time, most physicians in the United States had never before seen a case of PCP, primarily because most people's immune system can neutralize the causative bacteria. Even more puzzling was the fact that none of the men with PCP had any apparent reason to be immune deficient.[5]

Case reports of homosexual men presenting with strange infections were published in the medical journals. The lay press, too, began to cover stories about the strange disease. The Associated Press and the *Los Angeles Times,* for example, were the first to publish articles in response to the CDC's *Morbidity Mortality Weekly Review* (*MMWR*) June 5, 1981, report. The *New York Times* published its first news story on HIV/AIDS on July 3, 1981.[10] Soon thereafter, the CDC convened a Task

Force on Kaposi's Sarcoma and Opportunistic Infections (OIs) to determine the etiology of these diseases and to identify those who might be at greatest risk for developing these cancers and infections. The task force was urgently trying to figure out the cause of these diseases, because the case fatality death rate among those infected was very high: 40 percent of those infected died. In 1982, the CDC formally established the term AIDS and declared the new disease an epidemic. Also in 1982, the first congressional hearings were held on HIV/AIDS.[10]

Most physicians knew by the end of 1981 that these OIs, previously typical only in immunosuppressed patients, had to be linked to something new and different. These otherwise unexplainable occurrences of certain illnesses initially were termed "gay cancer," because the overwhelming majority of those stricken with these diseases were gay men, but were soon renamed GRID (Gay-Related Immune Deficiency).[11] Almost all of those affected were urban homosexual men; no cases at the time had been reported from outside the homosexual community. The belief was that an infectious agent that was sexually transmitted between gay men caused the disease and that HIV/AIDS could affect only this particular subgroup of society.

Soon, however, individuals outside of the gay community began presenting with similar types of infections. It became clear that a new hypothesis was needed to explain the disease pattern in heterosexual men. The first cases of PCP in injecting-heroin drug users were documented in December 1981. Reports of 10 cases of hemophiliacs with AIDS were publicized in 1982, and by 1985, AIDS had been noted in at least 70 individuals who had received blood transfusions.[5] Arthur Ashe, the champion tennis player, was infected as a result of blood transfusions and eventually died of AIDS. An American teenager, Ryan White, also was diagnosed with AIDS. A hemophiliac, Ryan, too, had received frequent transfusions of blood-clotting proteins purified from human blood products. Ryan was considered an "innocent victim," implying that there were "guilty victims"—that is, homosexuals and intravenous (IV) drug users, individuals who could be blamed for contracting AIDS because of their immoral lifestyle.[12]

The fact that hemophiliacs and recipients of blood transfusions also were contracting the disease seemed to indicate that the infectious agent was most likely transmitted via blood. Data obtained in 1985 (before the availability of HIV testing) showed that 90 percent of individuals transfused with HIV-infected blood became infected with the virus.[13] The CDC estimated that, in the United States in the early years of the epidemic, thousands of individuals contracted transfusion-associated AIDS as a result of untested HIV-infected blood.[14] Clearly, the illness extended beyond people sharing infected needles to inject drugs and engaging in unprotected anal sex.

Another population group, Haitians, appeared also to be at high risk for contracting HIV/AIDS. Probably because of fear and ignorance, in July 1982, being Haitian was officially included as a risk factor for AIDS, thus stigmatizing the Haitian American community. The CDC was accused of racism, and Haiti's tourism industry (and, therefore, the majority of Haiti's economy) significantly suffered.[15] It took three long years before the CDC would remove "Haitian" as an AIDS risk factor, following the accumulation of evidence that Haitian transmission could be traced to heterosexual sex and exposure to contaminated needles.

Fear surrounds most infectious diseases, but AIDS was exceptionally terrifying because once a person started to show signs of illness, they typically did not survive

for much more than a year. As a result of the discovery of AIDS cases in women, children, and other cohorts not considered to be at high risk for this disease, experts modified their disease theories and concluded that personal behaviors were not simply exacerbating a previously seen pathogen, but that an entirely new virus must have emerged. With limited information available at the time, the CDC referred to AIDS in broad and vague terms. So much more needed to be understood about the etiology of the disease, as well as about its transmission, prevention, and cure.

By the mid- to late 1980s, the AIDS epicenters were San Francisco and New York City, although San Francisco had the distinction of having the nation's highest incidence of AIDS until 1994.[16] In the early stages of the epidemic, the number of cases doubled annually. In January 1983, 1,501 individuals had been diagnosed with AIDS, but by the end of the decade, more than 100,000 AIDS cases had emerged and more than 58,000 individuals had died. By 1992, the number of AIDS cases soared to more than 200,000, and estimates indicated that at least 1 million were infected with HIV.[9] The rapid spread of the disease was breathtaking. By 1994, AIDS had become the leading cause of death in the United States among 25- to 44-year-olds. By 2004, the states with the most serious AIDS burden also included Delaware, Georgia, Louisiana, Maryland, Mississippi, and Washington, D.C. The northern Midwestern section of the country had the fewest reported cases.[17]

During the 1980s, there was a shift in who was getting infected. For example, in January 1982, homosexuals represented 96 percent of AIDS cases, whereas IV drug users represented 3 percent. In just 20 months, however, the proportion changed: 71 percent were homosexuals, and 17 percent were IV drug users.[5] Also, the disease was more prevalent among racial minority groups that now account for almost three-quarters of new AIDS cases. In 2003, more than 50 percent of the new HIV diagnoses and 62 percent of children born to HIV-infected mothers were African American.[18] In fact, minority women now are the fastest growing group to be affected by AIDS. Between 1985 and 2003, the percentage of women infected in the United States more than tripled, from 8 percent to 27 percent of the total HIV-positive population. An astonishing 72 percent of all new female HIV cases are African American, and AIDS is now one of the top three causes of death for African American women age 35 to 44 years old.[18] Heterosexual sex accounts for most new HIV cases, with the remainder resulting from injecting drug use. Low-income women, in particular, were more likely to be infected by a husband or steady partner than by a casual sexual partner (see table 13.1). Table 13.2 shows estimated HIV and AIDS diagnoses by race or ethnicity and by year, and table 13.3 shows estimated adult and adolescent males living with AIDS by race or ethnicity by exposure category. In the following three tables, it can be seen that African Americans have a much higher rate of both HIV and AIDS than any other race or ethnic group.

The spread of HIV/AIDS within the United States and around the world was helped by collective denial, silence, and ignorance. The late Jonathan Mann, AIDS researcher, champion of human rights, and director of the United Nations AIDS program, summarized the situation succinctly:

> The dominant feature of [the early years] was silence, for the human immunodeficiency virus (HIV) was unknown and transmission was not accompanied by signs or symptoms salient enough to be noticed. While rare, sporadic case reports of AIDS and seroarchaeological

Table 13.1
Estimated Adult and Adolescent Females Living with AIDS by Race/Ethnicity and Exposure Category, 2003

	Exposure category		
Race/ethnicity	Injection drug use	Heterosexual contact	Other
White, not Hispanic	7,147	10,313	529
Black, not Hispanic	18,164	36,791	1,474
Hispanic	5,802	11,561	416
Asian/Pacific Islander	102	491	51

Source: Centers for Disease Control and Prevention. *HIV/AIDS Surveillance Report.* Vol. 16. Atlanta, GA: US Department of Health and Human Services; 2004.

Table 13.2
Estimated HIV and AIDS Diagnoses by Race/Ethnicity and Year

	Year of diagnosis			
Race/ethnicity	2001	2002	2003	2004
	Estimated HIV diagnoses			
White, not Hispanic	11,242	11,352	11,097	11,806
Black, not Hispanic	21,556	20,237	19,310	19,206
Hispanic	7,714	6,964	7,078	6,970
Asian/Pacific Islander	279	319	367	394
	Estimated AIDS diagnoses			
White, not Hispanic	11,052	11,604	11,657	12,013
Black, not Hispanic	19,473	19,934	20,685	20,965
Hispanic	7,974	7,907	8,632	8,672
Asian/Pacific Islander	381	440	478	488

Source: Centers for Disease Control and Prevention. *HIV/AIDS Surveillance Report.* Vol. 16. Atlanta, GA: US Department of Health and Human Services; 2004.

Table 13.3
Estimated Adult and Adolescent Males Living with AIDS by Race/Ethnicity and Exposure Category, 2003

	Exposure category			
Race/ethnicity	Male-to-male sexual contact	Injection drug use	Male-to-male and drug use	Heterosexual contact
White, not Hispanic	34,797	13,137	11,366	529 I
Black, not Hispanic	52,120	34,797	9,174	21,565
Hispanic	33,717	18,472	4,361	8,204
Asian/Pacific Islander	2,445	314	162	387

Source: Centers for Disease Control and Prevention. *HIV/AIDS Surveillance Report.* Vol. 16. Atlanta, GA: US Department of Health and Human Services; 2004.

studies have documented human infections with HIV prior to 1970, available data suggest that the current pandemic started in the mid- to late 1970s. By 1980, HIV had spread to at least five continents (North America, South America, Europe, Africa and Australia). During this period of silence, spread was unchecked by awareness or any preventive action and approximately 100,000–300,000 persons may have been infected.[3]

THE POLITICS OF AIDS IN THE UNITED STATES IN THE 1980s

Initially, a diagnosis of AIDS was literally a death sentence. Although it took years from time of infection to the development of full-blown AIDS, survival time following an AIDS diagnosis was a mere one and a half years.[19] For many reasons, however, finding a cure for this disease was not a high priority. For the first years of the epidemic, President Reagan never once publicly mentioned AIDS. By the time he did in 1987, 12,000 Americans had lost their lives to the disease. Reagan urged the public not to panic because AIDS was primarily confined to gay men and IV drug users. He did not sympathize with the victims or acknowledge the government's delayed and inadequate response. His focus was to promote abstinence-only education and to bar HIV-positive visitors from entering the country.[20]

The conservative Reagan administration was so opposed to the gay lifestyle and sexual practices that they had no idea how to officially respond to a disease that predominantly affected the homosexual community and IV drug users. In the first term of his presidency, when a reporter asked for a reaction about the CDC's announcement that officially declared AIDS as an epidemic, Press Secretary Larry Speakes uneasily tried to avoid the question by replying "What's AIDS?" and insisting, "I don't know anything about it."[21] Reagan, paralyzed in silence and inaction, grossly underestimated AIDS. Each year between 1982 and 1986, the amount of money that the Reagan administration requested for combating AIDS was less than half of the amount that Congress actually appropriated (which many would claim was not enough in the first place).[22] Many prominent scientists at the time frustratingly spoke of the federal government continuously rejecting their pleas for grants and funding to study the epidemic. In 1986, the National Academy of Sciences came out with harsh criticism of the inadequate response to the AIDS crisis by the U.S. government and called for more funding to find a cure. It took another two years before the National Institutes of Health (NIH) established an Office of AIDS Research. By that time 40,000 people in the United States had died of AIDS. To his credit, Dr. C. Everett Koop, the U.S. Surgeon General from 1982 to 1989, wrote a very accurate, comprehensive, and frank AIDS pamphlet, and indeed, refused to bow to the strong resistance to keeping the very blunt sex and condom talk in the brochure. Dr. Koop was praised by many public health experts for breaking his previous silence and for providing the public with the factual information.[23] "Understanding AIDS" was the first public health alert aimed at reaching the entire population.

While the executive branch chose to ignore the situation, the judicial branch was quite active. In 1986, the Justice Department ruled that anyone with or suspected as having HIV could be legally be fired from their job. That same year, 20 states introduced bills that would prevent individuals infected with AIDS from holding any job that involved food handling. Transmitting the virus also became a crime, and mandatory testing of prostitutes was initiated. In 1987, it became illegal for any

HIV-positive individual to visit or immigrate to the United States. Clearly, the fear of this disease was a driving force in the formulation of laws and legislation.

THE PRIVATE SECTOR TAKES THE LEAD ON AIDS

While the government preferred to ignore the growing crisis, those living with HIV/AIDS needed both support and medications not only to deal with the psychological aspects of the disease, but also to stop the spread of the virus by killing whatever bacteria, virus, parasite, or fungus was causing their OIs. Six men recognized the need for emotional support and founded the first AIDS support organization, Gay Men's Health Crisis (GMHC) in New York City in 1982.[24] The objective was to provide public health education, social support, and counseling services. The debut of the group's AIDS hotline, in the home of a GMHC volunteer, received more than 100 calls in the first night. (The following year, the CDC established the National AIDS Information Line.)[25]

The foresight and innovation of the GMHC led also to the creation of a Buddy Program through which GMHC volunteers assist people living with HIV/AIDS with their day-to-day needs. In its first year, GMHC raised $50,000 for research and distributed 50,000 free copies of its first newsletter to doctors, hospitals, clinics, and the Library of Congress. Across the nation, in San Francisco, another group of men formed The Kaposi's Sarcoma Research and Education Foundation (now known as the San Francisco AIDS Foundation) to provide direct services and educate the public about the new illnesses associated with AIDS. This foundation started a food bank for people with AIDS, held community information sessions, and distributed educational materia1.[26] Most public information campaigns in these years were provided completely by organizations such as GMHC and the San Francisco–based foundation. It is truly amazing how fast these communities stepped in to protect and take care of each other amid fear and hopelessness.

It took three long years for the government to appreciate the scope and importance of these grassroots organizations, and finally in 1984, federal funds were made available for community-based AIDS organizations. Until then, these all-encompassing havens of support survived strictly on donations, fundraisers, and volunteerism. Unfortunately, three years later, the funds were rescinded. Legislation introduced by Sen. Jesse Helms in 1987, which passed overwhelmingly, prevented the government from funding AIDS programs that "encourage or promote homosexual activity." The GMHC, the San Francisco AIDS Foundation, and many other community AIDS organizations designed to educate people about HIV prevention, AIDS prevention, and safe sex practices were run by the gay community for the gay community, and these organizations were restricted from receiving federal funds.

The AIDS community's extraordinary display of activist solidarity and patient empowerment eventually was responsible for spurring changes in AIDS drug development and distribution. In 1985, a group in San Francisco and another in New York City formed the first community-based drug-testing program. These grassroots research programs were unique in that they were formed by groups of gay men infected with AIDS in partnership with doctors. Both the AIDS patients and physicians felt that the federal testing program was moving too slowly. Doctors volunteered to administer experimental drugs and to keep records.[27]

DRUG BREAKTHROUGHS

Infected individuals were so desperate to get their hands on any medication that might possibly slow their disease that underground pharmacies began popping up all around the country to help people get experimental and unapproved drugs. People turned to buyers' clubs to purchase drugs that were still being tested in the United States, as well as tried to obtain drugs that had been approved and were being used in other countries.[28] These clubs were so well developed and extensive that they actually had drug companies abroad specifically manufacturing medications for them. So few options were available for AIDS patients at the time that some physicians would refer their patients to these underground buyers' clubs. In fact, given the situation, the FDA informally allowed drug clubs to bypass FDA regulations. In the summer of 1988, the FDA officially allowed Americans to import unapproved drugs from abroad in small amounts for personal use only.[29]

The first of the wonder drugs that have been shown to be effective against HIV/AIDS was zidovudine (azidothymidine, or AZT). The Burroughs Wellcome Company began selling AZT under the name Retrovir in 1987 at the astronomical price of $10,000 a year.[30] In protest of the exorbitant price demanded by Burroughs Wellcome and the FDA's slow process of approving drugs, the AIDS activist group ACT UP (AIDS Coalition to Unleash Power) staged their first mass demonstration on Wall Street on March 24, 1987.[31] ACT UP, formed in 1987 by the playwright and AIDS activist Larry Kramer, would become synonymous with nonviolent protest against the apathy and neglect of the government regarding fighting the AIDS epidemic. Partially because of ACT UP's perseverance, Burroughs Wellcome's AZT price came down by 20 percent. In 1992, the government finally responded to the activism and protests led and organized by ACT UP to increase drug-processing time, and started accelerated-approval interim licensing to quickly get drugs to people living with AIDS.

Since 1987, an additional government source has been available for AIDS care and treatment funds for low-income people: the AIDS Drug Assistance Programs (ADAPs). ADAPs are authorized under Title II of the Ryan White Comprehensive AIDS Resources Emergency (CARE) Act passed in 1990 to address the unmet health needs of individuals living with HIV/AIDS. Specifically, ADAPs are administered by each state, with the federal government giving each state a certain amount of funding to provide HIV/AIDS-related treatment and prescription drugs to underinsured and uninsured individuals.[32]

AZT and other single-combination drugs extended a patient's life only by a year or two primarily because of the nature of the HIV virus. HIV mutates and eventual accumulations of enough mutations make the virus resistant to the therapeutic effects of these first single-combination medications. It was not until a second class of antiretrovirals (ARVs) was discovered that AIDS treatments actually started showing rapid and dramatic health improvements in people living with AIDS. In 1995, the FDA approved the first protease inhibitor. Protease inhibitors are medications that attack the virus' ability to make the proteins it needs to infect other cells. The combination of both therapies, protease inhibitors and nucleoside analogues, led to an unimaginable success that would be termed highly active antiretroviral therapy (HAART).[33] HAART combines three drugs: two nucleoside analogues and one protease inhibitor.[34]

Physicians worldwide similarly have described the amazement, joy, and disbelief of the Lazarus Effect transformations that almost all patients undergo after only a couple weeks of HAART treatment.[35] Emancipated, exhausted, close to death patients regain their color, weight, energy, and, most important, their cell counts in just a short time on HAART. In the first three years of the HAART era, from 1996 to 1999, the annual number of AIDS-related deaths in the United States fell by 50 percent and decreased another 14 percent by the end of 2002.[36] HAART has dramatically increased survival rates, transforming AIDS from an acute crisis to a chronic disease.[37] In the Western developed nations, current ARV treatment regimens decrease AIDS-related deaths by 80 percent and lengthen lives by 13 years on average.[38] Because of cost, availability, and access, however, resource-poor nations have not seen the life-changing, life-saving effects of HAART. In 2000, 95 percent of the developing world lacked access to ARVs.[39] If in the United States, where the median annual income is $44,000, ARV treatment is unaffordable, medications for countries with average daily incomes of less than $1 was simply unthinkable.

AIDS ON A GLOBAL SCALE

Not a region in the world is untouched by AIDS. But, no region has been harder hit by this epidemic than Africa. There is no single reason as to why the HIV/AIDS epidemic is rampant in Sub-Saharan Africa. Transmission is primarily a result of unprotected heterosexual contact, but poverty, social instability, high levels of sexually transmitted infections, low status of women, sexual violence, ineffective leadership, rapid urbanization and modernization, high rates of migratory labor, and decline of social services are all contributing factors to the region's epidemic.[40]

Without question, AIDS has and continues to ravage Sub-Saharan Africa with speed and scope unseen and unimaginable by any other region of the world. Although Sub-Saharan Africa contains only 11 percent of the world's population, this region is home to 64 percent of all people in the world living with HIV/AIDS (26 million people), 63 percent of all new infections, and 74 percent of all AIDS-related deaths. When attempting to explain the rapid spread of AIDS in Africa, ignorance, fear, and widespread unprotected heterosexual sex immediately come to mind. In many countries, migrant truck drivers and mine laborers have been mainly responsible for the spread of HIV.

The damage that HIV has inflicted on this region of the world will continue to be seen for generations. Twelve million children have been orphaned in Sub-Saharan Africa as a result of AIDS.[41] AIDS has either stopped or reversed the life expectancy progress that had been achieved in some of the most affected countries of Botswana, Lesotho, Malawi, Mozambique, Swaziland, Zambia, and Zimbabwe. But, it is South Africa that is experiencing one of the most severe HIV epidemics in the world. In the 21st century, South Africa holds the dubious honor of being the country with more HIV infections than any other country in the world: the country counts a staggering 5.5 million cases of HIV, and almost 1,000 AIDS deaths every day.[42] In just the past 10 years, the country's prevalence has skyrocketed from 1 percent to a shocking 25 percent.

Politics most certainly has played a role in the proliferation of this disease in South Africa. Between 1993 and 2000, during which time the country experienced

massive political changes, HIV/AIDS went unchecked. AIDS "denialism" and misinformation were rampant. President Mbeki consistently refused to acknowledge that HIV is a cause of AIDS, even after his own son died of the disease. He claimed that ARVs were harmful and unsafe. The high levels of new HIV infections occurring in South Africa reflect the difficulties that AIDS education and prevention campaigns have faced. The social and political climate has not proved accommodating to safe sex messages. An estimated 6 million South Africans are expected to die from AIDS-related diseases over the next 10 years.[43] The average life expectancy in South Africa has dipped below age 50.

Movement between the African continent and the European continent contributed to the introduction of HIV into the Western European population. The first European HIV/AIDS cases were clustered in individuals either of African descent or among people who had spent time in Africa. Unlike in America, 40 percent of these European AIDS cases with connections to Africa were young women (average age: 35 years old).

The link between Haiti and Africa contributed to the spread of the disease across the Atlantic Ocean. The spread of HIV from Kinshasa, Congo, to Haiti could be explained by hundreds of Haitian men who participated in the United Nations Educational, Scientific, and Cultural Organization (UNESCO) education technician program in the Congo between 1960 and 1975. All male participants were single and returned regularly to Haiti for vacation and holidays. The rate of infection was incredibly high, and by 1992, 60 percent of urban Haitian hospital beds were occupied by patients infected with HIV. The urban prevalence rate was 10 percent.[44] From Haiti, it is presumed that the disease spread to the rest of the Caribbean and to the United States as well. As was the case in Africa, transmission throughout the Caribbean was a consequence of unprotected heterosexual contact; however, it is also thought that gay tourism from the United States may have contributed to the spread of HIV/AIDS. Sex tourism provided an easy and ready vehicle for the spread of this deadly disease.

While the focus of the epidemic was on Africa, Haiti, and the United States, HIV/AIDS had not been reported in some of the most populous countries of the world: Russia, India, and China. In time, this situation was to change dramatically and with disastrous consequences. In the 21st century, the AIDS epidemic is full blown in the Russian Federation, as well as in Eastern Europe and Central Asia.[45] Between 2003 and 2005, the number of people living with HIV increased by 25 percent to 1.6 million, and the number of AIDS deaths doubled to 62,000.[46] The Eastern European countries' epidemic is being driven by IV drug use within the male population ages 15 to 29 years old, but as heterosexual transmission increases, HIV prevalence among women is becoming an increasing problem as well. The Central Asian countries of Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan are located where the east and west drug-trafficking routes meet, and thus IV drug use is driving the epidemic in these countries. In contrast, the spread of AIDS in the Czech Republic, Hungary, and Slovenia is primarily fueled by sex between men.

India is on the brink of an unimaginable epidemic without effective, rapid, and widespread interventions.[47] The spread of HIV/AIDS has been fueled by migrant laborers and long-distance truck drivers who engage in sex with multiple partners (sex workers) and who then infect their spouses. Widespread stigma and discrimination against the infected, taboos against discussing sex, the limited control of

women to protect themselves against infection, and punishing poverty all contribute to the AIDS problem in India.

Because of its population size, China had the potential to increase the number of global AIDS cases by 13 million, or 33 percent of the current number of cases worldwide.[48] Initially, Chinese officials rarely reported HIV/AIDS cases, and they considered AIDS to be a foreign disease (only 22 cases were reported in China by 1988). The Chinese government's response to the global epidemic was to ban not only the import of blood products but also the import of secondhand clothing, and to prohibit HIV-positive foreigners from entering the country. China's efforts, however, did not stop the virus from infecting individuals primarily because the disease had taken root in the population as a result of poor screening of blood products. Selling one's blood was a popular means of getting money for many rural Chinese. As such, it is not surprising that 80 percent of China's AIDS cases in the 21st century are among rural residents.[49] Although thus far HIV/AIDS has been localized to a few high-risk groups, the virus now appears to be spreading, with heterosexual relations being the dominant means of transmission. The underestimation of AIDS cases is a major problem in China, and in the 21st century, it is assumed that China has one of the fastest growing AIDS epidemics in the world because of the sheer number of people in this country. Conversely, countries with large Muslim populations have managed to keep infection rates low in comparison with the rest of the world.

In Southeast Asia, Cambodia, Myanmar, and Thailand have the highest rates of HIV infections. Myanmar and Thailand both have epidemics resulting from a combination of IV drug use and commercial sex, whereas Cambodia's epidemic is primarily due to transmission from commercial sex. The epidemic in Thailand began spreading rampantly in the late 1980s.[50] The exponential rise in Thai infection rates between 1987 and 1989 likely was due to HIV spreading into the general population mainly by sex workers and IV drug users. Unlike so many other countries, however, Thailand has been able to gain relative control of its rate of infection. Thailand had been on the front lines of the AIDS epidemic since the 1980s. Thanks to model prevention and public education programs, by the late 1990s, infection rates had either leveled off or started to decline in the various sectors of society that normally are surveyed, although not before close to 1 million people were infected. Thailand has become a pioneer in the distribution of low-cost ARVs, which are available to all for less than $1 a day. At the same time, little headway has been made in easing the harsh stigma associated with AIDS. As more individuals are living longer, they are becoming outcasts in society, which poses a new challenge: What should be done with individuals who have AIDS who are rejected by their families and who cannot find work?

Over the past 25 years the AIDS pandemic has exploded to rank alongside of the influenza pandemic of the 1920s and the bubonic plague in terms of fatalities. In 2005 alone, 3.1 million people lost their lives to AIDS, 4.9 million people became newly infected (700,000 of those were children under 15 years old), and 40.3 million people were living with HIV.[51] As AIDS spreads into new societies, the face of the epidemic continues to change. Worldwide, women are more at risk for HIV infection than men. By the year 2003, women accounted for 50 percent of all people in the world living with HIV.[52] The greater vulnerability and risk witnessed in the female population is exacerbated by many political, cultural, and social factors, but

mostly is a result of sexual violence, gender inequalities in terms of negotiating sex and condom usage, and a female's lack of financial independence. Tragically, women are even biologically more susceptible to HIV infection than men; male-to-female transmission is twice as likely to occur as female-to-male transmission. Women also bear the majority of the impact of AIDS in the sense that they are the ones that act as caretakers for the sick and are more likely to experience discrimination and stigma as a result of the illness, including losing their jobs, an income, and truncated schooling.

THE RACE FOR THE CURE: GLOBAL POLITICS INTRUDES

The first Western world leader to acknowledge publicly not only the impact of AIDS on the developing world, but also the moral duty and necessity of the international community to respond with aid, was French President Jacques Chirac. Chirac, speaking at the International Conference on Sexually Transmitted Disease (STD)/AIDS in Africa in 1997, denounced the fact that where one lives in the world determines whether one can or cannot get medical treatment to prevent certain, untimely death. He went on to call for all nations and people to do what they can to ensure that the benefits of new treatment are extended to deprived populations, and formed the International Therapeutic Solidarity Fund (ITSF). Unfortunately, the world did not stand up to applaud, nor did they line up to donate to the ITSF. Rather, most countries reacted by either reemphasizing the impossibility of providing prescription drugs universally worldwide or simply by continuing to ignore the vastness and gravity of the global AIDS status.

By the turn of the 21st century, the momentum for providing ARVs globally was still in its infancy. To most experts, putting millions of people outside of the industrialized developed world on ARVs did not seem feasible logistically or financially. The health care systems of most developing countries did not have the infrastructure, the workforce, or the budget to initiate complex treatments at more than $12,000 per year (the average price of a HAART regimen in 1996). Instead of looking for a solution, public officials were continuously and repetitively legitimizing their inaction with excuses. The world was paralyzed by what seemed to be too great and expensive a task—the potential to save millions of lives. These excuses would run dry and, since then, not only have the Africans (and other members of poor developing countries) proven the Western Hemisphere wrong by adhering to the strict treatment regimens, but they have shocked and embarrassed these nations by demonstrating even *better* adherence rates than people on ARVs in the Western world (90 percent on average verses 70 percent in the United States). Slowly and painstakingly, AIDS has been begun to transform the international mind-set about health, inequities, and human rights.

Under the United Nations, since 1986, the World Health Organization (WHO) has been primarily responsible for AIDS prevention and treatment.[53] During the 1990s, however, many UN agencies were simultaneously, yet separately, addressing different aspect of the AIDS pandemic. By 1996, it became necessary to coordinate their efforts and the joint United Nations Program on HIV/AIDS (UNAIDS) was formed. From the beginning, UNAIDS faced financial resistance from countries all around the world even though AIDS was recognized as a global threat.

The inequities in access to treatment became glaringly obvious to the world community at the XIII International AIDS Conference held in Durban, South Africa, in 2000. AIDS activists introduced the world to their AIDS crisis when thousands of South Africans demanding treatment marched through the streets of Durban in Africa's largest AIDS march.[54] In Western Europe and the United States, AIDS had become a manageable disease, whereas in Sub-Saharan Africa HIV and AIDS are rampant, devastating communities at a pace unimaginable to developed countries.

Activists not only in South Africa but also in Thailand and Brazil were the first to demand their rights to affordable AIDS drugs.[55, 56] In the case of Brazil, the 1988 constitution clearly outlined health as a human right. The government had been providing free treatment and chemoprophylaxis for OIs since the early 1990s. In 1996, Brazil expanded its program and became the first developing country to provide free universal ARV treatment through its public health system. Brazil's biggest challenge, however, was standing up to the United States and its powerful pharmaceutical industry's attempts to protect profits and enforce intellectual property patents on their medications. The Brazilian government successfully resisted not only the World Bank's demand to stop AZT distribution as a condition of a loan agreement, but also the United States' threat to challenge the country's manufacture of AZT before the World Trade Organization (WTO).[57]

In 2001, a WTO conference held in Doha, Qatar, ruled that countries facing a national public health emergency (for example, the status of AIDS in *many* countries) could manufacture generic drugs for a royalty fee to the patent holder (issue compulsory licensing) or import generic drugs (parallel importing) without permission from the patent holder. This trade agreement (known as the Doha Declaration) was complicated. The only way a country amid a national emergency could treat its citizens is to make the drug itself. In the poorest countries, the governments did not have the means to do this. Only three countries have issued compulsory licenses (Malaysia, Mozambique, and Zimbabwe). The United States and the multinational drug companies are adamant about patent protection, not because they are worried about competing with the generic companies in the poor Sub-Saharan markets, but rather because they are worried about parallel importing of the generics into wealthy countries, thereby upsetting their markets in the Western Hemisphere.

Although falling prices for ARVs hold great significance in changing the course of the epidemic, ARVs still are not free, and thus remain out of reach for many of the world's poorest citizens. It was quickly becoming obvious that the global community needed to substantially increase funding for health programs in poor nations.[58] In April 2001, during the first African Summit on HIV/AIDS, TB (Tuberculosis), and Other Infectious Diseases, UN secretary-general, Kofi Annan, proposed the creation of an international body dedicated to fighting HIV/AIDS. The Global Fund to Fight AIDS, TB, and Malaria was established in 2002 as a public-private partnership aimed at rapidly mobilizing funds for the fight against these three diseases in impoverished countries.

Around the same time as establishment of the Global Fund, the former U.S. president Bill Clinton established the Clinton Foundation HIV/AIDS Initiative (CHAT) with the mission of further lowering ARV prices. CHAT partners with governments and drug companies to secure supplies of ARVs at the lowest prices

possible. Because the price of raw chemicals for ARVs are scale dependent, CHAT business experts partnered with many buyers to guarantee the purchase of chemicals produced in large batches for less. Dropping ARV prices as much as possible, although it is a low-margin business, can still be profitable (as CHAT has proved). CHAT was able to negotiate agreements for large-volume purchases of the lowest possible priced ARVs by the Global Fund and the World Bank with many different generic suppliers. These prices are available to more than 48 countries (representing 70 percent of those living with AIDS), and currently 400,000 people have been put on ARVs under CHAT agreements.[59]

On November 30, 2006, former president Clinton announced that two Indian pharmaceutical companies agreed to cut the prices of HIV and AIDS treatment for children, thus making the drugs more economically accessible worldwide. Treatment would cost less than $60 a year per child, allowing an additional 100,000 HIV-positive children to receive care.[60]

The administration under George W. Bush, too, became involved in providing emergency funding to fight HIV/AIDS in resource-poor countries. With the announcement of the Presidential Emergency Plan for AIDS Relief (PEPFAR), President Bush committed $15 billion over five years (2003–2008) to fight the AIDS epidemic in 15 focus countries. Within these countries, PEPFAR's goals were to provide care to 10 million infected people, including orphans and vulnerable children. Critics have come out in opposition to many aspects of PEPFAR. PEPFAR is a unilateral project, and many people feel that the AIDS epidemic would be best addressed in a multilateral, coordinated fashion. Another aspect of PEPFAR that many people protest against is the condition that PEPFAR funds can be used only to purchase medications that are approved by the FDA. Meanwhile the Global Fund, governments, nongovernmental organizations, and all other organizations purchasing ARVs for poor nations buy drugs approved by the WHO prequalification process.

The distribution of ARVs to resource-poor countries has been sluggish. To jumpstart the movement, WHO and UNAIDS joined forces in 2003 to cosponsor the "3 by 5 Initiative," whose goal is to make universal access of HIV/AIDS prevention and treatment accessible as a human right for those in need.[61] Thanks to the perseverance and humanitarian efforts of Treatment Action Campaign, *Médecins Sans Frontières*, CHAT, UNAIDS, WHO, and all the activists and governments that have stood up to patent pressures, the price of ARVs has fallen from an annual price of approximately $12,000 to less than $140 in less than a decade. Additionally, the cost of the tests to monitor AIDS patients has decreased by 80 percent.

SO NOW WHAT?

The search for an AIDS vaccine rallied scientists around the globe. Yet, only 1 percent of the total spending on health product development is spent on the AIDS vaccine.[62] The majority of researchers exploring a vaccine are affiliated with government agencies: only two pharmaceutical companies currently are investigating an HIV vaccine primarily because vaccinations are not cost-beneficial for big pharmaceuticals. In fact, the chief of AIDS research at the NIH commented that he does not think drug companies have the incentive to create a vaccination, and they mostly likely will wait until the government develops one. The Pharmaceutical Research

and Manufacturers of America, however, insists that the drug companies are firmly committed to developing a vaccine.[63]

The newly formed Global HIV Vaccine Enterprise, a new alliance of agencies suggested by the Gates Foundation in 2003 and endorsed by the heads of the G8 nations at their summit in 2004, is beginning to address the shortfalls of HIV vaccine research.[64] The HIV Vaccine Initiative, for the first time ever in a research endeavor, has advocated the sharing of intellectual property, specimens, and data among the international partners. In fact, these are conditions for the grant awardees. Another mandate is the continued evaluation of progress and failures to ensure that only the leads with the most promise are explored. The initiative addresses all facets of vaccine development, from basic science to developing the infrastructure necessary to test the efficacy of vaccination candidates. Hopefully, this coordinated, collaborative initiative will provide the leadership, organization, and direction that vaccine research has lacked thus far.

From the beginning of the epidemic, AIDS has been a controversial, politically charged issue that has made class, race, gender, socioeconomic status, and geographic inequities painfully obvious. Fear, stigma, ignorance, and apathy about AIDS have increased human rights violations against people living with HIV/AIDS, and human rights violations facilitate the spread of AIDS. The UN Commission on Human Rights, after declaring health as a human right, began to take notice, and in 2000 and 2001, it adopted two resolutions addressing HIV/AIDS that clearly stated that people around the world living with AIDS had the right to ARV treatment.

CHALLENGES AHEAD

In the 21st century, progress fortunately has been made in terms of treating those who are HIV positive and those who have full-blown AIDS. But, much remains to do in terms of treatment, prevention, and education. Less than 10 years after the development of the "cocktail" of drugs now widely used to treat AIDS in the Western Hemisphere, only a small percentage of those in Africa and Asia who need the drugs have access to them. The single-most important impediment is the exorbitantly high cost of the medications. Regarding prevention, until recently, almost all foreign-funded AIDS programs in Africa and Asia were directed toward prevention. This prevention-only approach is not effective without treatment.[65] Most people infected with AIDS do not know their HIV status, which limits outreach and counseling and leads to the continued spread of the disease.

AIDS education must be an integral part of any country's fight against the disease. As recently as 2004, even after 25 years of AIDS in America, basic misconceptions about the disease still exist. Shockingly, nearly 37 percent of Americans still think HIV can transmitted by kissing, 22 percent by sharing a drinking glass, and almost 16 percent think that they can be infected by touching a toilet seat.[66] In other countries, especially the developing nations, myths, disinformation, and denial remain prevalent.

Despite the lack of knowledge about the disease, U.S. sentiment is to do more to combat the disease. Two-thirds of Americans (63 percent) said that the U.S. government is spending too little at home to fight HIV/AIDS—up from 52 percent in 2004. This willingness to spend more may stem from a belief that increased

spending on prevention (62 percent) and testing (59 percent) will lead to meaningful progress in slowing the epidemic. In addition, 6 in 10 Americans agree that the United States is a global leader and has a responsibility to help fight HIV/AIDS in developing countries—up from 44 percent in 2002. In addition, more than half (56 percent) think the United States is spending too little on HIV/AIDS in developing countries—up from 31 percent in 2002.

While there is increased support to do more, U.S. citizens seem to recognize the big challenges in confronting HIV/AIDS worldwide. Four in ten Americans think the world is losing ground on the epidemic; overwhelming majorities think most people with HIV in developing countries do not get needed medication (92 percent) and that most people at high risk do not have access to needed prevention services (81 percent). Meanwhile, as the world watches, the AIDS pandemic grows, outstripping prevention efforts and treatment programs, and causing tens of thousands of deaths each year.

REFERENCES

1. Gottlieb MS, et al. Pneumocystis pneumonia: Los Angeles. *MMWR*. 1981;30:250–252.
2. Mofenson LM. Interventions to prevent mother to child HIV transmission: antiretroviral prophylaxis trials in resource-rich and resource-limited settings. *UpToDate*. April 26, 2005.
3. Mann JM. AIDS: worldwide pandemic. In: Gottlieb MS, et al., eds. *Current Topics in AIDS*, vol. 2. New York: John Wiley and Sons; 1989.
4. McNeil DC. US urges HIV tests for adults and teenagers. *New York Times*. September 22, 2006: A1.
5. Leibowitch J. *A Strange Virus of Unknown Origin*. New York: Ballantine Books; 1985.
6. Vangroenweghe D. The earliest cases of human immunodeficiency virus type I group M in Congo-Kinshasa, Rwanda, and Burundi and the origin of acquired immune deficiency syndrome. *Philos Trans R Soc London*. 2001:356:923–925.
7. Keele BF, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*. 2006;313:523–526.
8. Altman KL. Chimp virus is linked to HIV. *New York Times*. May 26, 2006.
9. Arno PS, Feiden KL. *Against the Odds: The Story of AIDS Drug Development, Politics, and Profits*. New York: HarperCollins; 1992.
10. Kaiser Family Foundation. The global HIV/AIDS epidemic: a timeline of key milestones. http://www.kff.org.
11. Gay Men's Health Crisis. http://www.gmhc.org/aboutitimeline.html.
12. http://www.ryanwhite.com.
13. Kleinman S. Transfusion transmitted HIV infection and AIDS. *UpToDate*. August 20, 2004.
14. Centers for Disease Control and Prevention. Epidemiologic notes and reports possible transfusion-associated acquired immune deficiency syndrome, AIDS conference. *MMWR*. 1982;31:652–654.
15. Simon M, Haiti's Tourism Board. The stigma of DIAS is fatal. *New York Times*. November 29, 1983.
16. Cochrane M. *When AIDS Began: San Francisco and the Making of an Epidemic*. New York: Routledge; 2004.
17. Dowdle WR. The epidemiology of AIDS. *Public Health Rep*. 1983;98:308–312.
18. Elizabeth Glaser Pediatric AIDS Foundation. HIV/AIDS fact sheets: the global pandemic, the epidemic in the united states, and women, girls, HIV and AIDS. http://www.pedaids.org.

19. Ippolito G, et al. The changing picture of the HIV/AIDS epidemic. *Ann N Y Acad Sci*. 2001;946:1–12.
20. Hooper C. Critics unimpressed with Reagan's AIDS gambit. *United Press International*. April 6, 1987.
21. AIDS Treatment News. http://www.aidsnews.org.
22. Beegan D. Studds says Reagan has shown little concern over AIDS. Associated Press. September 19, 1985.
23. Office of the Surgeon General. http://www.surgeongeneral.gov/library/history/biokoop.httn.
24. Gay Men's Health Care. http:www.gmhc.org/about/timeline.html.
25. Centers for Disease Control and Prevention. Evolution of HIV/AIDS Prevention Programs: United States, 1981–2006. *MMWR*. 2006;55:597–603.
26. San Francisco AIDS Foundation. http:www.sfaf.org/aidstimeline/index_all.cfm?type+a.
27. Kolata G. Doctors and patients take AIDS drug trial into their own hands. *New York Times*. March 15, 1988.
28. Kolata G. Patients going underground to buy experimental drugs. *New York Times*. November 4, 1991.
29. Kolata G. A market for drugs: AIDS patients and their above-ground underground. *New York Times*. July 10, 1988.
30. US Food and Drug Administration. Approval of AZT. *Public Health Service*. March 20, 1987. http://www.fda.gov/bbs/topics/NEWS/NEW00217.html.
31. Rimmerman CA. ACT UP. http://www.thebody.com. Published 1998.
32. US Department of Health and Human Services, Health Resources and Services Administration. ADAP Fact Sheet. http://www.hab.hrsa.gov/programs/factsheet/adap.
33. Watensky RP, et al. The survival benefits of AIDS treatment in the United States. *Infect Dis*. 2006;194:1–19.
34. Pomerantz RJ, Horn DL. Twenty years of therapy for HIV-1 infection. *Nature Med*. 2003;9:867–873.
35. Schwartlander B, et al. The 10-year struggle to provide antiretroviral treatment to people with HIV in the developing world. *Lancet*. 2006;368:541–546.
36. Kellerman S, et al. Changes in HIV and AIDS in the United States: entering the third decade. *Curr HIV/AIDS Rep*. 2004;1:153–158.
37. Kalichman SX, ed. *Positive Prevention: Reducing HIV Transmission among People Living with HIV/AIDS*. New York: Kluwer Academic/Plenum Publishers; 2005.
38. Kim JY, Farmer P. AIDS in 2006: moving toward one world, one hope? *N Engl J Med*. 2006;355:645–647.
39. Cowley G. The life of a virus hunter. *Newsweek*. May 15, 2006: 63–65.
40. Joint United Nations Program on HIV/AIDS. Basic facts about the AIDS epidemic and its impact. http://www.unaids.org. Published November 2004.
41. Kaiser Family Foundation. HIV/AIDS policy fact sheet: HIV/AIDS epidemic in Sub-Saharan Africa. http://www.kff.org. Published October 2005.
42. Joint United Nations Program on HIV/AIDS. 2006 report on the global AIDS epidemic. http://www.unaids.org. Published 2006.
43. HIV/AIDS in South Africa. http://www.southafricainfo.org.
44. Pape J, Johnson WD. AIDS in Haiti: 1982–1992. *Clin Infect Dis*. 1993;17: S341–S345.
45. Ghys PD, et al. Growing in silence: selected regions and countries with expanding HIV/AIDS epidemics. *AIDS*. 2003;17:S45–S50.
46. Joint United Nations Program on HIV/AIDS, World Health Organization. AIDS epidemic update: special report on HIV prevention. http://www.unaids.org. Published December 2005.
47. Pais P. HIV and India: looking into the abyss. *Trop Med Int Health*. 1996;1:295–304.

48. Cohen J. HIV/AIDS in China: poised for a takeoff? *Science*. 2004;304(5676):1430–1432.
49. He N, Detels R. The HIV epidemic in China: history, response, and challenge. *Cell Res*. 2005;15:825–832.
50. Phoolcharoen W. HIV/AIDS prevention in Thailand: success and challenges. *Science*. 1998;280(5371):1873–1874.
51. Quinn TC. The global human immunodeficiency virus pandemic. *UpToDate*. September 17, 2004.
52. Joint United Nations Program on HIV/AIDS. 2004 report on the global AIDS epidemic executive summary. http://www.unaids.org. Published June 2004.
53. Joint United Nations Program on HIV/AIDS. International programs, initiatives, and funding issues. http://www.unaids.org. Published June 2005.
54. D'Adesky AC. *Moving Mountains: The Race to Treat Global AIDS*. New York: Verso; 2004.
55. Kazmin A. Thai victory on AIDS drug patent paves the way for others. *Financial Times*. February 20, 2004.
56. Teixeira PR, et al. Antiretroviral treatment in resource-poor settings: the Brazilian experience. *AIDS*. 2004;18:S5–S7.
57. Berkman A, et al. A critical analysis of the Brazilian response to HIV/AIDS: lessons learned for controlling and mitigating the epidemic in developing countries. *Am J Public Health*. 2005;95:1162–172.
58. Feachem RGA, Sabot 0J. An examination of the global fund at 5 years. *Lancet*. 2006;368:537–540.
59. Dugger CW. Clinton makes up for lost time in battling AIDS. *New York Times*. August 29, 2006.
60. Clinton cuts a deal for kids with AIDS. Associated Press. November 30, 2006.
61. World Health Organization. The '3 by 5' initiative. http://www.who.int.
62. The State of Global Research. http://www.iavi.org.
63. Solomon J. Sides differ over failure to develop AIDS vaccine. *Hartford Courant*. December 26, 2005: A7.
64. http://www.hivvaccineenterprise.org/plan/l.html.
65. Burkhalter H. The politics of AIDS: engaging conservative activities. *Foreign Affairs*. http://www.foreignaffairs.org. Published 2004.
66. Kaiser Family Foundation. Survey of Americans on HIV/AIDS. http://www.kff.org.

This page intentionally left blank

CHAPTER 14

Viral Hepatitis: Epidemiology, Prevention, Diagnosis, and Treatment

Brandon Aden, MD, MPH

Viral hepatitis remains an important public health concern in both developing and developed countries. Many illnesses and conditions can cause inflammation of the liver—for example, drugs, alcohol, chemicals, and autoimmune diseases. Additionally, viruses, for example, the virus of mononucleosis and the cytomegalovirus, can inflame the liver, resulting in symptoms that range from mild to severe. The few specific hepatitis viruses that primarily attack the liver include types A, B, C, D, E, F (not confirmed), and G. The most common hepatitis viruses, types A (HAV), B (HBV), and C (HCV), are the subject of this chapter. Hepatitis viruses are found in every part of the world and often cause infections ranging in severity from acute infections that are asymptomatic to fulminate and chronic infections, which in some instances lead to cirrhosis and hepatocellular carcinoma (HCC) and even death. Almost half (40 to 45 percent) of chronic liver disease in the United States can be attributed to viral hepatitis, with HCV leading to as many as 12,000 deaths each year and HBV contributing to more than 3,000 deaths each year.[1] This chapter focuses on the global and national burden of HAV, HBV, and HCV. Discussion of screening, prevention, and treatment highlights the tremendous effect vaccination has had on reducing the incidence of the disease.

BURDEN OF DISEASE

The World Health Organization (WHO) estimates that 1 in 12 persons (480 to 520 million) worldwide have chronic infection with either HBV or HCV,[2, 3] and the U.S. Centers for Disease Control and Prevention (CDC) estimates that more than 5 million U.S. residents have such chronic infections.[4] In the United States, an estimated 1.2 million Americans are living with chronic HBV and 4 million are living with chronic hepatitis C.[5–10] Worldwide an estimated 1 million deaths occur secondary to complications of either HBV or HCV.[11] Each year, an estimated 25,000 persons become infected with HAV, 43,000 with HBV, and 17,000 with HCV.[10] Yet these infections often go undetected and untreated because patients and physicians may be unaware of who is at risk or may fail to pursue testing.

Although prevention efforts have dramatically reduced the incidence of HAV, HBV, and HCV, chronic HBV and HCV remain prevalent not just in the United States, but also worldwide. Increasing international travel and immigration of infected individuals to the United States is adding to the spread of this disease. The vast majority of these infected individuals are asymptomatic and unaware of their infection status. As a consequence, in the United States, a growing cohort of individuals infected with HCV is reaching an age at which liver complications present. Approximately 15 to 40 percent of individuals with chronic HBV infection develop serious complications,[12] and 20 percent of persons with chronic HCV develop cirrhosis over 20 to 30 years.[13] Viral hepatitis is the most common reason for liver transplantation.

HEPATITIS A

HAV is one of the most frequently reported vaccine-preventable diseases. The WHO estimates 1.4 million cases of HAV infections occur annually worldwide.[2] Although mortality is rare, the cost of morbidity is high. In the United States, for example, the annual direct cost associated with HAV is $200 million, with indirect costs, including days of missed work, contributing to the economic burden of this disease.[14] HAV can affect anyone as it is spread from person to person, especially through fecal contamination of food and water. Food preparers who are infected, for example, can pass the virus if they do not wash their hands with soap and water after having a bowel movement. Eating raw or partially cooked shellfish contaminated with HAV is another way the virus spreads. Those at higher risk of HAV include travelers to countries with high rates of HAV, men who have sex with men (MSM), intravenous drug users (IDUs), individuals with chronic liver disease, and children living in communities with high rates of the disease.[1]

Diagnosis of HAV is through a blood test for IgM antibodies, which become detectable 5 to 10 days before the onset of symptoms and can persist for up to six months after infection. It is possible to experience mild or no symptoms whatsoever, but even if this is the case, the person's feces still will be infectious to others. Many people who become infected with HAV will have symptoms that include a short, mild, flu-like illness; nausea, vomiting, and diarrhea; loss of appetite and weight loss; and jaundice (yellow skin and whites of eyes, darker yellow urine, and pale feces). Jaundice typically develops 7 to 10 days after onset of symptoms. Interestingly, by the time jaundice develops, the other symptoms may have resolved or significantly improved. Jaundice can last for several weeks. There is no specific treatment for HAV as the virus will clear on its own in a few weeks without serious aftereffects. Complications with HAV infection are rare and permanent damage to the liver is unlikely. Once recovered, an individual is immune for life to HAV through the presence of the IgG antibody.[15]

In the United States, HAV rates historically varied cyclically with peaks every 10 to 15 years. The epidemiology of HAV infections has changed significantly since 1995, however, with the approval of an effective vaccine. In 1996, the CDC Advisory Committee on Immunization Practices (ACIP) recommended vaccination for individuals considered to be at risk for HAV infection, and in 1999, recommended vaccination for all children living in 11 states with rates greater than 20 cases per

100,000 as well as considered for children living in six states where the rates were between 10 and 20 cases per 100,000.[1, 16, 17] There was a tenfold decrease in reported incident cases between 1997 and 2007, with 31,582 cases of HAV reported in 1995 and an incident rate of 12 per 100,000. In 2007, only 2,979 cases were reported with an incident rate of 1 per 100,000.[1, 17, 18] This dramatic decrease can be directly attributed to the routine vaccination for HAV for all children in the United States as part of childhood vaccination schedule. The HAV vaccine is made from inactive HAV (synthetic) and has been shown to be highly effective in preventing HAV infection.

The rate of HAV infection has varied by race and ethnicity with rates significantly higher among Native Americans, Alaska Natives, and Hispanics. Before 1996, for example, the incidence rate among Native Americans was greater than 60 cases per 100,000, a rate far higher than that for any other group. After the introduction of the vaccine in 1996, rates plummeted for all ethnic and racial groups.[1, 17, 18] Globally, the risk of HAV infection varies by place. Risk predominately relates to sanitary conditions and hygiene; countries in Africa, South America, Mexico, Asia, and Southeast Asia in particular are categorized as high risk with a lifetime risk of infection in excess of 90 percent. Typically, in high-risk countries, infection more often than not occurs in childhood. In low-risk regions, including the United States, Canada, Western Europe, Australia, New Zealand, Japan, the Republic of Korea, and Singapore, the risk is largely confined to high-risk groups.[19–23] With the adoption of the HAV vaccine, logically high-risk countries should see a decrease in the incidence of HAV.

Since an effective vaccine exists, prevention of HAV infection should be a top global goal. HAV vaccine is available in two formulations and in a formulation in combination with HBV. HAVRIX (GlaxoSmithKline), VAQTA (Merck), and Twinrix (GlaxoSmithKline) are all on the market. HAV immunization is given in a series of injections at time zero and six months and one year later. Universal childhood vaccination for HAV is now recommended in the United States with good effect, and this should be a public health goal on the world level. Adults without a history of vaccination who are traveling to countries where HAV is endemic or whose sexual practices are likely to put them at risk should be offered vaccination. Expedited vaccination protocol with Twinrix is approved with administration of vaccine at 0, 7, and 21 days and at 12 months.[24, 25] The expedited vaccination protocol should be used for persons who are intending to travel to countries with high incidence and would not be able to complete the standard schedule before travel. If travel is expected to occur within the next two to four weeks, immunoglobulin can be administered to offer short-term protection. Individuals not previously vaccinated who are exposed to the virus should be given postexposure prophylaxis. Current recommendations are for persons 1 to 40 years of age to receive the age-appropriate HAV vaccination within two weeks of exposure. Individuals who are less than 12 months or more than 40 years of age, immunocompromised persons, persons with chronic liver disease, or persons allergic to the HAV vaccine should receive immunoglobulin.[26, 27]

HEPATITIS B

HBV is the most common of the viral liver infections with an estimated 350 million people infected globally of whom approximately 1 million die each year

from complications from HBV.[28–30] In the United States, 800,000 to 1.4 million people are estimated to be chronic HBV carriers; every year, 3,000 to 5,000 Americans die from complications of chronic HBV infections.[10, 31] Chronic infection with HBV results in liver cell damage, scarring of the liver (fibrosis and cirrhosis), and increased risk of HCC. HBV infection has two phases: acute and chronic. Acute HBV occurs shortly after exposure to the virus and includes the first six months of infection. Few individuals during this period develop a severe, life-threatening form of acute hepatitis called fulminant hepatitis. Chronic HBV is an infection with HBV that lasts longer than six months. Once the infection becomes chronic, it is typically lifelong. Of those who develop a chronic infection, 15 to 25 percent will develop chronic liver disease, including fibrosis, cirrhosis, and HCC.[10] In addition to those with active disease, individuals may develop a carrier state in which they do not have active disease but may transmit the infection. The risk of the infection becoming chronic is highly dependent on the age of acquisition. Those who acquire the infection perinatally or as young children have significant risk of developing chronic infection (70 to 90 percent), whereas those who acquire the infection as an adolescent or an adult have a much lower risk of developing a chronic infection (10 to 20 percent).[32]

HBV is transmitted through blood and bodily fluids by percutaneous or mucosal exposure. The virus may be transmitted through intravenous drug use, vertical transmission from infected mother to infant at time of delivery, sexual contact as well as nonsexual contact, and by health care exposure (for example, needle sticks, dialysis, blood transfusions).[1, 17, 18] Many individuals with an acute HBV are asymptomatic, although the risk of symptoms increases with age of acquisition. Those who do develop symptoms may experience mild and flu-like symptoms, loss of appetite, nausea, fatigue, muscle, or joint aches, and possibly jaundice.[10] If jaundice does develop, it typically will occur one to two weeks after the onset of initial symptoms. In the majority of cases, however, jaundice does not occur. Testing for HBV should include the following: HBsAg, anti-HBc, and anti-HBs. During the acute infection, HBsAg will be the initial first positive test at four weeks, with IgM anti-HBc developing later in the course of the disease.[33, 34]

The ACIP, in the mid-1980s, made it first recommendation for vaccination of people at high risk for HBV—for example, individuals having sex with someone infected with HBV; a man having sex with a man; persons living in the same house with someone who has chronic HBV infection; persons whose job involves contact with human blood; IDUs; and persons traveling to areas where HBV is endemic. The incidence of HBV has decreased steadily with the introduction of the HBV vaccine, especially after 1991 when the universal childhood vaccination for HBV was introduced.[35, 36] Before universal vaccination, approximately 300,000 individuals, including 24,000 infants, were infected with HBV annually. Between 1997 and 2007, the reported incidence of HBV decreased from 8.5 cases per 100,000 to 1.5 cases per 100,000. The most significant decrease was among individuals under the age of 15; the incidence decreased from 1.2 cases per 100,000 to 0.02 cases per 100,000. Yet, while the availability of the vaccine has contributed significantly to the decrease in the incidence of HBV, a substantial number of cases are diagnosed each year, especially among the high-risk population. Specifically, in 2007, individuals between the ages of 25 and 44 years have the highest incidence of HBV.

Incidence among ethnic and racial groups shows a different pattern, however. Among those groups with higher incidence (Asian and Pacific Islanders), vaccination has reduced the number of new cases substantially.[1, 17, 20, 28–30, 37–40]

Globally, an estimated 2 billion people have been exposed to HBV infection. The incidence and prevalence is highest in Africa, Asia, and the Western Pacific. Incidence and prevalence is lowest in North America, Australia, and Western Europe. Interestingly, HBV genotypes differs by region, with genotype A and D found worldwide, B and C in East and Southeast Asia, D in Western Africa, F among the indigenous populations of Central and South America, and G in the United States.[28–30, 41]

Currently, there are no recommendations for treatment during an acute HBV infection, but patients should be monitored for spontaneous clearance of the virus. The goal of treatment in chronic infection is to suppress replication and thereby limit the immune system–mediated hepatic inflammation and progression to cirrhosis and development of HCC. Current modalities of treatment include interferon-alpha or nucleos(t)ide analog therapy. Historically, both interferon-alpha-2a and interferon-alpha-2b have been used to treat HBV, but they are no longer first-line therapy. Currently five nucleos(t)ide analogues are available for HBV treatment: adefovir, entecavir, lamivudine, telbuvidine, and tenofovir. After evaluation, treatment is recommended if HBV viral DNA (deoxyribonucleic acid) levels are greater than 2,000 IU/ml or alanine aminotransferase (ALT) levels are above the upper limit of normal (ULN), and if the individual has moderate to severe necroinflammation or fibrosis on liver biopsy. In those individuals who are either HBeAg+ or HBeAg−, the primary endpoint is a sustained HBsAg loss with or without seroconversion of HBsAb. Among HBeAg+ patients, attaining complete viral suppression and seroconversion with HBeAg− or HBeAb+ also is acceptable. Among patients who are HBeAg−, who do not have loss of HBsAg, or patients with cirrhosis, long-term treatment should be maintained with goal of suppressing the virus.[42]

Current recommendations are for the use of entecavir or tenofovir as monotherapy because both are potent inhibitors of HBV with high barriers to resistance. Telbuvidine is potent medication, but adefovir and lamivudine are less efficacious and all have a low barrier to resistance.[42–62] All patients with chronic HBV infection should be evaluated at regular intervals with imaging as they are at increased risk of developing HCC, and this risk is independent of the level of fibrosis in the liver.[63–65]

Patients with viral coinfections require special consideration. Of the estimated 40 million individuals worldwide who are HIV positive, an estimated 2 to 4 million also have chronic HBV infection. Among those with HIV infection in the United States and Western Europe, between 6 and 14 percent are HIV positive.[37, 42, 66–68] The indications for treatment of HBV for HIV positive individuals are the same as mono-infected individuals. But, beginning treatment for HIV in individuals previously untreated actually may lead to hepatitis flare resulting from immune reconstitution; therefore, treatment for HIV and HBV should be started simultaneously, typically with tenofovir-based therapy. If HIV-positive patients are not to be treated for their HIV, adefovir or telbuvidine should be used for HBV treatment because they do not have activity against HIV. Lamivudine, entecavir, and tenofovir are contraindicated as single agents.[69–76] Typically, patients with HCV coinfection have low or undetectable levels of HBV DNA, and HCV is primarily responsible for

active inflammation in the liver. Such patients should undergo standard treatment of HCV, which is interferon-alpha based. If, after treatment for HCV, sustained virologic response (SVR) is achieved, but the HBV reactivates, treatment of the HBV should be with a nucleos(t)ide analogue agent.[77, 78] HDV is responsive to treatment with interferon-alpha, but not the nucleos(t)ide analogues, thus patients with HDV coinfection should be treated with interferon-alpha.

Other conditions may require special consideration. Although the majority of adults with acute HBV infection recover spontaneously, some patients might develop fulminant hepatitis or have a protracted severe course and require treatment. Treatment for pregnant women generally is not recommended; however, evidence indicates that treatment during the third trimester among women with high viremia may reduce the risk of intrauterine and perinatal transmission when given in addition to immune globulin and HBV vaccine. After delivery, women should be monitored closely as exacerbation of the HBV infection may occur. HBV infected mothers may breastfeed.[79–81]

Fortunately, the HBV vaccine is now part of the routine vaccination for all children born in the United States. Five versions of the vaccine are available; two of which are HBV alone and three others that are administered in combination with other vaccines. Recombivax HB (Merck); Enerix-B (GlaxoSmithKline); Comvax (Merck), which is HBV plus *Haemophilus influenzae* type b; Pediarix (GlaxoSmithKline), which is HBV plus diphtheria, tetanus, acellular pertussis adsorbed, and inactivated poliovirus vaccine; and Twinrix (GlaxoSmithKline), which is HBV plus HAV. Infants born to mothers who are chronically infected with HBV, particularly mothers who are known to be HBeAg+, should receive Hepatitis B Immune Globulin (HBIG) and a single dose of HBV vaccine. The infants should then continue the HBV vaccine series as scheduled. If a child is born to a mother of unknown status, a single dose of HBIG can be given. HBV vaccination should be given to those living in the same household as a chronically infected individual; those who are exposed to blood through their job; drug users; and those who have multiple sex partners. Vaccination provides protection for more than 15 years. HBV booster shots are not necessary.[10, 35, 36, 38]

HEPATITIS C

HCV is a slowly progressing disease that can lead to serious liver damage. It is estimated that 3 percent of the world's population, or 180 to 200 million, are infected with HCV.[5–9] Most cannot afford treatment and most will become chronically infected and risk developing cirrhosis and liver cancer. The incidence and prevalence of HCV infections in the United States and Western Europe is relative low (1 percent to 2 percent). Eight countries, Bolivia, Burundi, Cameroon, Egypt, Guinea, Mongolia, Rwanda, and Tanzania, have an HCV prevalence above 10 percent and another seven countries (predominantly in Africa and Southeast Asia) have an HCV prevalence between 5 percent and 10 percent.[82–84] In the United States, it is estimated that 1.8 percent of the population, or at least 4 million individuals, are infected with HCV.[6, 7] It is estimated that approximately 36,000 new infections occur each year, contributing to 10,000 to 20,000 deaths annually.[85] The true incidence of this viral disease, however, is difficult to estimate

because most cases are asymptomatic until they reach a more serious point and require medical treatment.

HCV infections can and do develop into chronic infection. Approximately 80 percent of HCV acute infections develop into chronic infections. As such, HCV is the leading reason for liver transplantation, accounting for 40 to 50 percent of transplants in the United States. Projected mortality from HCV-related complications is expected to triple during the next 10 to 20 years, overtaking the number of deaths from AIDS. Between 2010 and 2019, 193,000 deaths and $10.7 billion in direct medical costs are estimated.[86–88] Among persons with HIV, HCV is the fastest growing opportunistic infection, affecting some 30 percent of persons with HIV in the United States, including 72 to 95 percent of IDUs with HIV.[89] An estimated 70 to 90 percent of IDUs in the United States are infected with HCV,[90, 91] although the prevalence among those who have been injecting for less than five years has fallen considerably since the implementation of education, outreach, and needle exchange programs.[92, 93]

Most of those infected with HCV do not have symptoms, and if symptoms are present, they generally are mild—for example, may include nausea, vomiting, fatigue, malaise, arthralgias, myalgias, headache, cough, coryza, and low-grade fevers. Although the majority of patients with acute HCV infection do not develop symptoms, it is important to recognize the likelihood of acute infection because treatment administered during this the acute phase (the first six months after infection) improves outcomes compared with treatment initiated after six months postinfection. Diagnosis of HCV includes testing for HCV antibodies through a simple and specific blood test; the enzyme immunoassay test (EIA) that detects anti-HCV is a highly sensitive test (sensitivity at 95 percent). If the test is negative, the infection could be present but antibody levels may not be high enough to be detected. The HCV antibody does not develop until three to eight weeks after initial infection, a "window period." During this time period, HCV RNA testing can be done using either qualitative or quantitative test. Qualitative testing is a more sensitive assay and can be done with polymerase chain reaction (PCR) or transcription-mediated amplification (TMA) testing.[94] Liver enzyme levels should be tested as well. ALT and aspartate aminotransferase (AST) levels are released when liver cells are injured or die and may be elevated significantly during the acute infection period, but they also often vary significantly in the first six months of infection, typically stabilizing after six months.

The primary risk factors for HCV include IDUs, those with tattoos or body piercing done with unsterile instruments, anyone who had a blood transfusion before 1992, and hemodialysis patients. Vertical transmission during childbirth appears to be rare, although the risk is higher for mothers with HIV/HCV coinfection.[95–98] Transmission through sexual intercourse is less well understood with the risk generally considered to be very low. MSM appear to be at higher risk of sexual transmission of HCV, particularly if they have HIV infection.[66, 99] As this list indicates, the primary sources of HCV infection reflect exposure to blood or blood products, including transfusion of blood products that have not undergone viral inactivation and other activities that break the skin. The incidence rate of HCV infection seemed to have plateaued since 2003, although there was an increase in reported cases in 2007.[1] Whether this is the beginning of an upward trend remains to be seen.

The goal of treatment is to obtain a SVR, defined as HCV RNA being undetectable six months after completing treatment, which is considered a cure. The current treatment for HCV infection is composed of a combination of two medications: pegylated interferon alfa-2a or alfa-2b and ribavirin. Treatment regimens depend on the peginterferon formulation with peginterferon alfa-2b weight based at 1.5 μg/kg per week plus ribavirin 800 mg per day for patients less than 65 kg, 1,000 mg per day for patients 65 to 85 kg, 1,200 mg per day for patient 85 to 105 kg, or 1,400mg per day for patients more than 105kg. Peginterferon alfa-2a is a fixed dose at 180 μg per week plus ribavirin 1,000 mg per day for patients less than or equal to 75 kg or ribavirin 1,200 mg per day for patients more than 75 kg.

Interferon and ribavirin both have significant toxicities, including pancytopenia, as well as psychiatric side effects. Ribavirin is a significant teratogen when taken by men or women. The length of treatment is dependent on the HCV genotype and is the same for either regimen. Patients with genotype 1 or 4 should be treated for 48 weeks, while patients with genotype 2 or 3 can be treated for 24 weeks. Current therapy cure rates are 40 to 55 percent, with rates as high as 70 to 95 percent for patients with genotype 2 or 3 and as low as 20 percent in African Americans and individuals with HIV coinfections. Generally, patients with low viral load, nongenotype 1, who are non-African American, have a lower body mass index, are female, or are in the earlier stages of fibrosis tend to have higher chance of cure.[100] Recently, a specific genetic variation related to the IL28B gene has been identified that also appears to be significant in predicting patient's response to treatment.[101]

During the acute phase of infection (the first six months) patients have significantly higher chances of attaining an SVR with treatment. Approximately 20 percent of those with HCV infection, however, will clear the virus during this period without treatment; the majority will do so between 8 and 12 weeks following infection. Thus, it is reasonable to wait until 12 weeks after infection before starting treatment. Patients who are symptomatic or have jaundice are more likely to spontaneously clear the infection. Because of the significant advantages of early treatment, those with genotype 1 infection, high viral loads (more than 800,000 log10 IU/mL), or HIV infection should be considered for earlier treatment. SVR rates range between 80 and 98 percent when treatment is started in the acute phase, even for those with HIV coinfection. Optimal timing of onset and length and regimen of treatment during acute infection remains unclear. Evidence suggests that starting treatment 24 weeks after treatment leads to poorer rates of SVR, with rates much improved if treatment is initiated within 16 weeks of infection.[102] Current recommendations call for the use of peginterferon at standard dosing. Treatment should be continued for 24 weeks, although early termination at 12 weeks could be considered for patients with genotype 2 or 3.[102–122]

As with the other hepatitis viruses, special care must be taken when treating patients with comorbid conditions, especially HIV status. Approximately 25 percent of those infected with HIV have chronic HCV infections, and approximately 8 percent of those with chronic HCV infection also have HIV infection. Given this situation, patients with HCV infection should be evaluated for HIV infection and vice versa. Individuals with HCV/HIV appear to have more rapid progression of liver disease; therefore, early treatment should be strongly considered in this population. Treatment is with standard peginterferon and ribavirin dosing, and it should be

continued for 48 weeks among those with genotype 2 or 3. Before initiation of treatment of HCV, optimization of antiretroviral therapy should be achieved for those for which HIV treatment is indicated. It is still unclear whether treatment for HIV infections should be initiated for patients who do not meet current criteria for beginning highly active antiretroviral therapy (HAART). Although peginterferon therapy may lead to leukopenia, the percentage of white blood cells that are CD4 typical remains stable and development of opportunistic infections has not been observed.[123–139]

Individuals who are in methadone maintenance programs or actively using intravenous drugs often are excluded from treatment because of multiple concerns on both the part of providers and patients. Current evidence suggests that these populations have rates of SVR and side effects similar to the general population, but few providers are treating this cohort. Concerns about compliance, follow-up, psychiatric side effects, lack of insurance, lack of access to medical providers, and patient fears of judgment or mistreatment often keep patients from even presenting to specialty providers for treatment.[140–162]

Multiple new treatment strategies currently are being studied with several new drugs undergoing investigation. These new drugs, specifically targeted antiviral therapies for HCV (STAT-C), have been developed to target specific events in the life cycle of the HCV virus. Drugs are in development to target the HCV serine protease NS3-NS4A and polymerase NS5, virus entry into host cells, and virus assembly and release. Those that target protease NS3-NS4A are furthest in clinical development and are being investigated in combinations with current standard treatment with evidence of significant improvements in SVR rates. As new agents are developed, therapy for HCV likely will evolve to develop interferon-free combination therapies of STAT-C agents.[163–165]

Unfortunately, no vaccine currently is available for HCV. Preventive measures thus have focused on reduction of risk of transmission. Many of these preventive measures are relatively simple with adequate screening of blood products and infection control measure in health care settings, such as hemodialysis centers. Additional measures are needed to curb the incidence of new cases in other high-risk groups, particularly IDUs and MSM. Measures that reduce the risk of HIV transmission, including needle exchange programs, have been shown to be somewhat effective, but as transmission of HCV persists in this disenfranchised group, additional preventive measures are needed. Education about risk of transmission through sexual activity, particularly among MSM with HIV infection, also may reduce transmission in this risk group.

CONCLUSION

Multiple new treatment strategies currently are being studied to prevent and to treat HAV, HBV, and HCV. In particular, several new drugs that specifically target antiviral therapies for HCV are being investigated. Most target specific events in the life cycle of the HCV virus. Over the decades, progress has been made in the area of securing the safety of the blood supply. Vaccines, too, have helped dramatically reduce the incidence of HAV and HBV. Yet, chronic HBV and HCV remain prevalent worldwide and shifting immigration patterns are bringing more chronically

infected individuals into contact with noninfected individuals. Part of the difficulty in preventing viral hepatitis is the fact that these infections remain asymptomatic for years, even decades, often with patients unaware of their infection. An estimated 65 percent of individuals with chronic HBV and 75 percent of individuals with HCV are unaware of their diagnosis.[166, 167] The longer that treatment is not initiated, the greater the risk of serious liver disease—for example, serious complications such as cirrhosis or HCC. The likelihood of the need for a liver transplant increases as the disease severity increases.

Much more needs to be done to prevent all forms of viral hepatitis, and improved screening is needed to detect patients with chronic infections. Reducing or eliminating exposure to contaminated blood supplies, initiating needle exchange programs, and providing educational material to those at potentially higher risk all should be investigated. The good news is that HAV and HBV are eminently preventable; vaccination should be routine and outreach to those at high risk could do much to reduce the number of infected individuals in the population.

REFERENCES

1. Daniels D, Grytdal S, Wasley A. Surveillance for acute viral hepatitis: United States, 2007. *MMWR Surveill Summ*. 2009;May 22;58(3):1–27.
2. World Health Organization. Hepatits: fact sheets. http://www.who.int/topics/hepatitis/ factsheets/en/index.html. Accessed November 11, 2009.
3. Lavanchy D. Chronic viral hepatitis as a public health issue in the world. *Best Pract Res Clin Gastroenterol*. 2008;22(6):991–1008.
4. Kuehn BM. Silent epidemic of viral hepatitis may lead to boom in serious liver disease. *JAMA*. 2009;302(18):1949–1950, 1954.
5. Dienstag JL, McHutchison JG. American Gastroenterological Association technical review on the management of hepatitis C. *Gastroenterology*. 2006;130(1):231–264.
6. Armstrong GL, et al. The prevalence of hepatitis C virus infection in the United States, 1999–2002. *Ann Intern Med*. 2006;144(10):705–714.
7. Edlin BR. Five million Americans infected with the hepatitis C virus: a corrected estimate. *Hepatology*. 2005;42(4 Suppl 1):213A.
8. Wasley A, Alter MJ. Epidemiology of hepatitis C: geographic differences and temporal trends. *Semin Liver Dis*. 2000;20(1):1–16.
9. Kim WR. The burden of hepatitis C in the United States. *Hepatology*. 2002;36(5 Suppl 1):S30–S34.
10. Centers for Disease Control and Prevention. Hepatitis statistics page. http://www.cdc.gov/hepatitis/statistics.htm. Published June 2, 2009. Accessed October 9, 2009.
11. Perz JF, et al. The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol*. 2006;45(4):529–538.
12. Chen CJ, et al. Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *JAMA*. 2006;295(1):65–73.
13. Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C: The OBSVIRC, METAVIR, CLINIVIR, and DOSVIRC groups. *Lancet*. 1997;349(9055):825–832.
14. Hepatitis Foundation International. The ABCs of hepatitis. http://www.hepfi.org/living/ liv_abc.html. Accessed December 10, 2009.
15. Centers for Disease Control and Prevention. The ABCs of Hepatitis. http://www.cdc.gov/ hepatitis/Resources/Professionals/PDFs/ABCTable.pdf. Accessed October 2, 2009.

16. Advisory Committee on Immunization Practices. ACIP issues recommendations for hepatitis A prevention with active or passive immunization. *Am Fam Physician*. 1997;55(3):973–976.
17. Wasley A, Grytdal S, Gallagher K. Surveillance for acute viral hepatitis: United States, 2006. *MMWR Surveill Summ*. 2008;57(2):1–24.
18. Wasley A, Miller JT, Finelli L. Surveillance for acute viral hepatitis United States: 2005. *MMWR Surveill Summ*. 2007;56(3):1–24.
19. Steffen R. Changing travel-related global epidemiology of hepatitis A. *Am J Med*. 2005;118(Suppl 10A):46S–49S.
20. Poovorawan Y, Chatchatee P, Chongsrisawat V. Epidemiology and prophylaxis of viral hepatitis: a global perspective. *J Gastroenterol Hepatol*. 2002;17(Suppl): S155–S166.
21. Nelson KE. Global changes in the epidemiology of hepatitis A virus infections. *Clin Infect Dis*. 2006;42(8):1151–1152.
22. Hendrickx G, et al. Has the time come to control hepatitis A globally? matching prevention to the changing epidemiology. *J Viral Hepat*. 2008;15(Suppl 2):1–15.
23. Lavanchy D. Viral Hepatitis: global goals for vaccination. *J Clin Virol*. 2009;Epub ahead of print.
24. De Serres G, et al. Ineffectiveness of the current strategy to prevent hepatitis A in travelers. *J Travel Med*. 2002;9(1):10–16.
25. Keystone JS, Hershey JH. The underestimated risk of hepatitis A and hepatitis B: benefits of an accelerated vaccination schedule. *Int J Infect Dis*. 2008;12(1):3–11.
26. Prevention of hepatitis A through active or passive immunization: recommendations of the Advisory Committee on Immunization Practices. *MMWR Recomm Rep*. 1996;Dec 27;45(RR-15):1–30.
27. Advisory Committee on Immunization Practices. Update: prevention of hepatitis A after exposure to hepatitis A virus and in international travelers. *MMWR*. 2007;56(41):1080–1084.
28. Alter MJ. Epidemiology of hepatitis B in Europe and worldwide. *J Hepatol*. 2003;39(Suppl 1):S64–S69.
29. Margolis HS, Alter MJ, Hadler SC. Hepatitis B: evolving epidemiology and implications for control. *Semin Liver Dis*. 1991;11(2):84–92.
30. Maynard JE. Hepatitis B: global importance and need for control. *Vaccine*. 1990;8(Suppl 1–3):S18–20.
31. Lin KW, Kirchner JT. Hepatitis B. *Am Fam Physician*. 2004;69(1):75–82.
32. McMahon BJ, et al. Acute hepatitis B virus infection: relation of age to the clinical expression of disease and subsequent development of the carrier state. *J Infect Dis*. 1985;151(4):599–603.
33. Lok AS, McMahon BJ. Chronic hepatitis B. *Hepatology*. 2001;34(6):1225–1241.
34. Hatzakis A, Magiorkinis E, Haida C. HBV virological assessment. *J Hepatol*. 2006;44(1 Suppl):S71–S76.
35. Mast EE, et al. A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: recommendations of the Advisory Committee on Immunization Practices part 1: immunization of infants, children, and adolescents. *MMWR*. 2005;54(RR–16):1–31.
36. Mast EE, et al. A comprehensive immunization strategy to eliminate transmission of hepatitis B virus infection in the United States: recommendations of the Advisory Committee on Immunization Practices Part II: immunization of adults. *MMWR*. 2006;55(RR–16):1–33, quiz CE 1–4.
37. Alter MJ. Epidemiology of viral hepatitis and HIV co-infection. *J Hepatol*. 2006;44(1 Suppl):S6–S9.

38. Edlich RF, et al. Hepatitis B virus: a comprehensive strategy for eliminating transmission in the United States. *J Long Term Eff Med Implants*. 2003;13(2):117–125.
39. Immunization Practices Advisory Committee. Hepatitis B virus: a comprehensive strategy for eliminating transmission in the United States through universal childhood vaccination. *MMWR*. 1991;40(RR–13):1–25.
40. Namgyal P. Impact of hepatitis B immunization, Europe and worldwide. *J Hepatol*. 2003;39(Suppl 1):S77–S82.
41. Rodriguez-Mendez ML, et al. Prevalence, patterns, and course of past hepatitis B virus infection in intravenous drug users with HIV-1 infection. *Am J Gastroenterol*. 2000;95(5):1316–1322.
42. European Association For The Study Of The LEASL Clinical Practice Guidelines: management of chronic hepatitis B. *J Hepatol*. 2009;50(2):227–242.
43. Liaw YF, et al. Lamivudine for patients with chronic hepatitis B and advanced liver disease. *N Engl J Med*. 2004;351(15):1521–1531.
44. Chang TT, et al. A comparison of entecavir and lamivudine for HBeAg-positive chronic hepatitis B. *N Engl J Med*. 2006;354(10):1001–1010.
45. Janssen HL, et al. Pegylated interferon alfa-2b alone or in combination with lamivudine for HBeAg-positive chronic hepatitis B: a randomised trial. *Lancet*. 2005;365(9454):123–129.
46. Lai CL, et al. A one-year trial of lamivudine for chronic hepatitis B: Asia Hepatitis Lamivudine Study Group. *N Engl J Med*. 1998;339(2):61–68.
47. Lai CL, et al. Telbivudine versus lamivudine in patients with chronic hepatitis B. *N Engl J Med*. 2007;357(25):2576–2588.
48. Lau GK, et al. Peginterferon Alfa-2a, lamivudine, and the combination for HBeAg-positive chronic hepatitis B. *N Engl J Med*. 2005;352(26):2682–2695.
49. Marcellin P, et al. Adefovir dipivoxil for the treatment of hepatitis B e antigen-positive chronic hepatitis B. *N Engl J Med*. 2003;348(9):808–816.
50. Marcellin P, et al. Long-term efficacy and safety of adefovir dipivoxil for the treatment of hepatitis B e antigen-positive chronic hepatitis B. *Hepatology*. 2008;48(3):750–758.
51. Marcellin P, et al. Peginterferon alfa-2a alone, lamivudine alone, and the two in combination in patients with HBeAg-negative chronic hepatitis B. *N Engl J Med*. 2004;351(12):1206–1217.
52. Marcellin P, et al. Tenofovir disoproxil fumarate versus adefovir dipivoxil for chronic hepatitis B. *N Engl J Med*. 2008;359(23):2442–2455.
53. Hadziyannis SJ, et al. Adefovir dipivoxil for the treatment of hepatitis B e antigen-negative chronic hepatitis B. *N Engl J Med*. 2003;348(9):800–807.
54. Hadziyannis SJ, et al. Efficacy of long-term lamivudine monotherapy in patients with hepatitis B e antigen-negative chronic hepatitis B. *Hepatology*. 2000;32(4 Pt 1):847–851.
55. Hadziyannis SJ, et al. Long-term therapy with adefovir dipivoxil for HBeAg-negative chronic hepatitis B. *N Engl J Med*. 2005;352(26):2673–2681.
56. Lai CL, et al. Entecavir versus lamivudine for patients with HBeAg-negative chronic hepatitis B. *N Engl J Med*. 2006;354(10):1011–1020.
57. Tassopoulos NC, et al. Efficacy of lamivudine in patients with hepatitis B e antigen-negative/hepatitis B virus DNA-positive (precore mutant) chronic hepatitis B: Lamivudine Precore Mutant Study Group. *Hepatology*. 1999;29(3):889–896.
58. Perrillo R, et al. Adefovir dipivoxil added to ongoing lamivudine in chronic hepatitis B with YMDD mutant hepatitis B virus. *Gastroenterology*. 2004;126(1):81–90.

59. Perrillo RP, et al. Predictors of HBeAg loss after lamivudine treatment for chronic hepatitis B. *Hepatology*. 2002;36(1):186–194.
60. Perrillo RP, et al. A randomized, controlled trial of interferon alfa-2b alone and after prednisone withdrawal for the treatment of chronic hepatitis B: the Hepatitis Interventional Therapy Group. *N Engl J Med*. 1990;323(5):295–301.
61. Wong DK, et al. Effect of alpha-interferon treatment in patients with hepatitis B e antigen-positive chronic hepatitis B: a meta-analysis. *Ann Intern Med*. 1993; 119(4):312–323.
62. Bonino F, et al. Predicting response to peginterferon alpha-2a, lamivudine and the two combined for HBeAg-negative chronic hepatitis B. *Gut*. 2007;56(5):699–705.
63. Beasley RP. Hepatitis B virus. The major etiology of hepatocellular carcinoma. *Cancer*. 1988;61(10):1942–1956.
64. Bruix J, Llovet JM. Hepatitis B virus and hepatocellular carcinoma. J Hepatol. 2003;39 Suppl 1:S59–63.
65. Yang HI, et al. Hepatitis B e antigen and the risk of hepatocellular carcinoma. *N Engl J Med*. 2002;347(3):168–174.
66. Gambotti L, et al. Acute hepatitis C infection in HIV positive men who have sex with men in Paris, France, 2001–2004. *Euro Surveill*. 2005;10(5):115–117.
67. Thimme R, Spangenberg HC, Blum HE. Hepatitis B or hepatitis C and human immunodeficiency virus infection. *J Hepatol*. 2005;42(Suppl 1):S37–S44.
68. Feldman JG, et al. Heterosexual transmission of hepatitis C, hepatitis B, and HIV-1 in a sample of inner city women. *Sex Transm Dis*. 2000;27(6):338–342.
69. Di Martino V, et al. Influence of HIV infection on the response to interferon therapy and the long-term outcome of chronic hepatitis B. *Gastroenterology*. 2002;123(6):1812–1822.
70. Hoffmann CJ, Thio CL. Clinical implications of HIV and hepatitis B co-infection in Asia and Africa. *Lancet Infect Dis*. 2007;7(6):402–409.
71. Puoti M, et al. Natural history of chronic hepatitis B in co-infected patients. *J Hepatol*. 2006;44(1 Suppl):S65–S70.
72. Soriano V, et al. Care of HIV patients with chronic hepatitis B: updated recommendations from the HIV-Hepatitis B Virus International Panel. *AIDS*. 2008; 22(12):1399–1410.
73. Soriano V, et al. Care of patients with chronic hepatitis B and HIV co-infection: recommendations from an HIV-HBV International Panel. *AIDS*. 2005;19(3): 221–240.
74. Rockstroh JK, et al. European AIDS Clinical Society guidelines for the clinical management and treatment of chronic hepatitis B and C coinfection in HIV-infected adults. *HIV Med*. 2008;9(2):82–88.
75. Rockstroh JK. Influence of viral hepatitis on HIV infection. *J Hepatol*. 2006;44(1 Suppl):S25–S27.
76. Sulkowski MS. Viral hepatitis and HIV coinfection. *J Hepatol*. 2008;48(2):353–367.
77. Chu CJ, Lee SD. Hepatitis B virus/hepatitis C virus coinfection: epidemiology, clinical features, viral interactions and treatment. *J Gastroenterol Hepatol*. 2008;23(4):512–520.
78. Liu CJ, et al. Ribavirin and interferon is effective for hepatitis C virus clearance in hepatitis B and C dually infected patients. *Hepatology*. 2003;37(3):568–576.
79. Terrault NA, Jacobson IM. Treating chronic hepatitis B infection in patients who are pregnant or are undergoing immunosuppressive chemotherapy. *Semin Liver Dis*. 2007;27 (Suppl 1):S18–S24.
80. van Zonneveld M, et al. Lamivudine treatment during pregnancy to prevent perinatal transmission of hepatitis B virus infection. *J Viral Hepat*. 2003;10(4):294–297.

81. ter Borg MJ, et al. Exacerbation of chronic hepatitis B infection after delivery. *J Viral Hepat*. 2008;15(1):37–41.
82. Simmonds P. Viral heterogeneity of the hepatitis C virus. *J Hepatol*. 1999; 31(Suppl 1):54–60.
83. Nguyen MH, Keeffe EB. Prevalence and treatment of hepatitis C virus genotypes 4, 5, and 6. *Clin Gastroenterol Hepatol*. 2005;3(10 Suppl 2):S97–S101.
84. Magiorkinis G, et al. The global spread of hepatitis C virus 1a and 1b: a phylodynamic and phylogeographic analysis. *PLoS Med*. 2009;(12):e1000198.
85. Centers for Disease Control and Prevention. Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. *MMWR Recomm Rep*. 1998;47(RR-19):1–39.
86. Armstrong GL, Alter MJ, McQuillan GM, Margolis HS. The past incidence of hepatitis C virus infection: implications for the future burden of chronic liver disease in the United States. *Hepatology*. 2000;31(3):777–782.
87. Davis GL, et al. Projecting future complications of chronic hepatitis C in the United States. *Liver Transpl*. 2003;9(4):331–338.
88. Deuffic-Burban S, et al. Estimating the future health burden of chronic hepatitis C and human immunodeficiency virus infections in the United States. *J Viral Hepat*. 2007;14(2):107–115.
89. Sulkowski MS, Thomas DL. Hepatitis C in the HIV-infected person. *Ann Intern Med*. 2003;138(3):197–207.
90. Garfein RS, et al. Viral infections in short-term injection drug users: the prevalence of the hepatitis C, hepatitis B, human immunodeficiency, and human T-lymphotropic viruses. *Am J Public Health*. 1996;86(5):655–661.
91. Lorvick J, et al. Prevalence and duration of hepatitis C among injection drug users in San Francisco, California. *Am J Public Health*. 2001;91(1):46–47.
92. Garfein RS, et al. Prevalence and incidence of hepatitis C virus infection among young adult injection drug users. *J Acquir Immune Defic Syndr Hum Retrovirol*. 1998;18(Suppl 1):S11–S19.
93. Tseng FC, et al. Seroprevalence of hepatitis C virus and hepatitis B virus among San Francisco injection drug users: 1998–2000. *Hepatology*. 2007;46(3):666–671.
94. Sarrazin C, et al. Detection of residual hepatitis C virus RNA by transcription-mediated amplification in patients with complete virologic response according to polymerase chain reaction-based assays. *Hepatology*. 2000;32(4 Pt 1): 818–823.
95. Ohto H, et al. Transmission of hepatitis C virus from mothers to infants: the Vertical Transmission of Hepatitis C Virus Collaborative Study Group. *N Engl J Med*. 1994; 330(11):744–750.
96. Wejstal R, Hermodsson S, Iwarson S, Norkrans G. Mother to infant transmission of hepatitis C virus infection. *J Med Virol*. 1990;30(3):178–180.
97. Wejstal R, et al. Mother-to-infant transmission of hepatitis C virus. *Ann Intern Med*. 1992;117(11):887–890.
98. Manzini P, et al. Human immunodeficiency virus infection as risk factor for mother-to-child hepatitis C virus transmission: persistence of anti-hepatitis C virus in children is associated with the mother's anti-hepatitis C virus immunoblotting pattern. *Hepatology*. 1995;21(2):328–332.
99. Alary M, et al. Lack of evidence of sexual transmission of hepatitis C virus in a prospective cohort study of men who have sex with men. *Am J Public Health*. 2005;95(3):502–505.
100. Ghany MG, et al. Diagnosis, management, and treatment of hepatitis C: an update. *Hepatology*. 2009;49(4):1335–1374.

101. Ge D, et al. Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance. *Nature*. 2009;461(7262):399–401.
102. Kamal SM, et al. Peginterferon alfa-2b therapy in acute hepatitis C: impact of onset of therapy on sustained virologic response. *Gastroenterology*. 2006;130(3):632–638.
103. Vogel W, et al. High-dose interferon-alpha2b treatment prevents chronicity in acute hepatitis C: a pilot study. *Dig Dis Sci*. 1996;41(12 Suppl):S81–S85.
104. Jaeckel E, et al. Treatment of acute hepatitis C with interferon alfa-2b. *N Engl J Med*. 2001;345(20):1452–1457.
105. Rocca P, et al. Early treatment of acute hepatitis C with interferon alpha-2b or interferon alpha-2b plus ribavirin: study of sixteen patients. *Gastroenterol Clin Biol*. 2003;27(3 Pt 1):294–299.
106. Gerlach JT, et al. Acute hepatitis C: high rate of both spontaneous and treatment-induced viral clearance. *Gastroenterology*. 2003;125(1):80–88.
107. Nomura H, et al. Short-term interferon-alfa therapy for acute hepatitis C: a randomized controlled trial. *Hepatology*. 2004;39(5):1213–1219.
108. Pimstone NR, et al. "Wait-and-see": an alternative approach to managing acute hepatitis C with high-dose interferon-alpha monotherapy. *Ann Intern Med*. 2004;141(6):W91–W92.
109. Delwaide J, et al. Treatment of acute hepatitis C with interferon alpha-2b: early initiation of treatment is the most effective predictive factor of sustained viral response. *Aliment Pharmacol Ther*. 2004;20(1):15–22.
110. Santantonio T, et al. Efficacy of a 24-week course of PEG-interferon alpha-2b monotherapy in patients with acute hepatitis C after failure of spontaneous clearance. *J Hepatol*. 2005;42(3):329–333.
111. Broers B, et al. Barriers to interferon-alpha therapy are higher in intravenous drug users than in other patients with acute hepatitis C. *J Hepatol*. 2005;42(3):323–328.
112. De Rosa FG, et al. Twelve-week treatment of acute hepatitis C virus with pegylated interferon-alpha-2b in injection drug users. *Clin Infect Dis*. 2007;45(5):583–588.
113. De Rosa FG, et al. Dose-dependent and genotype-independent sustained virological response of a 12 week pegylated interferon alpha-2b treatment for acute hepatitis C. *J Antimicrob Chemother*. 2006;57(2):360–363.
114. Corey KE, et al. Outcomes and treatment of acute hepatitis C virus infection in a United States population. *Clin Gastroenterol Hepatol*. 2006;4(10):1278–1282.
115. Wiegand J, et al. Early monotherapy with pegylated interferon alpha-2b for acute hepatitis C infection: the HEP-NET acute-HCV-II study. *Hepatology*. 2006;43(2):250–256.
116. Calleri G, et al. A short course of pegylated interferon-alpha in acute HCV hepatitis. *J Viral Hepat*. 2007;14(2):116–121.
117. Kamal SM, et al. Duration of peginterferon therapy in acute hepatitis C: a randomized trial. *Hepatology*. 2006;43(5):923–931.
118. Vogel M, et al. Treatment of acute hepatitis C infection in HIV-infected patients: a retrospective analysis of eleven cases. *J Viral Hepat*. 2005;12(2):207–211.
119. Vogel M, et al. Pegylated interferon-alpha for the treatment of sexually transmitted acute hepatitis C in HIV-infected individuals. *Antivir Ther*. 2006;11(8):1097–1101.
120. Dominguez S, et al. Efficacy of early treatment of acute hepatitis C infection with pegylated interferon and ribavirin in HIV-infected patients. *AIDS*. 2006;20(8):1157–1161.
121. Serpaggi J, et al. Sexually transmitted acute infection with a clustered genotype 4 hepatitis C virus in HIV-1-infected men and inefficacy of early antiviral therapy. *AIDS*. 2006;20(2):233–240.

122. Matthews GV, et al. Characteristics and treatment outcomes among HIV-infected individuals in the Australian Trial in Acute Hepatitis C. *Clin Infect Dis*. 2009; 48(5):650–658.
123. Sherman KE, et al. Hepatitis C virus prevalence among patients infected with human immunodeficiency virus: a cross-sectional analysis of the US Adult AIDS Clinical Trials Group. *Clin Infect Dis*. 2002;34(6):831–837.
124. Bica I, et al. Increasing mortality due to end-stage liver disease in patients with human immunodeficiency virus infection. *Clin Infect Dis*. 2001;32(3):492–497.
125. Bonacini M, Lin HJ, Hollinger FB. Effect of coexisting HIV-1 infection on the diagnosis and evaluation of hepatitis C virus. *J Acquir Immune Defic Syndr*. 2001;26(4):340–344.
126. Graham CS, et al. Influence of human immunodeficiency virus infection on the course of hepatitis C virus infection: a meta-analysis. *Clin Infect Dis*. 2001;33(4):562–569.
127. Sulkowski MS. The HIV-coinfected patient: managing viral hepatitis. *J Acquir Immune Defic Syndr*. 2007;45(Suppl 2):S36–S37.
128. Labarga P, et al. Hepatotoxicity of antiretroviral drugs is reduced after successful treatment of chronic hepatitis C in HIV-infected patients. *J Infect Dis*. 2007;196(5):670–676.
129. Laguno M, et al. Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for treatment of HIV/HCV co-infected patients. *AIDS*. 2004;18(13):F27–36.
130. Chung RT, et al. Peginterferon Alfa-2a plus ribavirin versus interferon alfa-2a plus ribavirin for chronic hepatitis C in HIV-coinfected persons. *N Engl J Med*. 2004; 351(5):451–459.
131. Torriani FJ, et al. Peginterferon Alfa-2a plus ribavirin for chronic hepatitis C virus infection in HIV-infected patients. *N Engl J Med*. 2004;351(5):438–450.
132. Carrat F, et al. Pegylated interferon alfa-2b vs standard interferon alfa-2b, plus ribavirin, for chronic hepatitis C in HIV-infected patients: a randomized controlled trial. *JAMA*. 2004;292(23):2839–2848.
133. Soriano V, et al. Relapses of chronic hepatitis C in HIV-infected patients who responded to interferon therapy: Hepatitis/HIV Spanish Study Group. *AIDS*. 1997;11(3):400–401.
134. Alvarez D, et al. Zidovudine use but not weight-based ribavirin dosing impacts anaemia during HCV treatment in HIV-infected persons. *J Viral Hepat*. 2006;13(10):683–689.
135. Hoggard PG, et al. Effects of drugs on 2',3'-dideoxy-2',3'-didehydrothymidine phosphorylation in vitro. *Antimicrob Agents Chemother*. 1997;41(6):1231–1236.
136. Lafeuillade A, Hittinger G, Chadapaud S. Increased mitochondrial toxicity with ribavirin in HIV/HCV coinfection. *Lancet*. 2001;357(9252):280–281.
137. Landau A, et al. Lack of interference between ribavirin and nucleosidic analogues in HIV/HCV co-infected individuals undergoing concomitant antiretroviral and anti-HCV combination therapy. *AIDS*. 2000;14(12):1857–1858.
138. Valerio L, et al. Baseline CD4 cell count and outcome of pegylated interferon plus ribavirin therapy in HIV/hepatitis C virus-coinfected patients. *J Acquir Immune Defic Syndr*. 2008;47(1):50–55.
139. Opravil M, et al. Effect of baseline CD4 cell count on the efficacy and safety of peginterferon Alfa-2a (40KD) plus ribavirin in patients with HIV/hepatitis C virus coinfection. *J Acquir Immune Defic Syndr*. 2008;47(1):36–49.
140. Edlin BR. Prevention and treatment of hepatitis C in injection drug users. *Hepatology*. 2002;36(5 Suppl 1):S210–S219.

141. Edlin BR, Carden MR, Ferrando SJ. Managing hepatitis C in users of illicit drugs. *Curr Hepatitis Rep*. 2007;6(2):60–67.
142. Edlin BR, et al. Overcoming barriers to prevention, care, and treatment of hepatitis C in illicit drug users. *Clin Infect Dis*. 2005;40(Suppl 5):S276–S285.
143. Edlin BR, et al. Is it justifiable to withhold treatment for hepatitis C from illicit-drug users? *N Engl J Med*. 2001;345(3):211–214.
144. Sylvestre DL. Treating hepatitis C in methadone maintenance patients: an interim analysis. *Drug Alcohol Depend*. 2002;67(2):117–123.
145. Sylvestre DL, Clements BJ. Adherence to hepatitis C treatment in recovering heroin users maintained on methadone. *Eur J Gastroenterol Hepatol*. 2007;19(9):741–747.
146. Sylvestre DL, et al. The impact of barriers to hepatitis C virus treatment in recovering heroin users maintained on methadone. *J Subst Abuse Treat*. 2005;29(3):159–165.
147. Schaefer M, et al. Hepatitis C treatment in "difficult-to-treat" psychiatric patients with pegylated interferon-alpha and ribavirin: response and psychiatric side effects. *Hepatology*. 2007;46(4):991–998.
148. Schaefer M, Mauss S. Hepatitis C treatment in patients with drug addiction: clinical management of interferon-alpha-associated psychiatric side effects. *Curr Drug Abuse Rev*. 2008;1(2):177–187.
149. Schaefer M, et al. Adherence and mental side effects during hepatitis C treatment with interferon alfa and ribavirin in psychiatric risk groups. *Hepatology*. 2003;37(2):443–451.
150. Backmund M, Meyer K, Edlin BR. Infrequent reinfection after successful treatment for hepatitis C virus infection in injection drug users. *Clin Infect Dis*. 2004;39(10):1540–1543.
151. Backmund M, Meyer K, von Zielonka M, Eichenlaub D. Treatment of hepatitis C infection in injection drug users. *Hepatology*. 2001;34(1):188–193.
152. Matthews AM, et al. Relapse to opioid use after treatment of chronic hepatitis C with pegylated interferon and ribavirin. *Am J Psychiatry*. 2006;163(8):1342–1347, quiz 479.
153. Matthews G, Kronborg IJ, Dore GJ. Treatment for hepatitis C virus infection among current injection drug users in Australia. *Clin Infect Dis*. 2005;40(Suppl 5):S325–S329.
154. Stein MD, Friedmann P. Need for medical and psychosocial services among injection drug users: a comparative study of needle exchange and methadone maintenance. *Am J Addict*. 2002;11(4):262–270.
155. Stein MD, Maksad J, Clarke J. Hepatitis C disease among injection drug users: knowledge, perceived risk and willingness to receive treatment. *Drug Alcohol Depend*. 2001;61(3):211–215.
156. Chainuvati S, et al. Comparison of hepatitis C treatment patterns in patients with and without psychiatric and/or substance use disorders. *J Viral Hepat*. 2006;13(4):235–241.
157. Litwin AH, et al. Successful treatment of chronic hepatitis C with pegylated interferon in combination with ribavirin in a methadone maintenance treatment program. *J Subst Abuse Treat*. 2009;37(1):32–40.
158. Grebely J, et al. Treatment uptake and outcomes among current and former injection drug users receiving directly observed therapy within a multidisciplinary group model for the treatment of hepatitis C virus infection. *Int J Drug Policy*. 2007;18(5):437–443.
159. Grebely J, et al. Barriers associated with the treatment of hepatitis C virus infection among illicit drug users. *Drug Alcohol Depend*. 2008;93(1–2):141–147.

160. Grebely J, et al. Directly observed therapy for the treatment of hepatitis C virus infection in current and former injection drug users. *J Gastroenterol Hepatol*. 2007;22(9):1519–1525.
161. Seal KH, et al. Hepatitis C treatment candidacy and outcomes among 4318 US veterans with chronic hepatitis C virus infection: does a history of injection drug use matter? *J Clin Gastroenterol*. 2007;41(2):199–205.
162. van Thiel DH, Anantharaju A, Creech S. Response to treatment of hepatitis C in individuals with a recent history of intravenous drug abuse. *Am J Gastroenterol*. 2003;98(10):2281–2288.
163. Pereira AA, Jacobson IM. New and experimental therapies for HCV. *Nat Rev Gastroenterol Hepatol*. 2009;6(7):403–411.
164. McHutchison JG, et al. Telaprevir with peginterferon and ribavirin for chronic HCV genotype 1 infection. *N Engl J Med*. 2009;360(18):1827–1838.
165. Hezode C, et al. Telaprevir and peginterferon with or without ribavirin for chronic HCV infection. *N Engl J Med*. 2009;360(18):1839–1850.
166. Hagan H, et al. Self-reported hepatitis C virus antibody status and risk behavior in young injectors. *Public Health Rep*. 2006;121(6):710–719.
167. Lin SY, Chang ET, So SK. Why we should routinely screen Asian American adults for hepatitis B: a cross-sectional study of Asians in California. *Hepatology*. 2007;46(4):1034–1040.

CHAPTER 15

Still a Hidden Epidemic: The Challenges of Sexually Transmitted Disease Control

Elizabeth A. Torrone, MSPH, PhD, and Thomas A. Peterman, MD, PhD, MSc

In 1919, a New Jersey billboard asked citizens to "Fight the enemies at home! Venereal Disease!"[1] In the 1940s, Hank Williams and Woody Guthrie sang "hillbilly operas" about the risks of venereal disease (VD) as part of a Columbia University radio project.[2] In 1973, Walt Disney produced an animated film "VD: Attack Plan!" targeting adolescents.[3] In 2009, a sexually transmitted disease (STD) awareness campaign "GYT: Get Yourself Tested" was launched with television commercials airing on MTV and an Internet site with interactive games and links to search for testing sites.[4] Despite such media campaigns, STDs remain a "hidden epidemic."[5] STD education and prevention messages often are censored in classrooms, in media, and in research. The asymptomatic nature of many STDs helps keep the epidemic hidden, making prevention and control a continue challenge in the 21st century.

More than 35 pathogens currently are classified as causing STDs, including bacteria, viruses, fungi, and parasites.[6] Most of these organisms are primarily transmitted by sexual contact; however, some are transmitted sexually only occasionally and many are transmitted congenitally. Manifestations depend on the organism and may include ulcers, genital inflammation and discharge, rash, arthritis, enteric, and neurologic syndromes. Yet, most often, infections are asymptomatic. Even with symptoms, they may resolve without treatment and often without clearing the infection. Asymptomatic infections are particularly challenging for STD control efforts because they can remain infectious, which means that they can be spread to others, and can lead to chronic disease, such as infertility and cancer.

This chapter presents a brief overview of STD and then focuses on syphilis to illustrate the challenges in STD prevention and treatment.

BURDEN OF DISEASE

Worldwide, more than 448 million people acquire a curable STD each year (see table 15.1).[7] When viral infections such as human immunodeficiency virus (HIV) are included, the disease burden grows. An estimated 33 million people are

Table 15.1
Estimates of Incidence and Prevalence of Common STD

	Global (millions)		United States (millions)	
Pathogen	Annual incidence	Prevalence	Annual incidence	Prevalence
Trichomonas vaginalis	248.5[7]	—	7.4[9]	—
Chlamydia trachomatis	102.2[7]	—	2.8[9]	—
Neisseria gonorrhoeae	87.7[7]	—	0.7[9]	—
Treponema pallidum	10.6[7]	—	0.04[9]	—
Herpes simplex virus type 2	23.6[7]	536[7]	1.6[9]	45[9]
Hepatitis B	4.6[106]	368[106]	0.04[107]	0.8–1.4[107]
HIV	2.7[108]	33[108]	0.06[109]	1.1[110]

Source: Authors' compilation based on sources as noted.
Note: HIV = human immunodeficiency virus; STD = sexually transmitted disease.

living with HIV and 2.7 million new HIV infections are acquired each year. In the United States, almost 11 million curable infections occur each year. To highlight the burden of STDs, collectively they account for 86 percent of infections reported to the U.S. Centers for Disease Control and Prevention (CDC); the most frequently reported infections include *Chlamydia trachomatis* (chlamydia), *Neisseria gonorrhoeae* (gonorrhea), and HIV.[8] The most prevalent STDs, herpes simplex virus type 2 (HSV-2) and human papillomavirus (HPV), are not routinely reported. It is estimated that each year 1.6 million new HSV-2 infections occur in North America more than 6 million new HPV infections occur in the United States.[9]

The burden of STD is not evenly distributed across the population; it is clustered in marginalized populations. In the United States, rates are highest in young adults and in minority populations. Although young adults make up only one-fourth of the sexually active population, they acquire almost half of new sexually transmitted infections.[9] One in four young women ages 14 to 19, for example, have a sexually transmitted infection (most have HPV) and the prevalence is almost double in young black women.[10] The rate of reported chlamydia among blacks is more than eight times the rate among non-Hispanic whites (1,398.7 and 162.3 per 100,000, respectively). For gonorrhea, the reported rate among blacks is 19 times the rate among non-Hispanic whites (662.9 and 34.7 per 100,000, respectively). High rates also are seen in Hispanic and American Indian and Alaska Native populations.[11] Reasons for disparities are complex, difficult to identify, and multifactorial, and involve both social and structural determinants.[12]

Deaths due to STDs are decreasing primarily because of expanded treatment coverage for HIV. Still, AIDS accounts for more than 2 million deaths per year worldwide,[13] and HIV disease is the leading cause of death for black females ages 25 to 34 in the United States.[14] Approximately 270,000 women, the overwhelming majority of whom live in developing countries, died from HPV-related cervical cancer in 2002.[15] Although syphilis-related deaths have decreased significantly after the discovery of penicillin, an estimated 500,000 fetal and neonatal deaths are attributed to untreated maternal early syphilis worldwide.[16] Untreated maternal gonococcal and chlamydial infections result in infant blindness in as many as 4,000 newborns

yearly, again, primarily in the developing countries around the world.[17] Morbidities from STD also result in substantial treatment costs. The total direct cost to treat STD (not including HIV) in the United States, for example, was estimated to be $7.5 billion in 1994 dollars.[5] Lifetime direct medical costs for a single case of HIV are estimated to be almost $170,000.[18] The economic burden is even higher when indirect costs, such as productivity-related losses, are included. Worldwide, STD caused a loss of more than 11 million Disability-Adjusted Life Years in 2004.[19]

GENERAL FRAMEWORK FOR STD CONTROL

STD control programs vary by location, but generally include surveillance, clinical services (including case diagnosis and treatment), and prevention services. Such programs primarily are designed to reduce disease incidence and to reduce complications from existing infections. In the United States, STD control efforts involve the public and private sectors at local, state, and national levels. Interventions occur at the individual level (for example, prevention counseling in STD clinics), community level (for example, social marketing campaigns), and structural or policy level (for example, legislation to permit partner-delivered treatment). STD control programs primarily intervene on duration of infectiousness through case finding, treatment, and partner notification. These interventions prevent transmission from infectious persons (primary prevention) and prevent more severe health consequences for the infected (secondary prevention).

Few STD prevention strategies have been evaluated rigorously in randomized controlled trials. Some evaluations, however, have used surveillance systems to document changes in disease rates that suggest these control efforts work. For example, a national gonorrhea control program was initiated in the United States in the mid-1970s, which combined widespread screening, aggressive partner notification, and treatment conducted by local health department staff. In 1976 alone, the 1 million reported cases of gonorrhea resulted in 319,029 postdiagnosis interviews, 82,277 new cases found and treated through partner interviews, and another 98,760 partners were prophylactically treated.[20] After initiation of the program, U.S. gonorrhea rates decreased from 467.7 to 121.8 per 100,000 population between 1975 and 1996.[11]

A framework for STD control can be built using the formula for the reproductive rate of disease: $R0 = \beta cd$. R0 is defined as the average number of secondary cases an infected case will cause in a population without immunity, β is the probability of transmission, c is the average rate of new partnerships, and d is the duration of infectiousness.[21] Intervening on any of the three variables will modify R0 (see table 15.2). When R0 is below 1, programs are successfully removing more cases from the population than are being created. Although the case finding-partner notification framework has been effective in controlling bacterial infections, control programs for noncurable STDs may need to rely more heavily on reducing β (transmission probability) or c (number of contacts) through behavior change interventions.

CHALLENGES OF STD CONTROL PROGRAMS

Programs often are grouped together as "STD control," however, because of heterogeneity in pathogen type, disease manifestation, disease incidence, and

Table 15.2
Selected Examples of STD Control Program Activities by Reproductive Rate Determinant and Intervention Level

	Intervention example		
Determinant	Individual level	Community level	Policy level
Reduce probability of pathogen transmission (β)	Postexposure prophylaxis after exposure to Hepatitis B	Internet outreach on sex-seeking Web sites promoting condom use	Condom availability in correctional facilities
Minimize the number of susceptible partners (c)	Prevention counseling to reduce concordant sexual partnerships	School-based education promoting sexual activity delay	HPV vaccine recommendation for 11- to 12-year-old girls
Shorten the duration of infectiousness (d)	Notification and treatment of exposed partners	Awareness campaigns for recognizing symptoms of STD	Recommendation for routine STD screening at health care visits

Note: STD = sexually transmitted disease.

epidemic phase, diagnostic testing, and treatment availability, it might be prudent to have an individualized control program specific to the disease as well as to the population currently driving the epidemic. Although the overarching goal of reducing incidence and minimizing negative health consequences is universal, strategies vary. For example, chlamydia is usually asymptomatic and, if untreated, can lead to pelvic inflammatory disease, ectopic pregnancy, and infertility. It is easily diagnosed by screening tests and can be treated with a single dose of an inexpensive antibiotic. Widespread screening programs effectively have been used as a control strategy. On the other hand, for a viral STD such as genital herpes, treatment can reduce symptoms and the likelihood of transmission, but it does not cure the infection. Consequently, control measures should focus on ensuring treatment coverage, creating education programs to prevent secondary transmission, and encouraging vaccine development. For those STDs for which vaccines are already available (that is, HPV and hepatitis B), programmatic effort should focus on expanding immunization coverage.

One feature common to all STDs is stigma based on sexual transmission, a major barrier for both prevention and control. STDs often are considered markers of immoral behavior. In the 1930s, patients with STD frequently were turned away from hospitals and private providers as not deserving of care.[22] Mass media messages contribute to the moralistic approach. In 1934, Thomas Parran, a commissioner of public health in New York, was censored for saying "syphilis" on a radio broadcast.[23] Seventy years later, television stations in Los Angeles refused to show antisyphilis public service announcements created by the county department of health, citing them as "inappropriate."[24] Censoring even occurs within the STD research world. In an article published in 1988, the *American Journal of Epidemiology* censored a questionnaire measuring sexual risk behaviors in men who have sex with men (MSM) with the footnote: "Some vernacular terms have been

deleted per the editor's suggestion." Deleted terms included "oral sex" and "screwing."[25] This kind of stigma can create barriers for both public health practitioners and patients.[26]

As a comprehensive review of the specific challenges to control programs for all STD is not feasible in a single chapter, we will use syphilis as a case study. Although any STD could have been used, we selected syphilis for its interesting history that includes being the stimulus for the first major STD campaign in the United States. We review the current epidemiology of syphilis and discuss challenges faced in five of the current key aspects to syphilis control: (1) screening, (2) partner notification, (3) provider education, (4) public awareness, and (5) surveillance. The chapter concludes by highlighting additional tools for prevention and control. We limit the discussion to control efforts in the developed world, primarily the United States. The majority of the issues covered are also applicable to control programs in developing countries; however, challenges often are amplified by poor infrastructure and low resources.

CASE STUDY OF SYPHILIS

Transmission and Clinical Manifestations

Syphilis is a bacterial infection caused by the spirochete *Treponema pallidum*. The pathogen is hypothesized to have mutated from a nonvenereal bacteria brought back to Spain in the late 15th century by Columbus' expedition to the New World.[27] In 1495, a syndrome marked by pustules and intense pain appeared in French mercenaries. The disease spread quickly across Europe and became known as the "great pox," marking the first recorded epidemic of the disease. By the 16th century it was generally accepted that the disease was spread sexually. We now know that the pathogen is spread by contact (usually sexual) with a mucosal or cutaneous lesion, not by the "passions of immoral sex" as conjectured by a French physician in 1527.[28] The pathogen also can be passed vertically, either in utero or at time of delivery. Sexual transmission probability is estimated to be 30 percent during a single sexual act with an infectious partner.[29] Congenital transmission probability varies by duration of in utero exposure and stage of mother's infection.[30]

The general natural history of the disease is well-known, due in part to a large preantibiotic prospective cohort conducted in Oslo, Norway, which followed untreated patients for almost 50 years.[31] The course of the disease is complicated and can span decades with multiple manifestations. Left untreated, the disease will progress through four stages: primary, secondary, latent, and tertiary. Primary syphilis is marked by a typically painless lesion appearing about three weeks after exposure. The lesion forms at the site of infection (usually genital, anal, or oral). The lesion is often missed, especially when it occurs on the anus or intravaginally, or is mistaken for hemorrhoids or a pimple. Even without treatment, the lesion will heal in a few weeks. Shortly after (or during) the primary stage, the pathogen will spread throughout the body leading to secondary syphilis. Systemic illness, such as sore throat, alopecia, fever, and rash may occur. Similar to primary syphilis, symptoms may be missed or ignored by the patient or be misdiagnosed, and symptoms will resolve without treatment. Latent ("hidden") stage syphilis is detectable only by

serologic testing, but it may lapse or relapse into secondary syphilis in the year following infection ("early latent" period). After this time, the stage is considered "late latent." The fourth stage, tertiary syphilis, may occur 10 to 20 years or more after an untreated infection and includes numerous clinical syndromes, such as skeletal and cardiovascular complications. Neurosyphilis (infection of the central nervous system) can occur during any stage of syphilis.

Syphilis is transmitted from lesions, which may be unrecognized by the individual, and only occurs during primary and secondary stages. Additionally, *T. pallidum* can cross the placenta and can be transmitted vertically during any stage of the mother's infection, but transmission probability is highest during early stages.[32]

Diagnosis

T. pallidum cannot be cultured or stained easily, but using darkfield microscopy, fluid from the surface of a lesion can be sampled and examined for presence of the bacterium.[33] Trained darkfield microscopists, however, rarely are available in clinic settings. During primary syphilis, serologic tests may be negative (test sensitivities range from 78 to 86 percent during this stage). Therefore, in the absence of darkfield microscopy, disease control efforts require presumptive treatment based on signs and symptoms. During secondary syphilis, serologic tests always are positive. Syphilis screening for latent infections traditionally starts with a nontreponemal test, such as the rapid plasma reagin (RPR) test or venereal disease research laboratory (VDRL) test, which detect antibodies to cardiolipin and are not specific for treponemal infection. Positive nontreponemal tests occur in 1 to 2 percent of individuals who have never had syphilis, so people with positive nontreponemal screening tests are then tested using a treponemal test. Treponemal tests detect antibodies specific to *T. pallidum*; a positive test indicates that treponemal infection has occurred at some point in the past. Individuals who test positive on both tests are considered to have untreated syphilis, unless they were previously treated. Even after treatment, treponemal tests usually remain positive, but nontreponemal tests often revert to negative. Titers on quantitative nontreponemal tests are used to monitor response to therapy and to detect reinfections. Health departments maintain registries of past positive syphilis test results and have staff to aid in interpretation.

Treatment

Original treatments for syphilis, such as mercury and arsenic, were toxic and mostly ineffective. In 1927, Julius Wagner-Jauregg won the Nobel Prize for his discovery that syphilis could be treated with malaria-induced fever (the malaria could be subsequently cured with quinine). Although the treatment appeared effective (a study from the Mayo Clinic reported "complete remission" for 50 percent of patients at a three-year follow-up), no randomized control trials were done and the treatment period was lengthy. When penicillin was introduced in the 1940s as an effective cure for syphilis, malaria treatments were abandoned.[34]

Unlike other bacterial STD pathogens, such as *Neisseria gonorrhoeae* that have developed resistance to a wide variety of drugs, *T. pallidum* remains sensitive to penicillin.[35] Current treatment guidelines are specific to stage of disease as well as for special populations (for example, patients coinfected with HIV).[36] The

recommended regimen for primary, secondary, and early latent stages is a single intramuscular injection of benzathine penicillin; three weekly injections are recommended for late latent syphilis. Neurosyphilis requires 10 to 14 days of parenteral penicillin treatment. Few options are available for patients with a penicillin allergy, but doxycycline, tetracycline, and ceftriaxone may be effective for treating primary and secondary syphilis infections and ceftriaxone may be effective for treating neurosyphilis. Currently, no alternative to penicillin is available for treating congenital syphilis or for treating pregnant women. As benzathine penicillin must be administered intramuscularly, a single-dose oral alternative is more attractive. Although a number of clinical trials have showed the effectiveness of oral treatments, in particular azithromycin,[37, 38] reports of azithromycin-resistance are concerning.[39]

Incidence

In the first half of the 20th century, syphilis was a common infection in the United States. Autopsy studies estimated that 5 to 10 percent of the population was infected. Similar to current disparities, studies estimated prevalence up to 25 percent, especially among lower socioeconomic groups.[32] Syphilis was one of the most common causes of dementia and, in the 1920s, and one in five patients in mental institutions were diagnosed with tertiary syphilis.[40] Recognizing the public health significance of the epidemic, Surgeon General Thomas Parran pressured Congress to pass the National Venereal Disease Control Act in 1938. The control program focused on routine testing, education, and stigma reduction and, combined with the introduction of penicillin, rates in the United States decreased dramatically (see figure 15.1).

In the following decades, rates of primary and secondary syphilis increased slightly about every 10 years, with peaks in the epidemic clustered by age, race, region, and gender.[41] For example, between 1985 and 1990, primary and

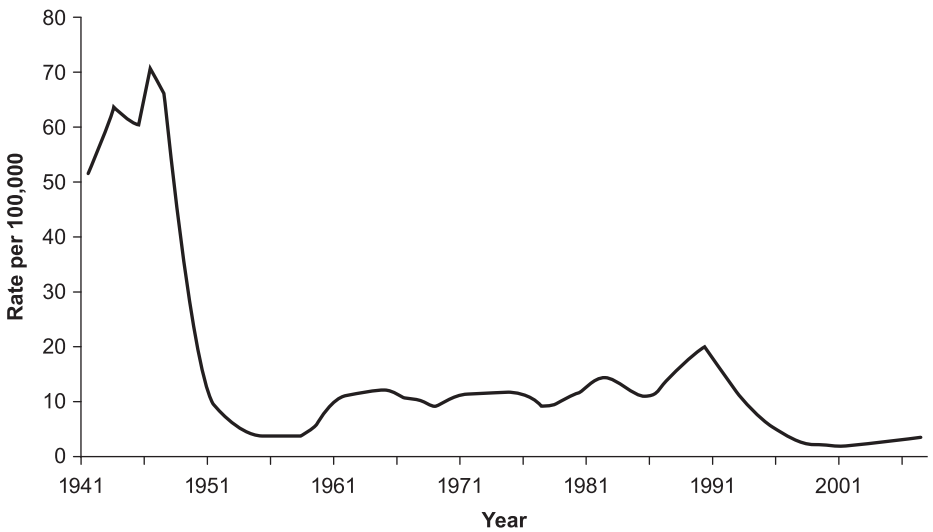

Figure 15.1 Primary and secondary syphilis rates, United States 1941–2007. (*Source:* CDC. Sexually Transmitted Disease Surveillance, 2007. Atlanta, GA: U.S. Department of Health and Human Services; 2008 December, 2008.)

secondary syphilis rates increased 75 percent with cases attributed to heterosexual transmission among African Americans in the southern United States.[42] Overall rates, however, steadily declined over the course of 50 years, and syphilis was considered a candidate for elimination in the United States in 1999.[43] Soon after the launch of a national elimination campaign, rates began to rise. From 2001 to 2007, overall rates of primary and secondary syphilis increased from 2.2 to 3.8 per 100,000 (see figure 15.2). This modest increase masks noteworthy gender differences: rates in women decreased 35 percent while rates in men increased 153 percent.[11, 44] The increase in male cases has been attributed to an increase in syphilis among MSM,[45] with outbreaks occurring in large U.S. cities.[46–49] In 2003, an estimated 62 percent of national cases of primary and secondary syphilis were attributed to MSM transmission.[50]

The changing epidemic of syphilis highlights an additional challenge of STD control: evolving risk profiles require continual modification of control programs, such as revised prevention messages and adjusted targeted screening campaigns.[41] For instance, key risk factors identified in the late 1980s and early 1990s epidemic were crack use and trading sex for drugs or money.[42] Interventions on these risk factors might include increasing access to drug treatment programs and targeted screening in correctional and detention facilities. These interventions may not be applicable to the young MSM driving the current epidemic. Additionally, inventions must take into account the social context of the epidemic, such as social norms, health care access, and community resources. An intervention targeting MSM in the Castro district of San Francisco might look different from an intervention targeting MSM in rural Mississippi.

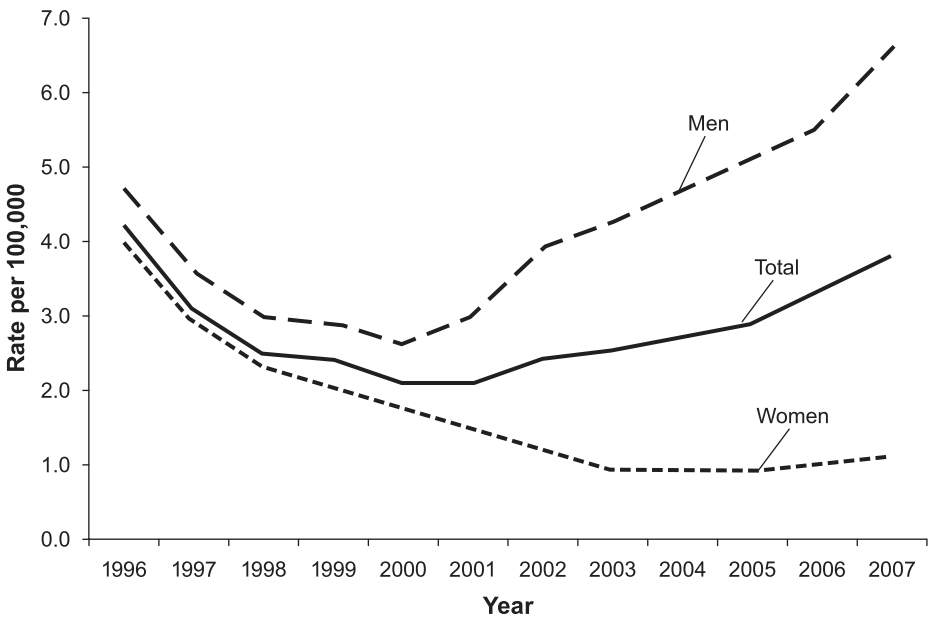

Figure 15.2 Primary and secondary syphilis rates by sex, United States, 1996–2007. (*Source:* CDC. Sexually Transmitted Disease Surveillance, 2007. Atlanta, GA: U.S. Department of Health and Human Services; 2008 December, 2008.)

KEY SYPHILIS CONTROL PRIORITIES

Screening

Identification of cases through screening and diagnostic tests has been the hallmark of syphilis control programs. In the original "platform for action" in 1937, Parran advocated for routine tests by insurance companies, employers, hospitals, and marriage license bureaus.[51] Widespread screening in the general population helped decrease late-stage complications from the disease including syphilis-related mortality. In the 21st century, complications from late-stage syphilis are uncommon, and screening primarily is used to find infectious cases to prevent further transmission. Unfortunately, finding infectious cases is challenging. Simply put, it is difficult to find cases when less than 0.1 percent of the population has infectious syphilis,[52] and screening the general population is neither sensible nor cost-effective. The one exception to this is antenatal screening. Syphilis is highly transmissible during pregnancy and the consequences are severe. As such, the U.S. Preventive Services Task Force (USPSTF) recommends routine screening during prenatal care,[53] and 90 percent of U.S. states mandate at least one syphilis test during pregnancy.[54] Although these recommendations require screening large, low-prevalence populations, the high-cost of complications from congenital syphilis make this screening program sensible and cost-effective.

One strategy to increase early syphilis diagnosis is to target screening to areas where high-risk populations may congregate, such as outreach in MSM bars and commercial sex venues. These "stand-alone" outreach campaigns, however, often are expensive as they require additional infrastructure and the yield may be low; only 0.2 percent of the 14,143 MSM screened in nonmedical settings in seven cities were newly diagnosed with primary or secondary syphilis.[55] One innovative way to target high-risk populations for community screening was pilot tested in San Francisco. In this program, men could request a syphilis test online, and then visit a local private lab to have their blood drawn. Results could then be accessed online. Over the course of a year, approximately 200 men used the service and the program reported an early syphilis positivity of 2.3 percent.[56] Although the positivity percent was higher than other community screening programs (self-selection bias), the absolute number of cases identified ($N = 5$) was low. Outreach screening, virtual or otherwise, is rarely cost-effective per case identified in low-prevalence settings.

An alternative control activity is to routinely screen at-risk populations in health care settings. About half of MSM with early syphilis also are HIV-infected,[57] and syphilis screening during HIV care visits presents a low-cost screening opportunity. In the United Kingdom, routine screening of HIV-infected outpatients identified 40 new cases of asymptomatic early syphilis (positivity of 1.6 percent).[58] In San Francisco, 1.8 percent ($N = 15$) of HIV-infected primary care patients had an incident syphilis infection diagnosed during routine testing.[59] During the crack-associated syphilis epidemic, screening during health care assessments in correctional settings increased access to the at-risk population. In Baton Rouge, Louisiana, in the late 1990s, about 25 percent of reported early syphilis was detected through jail screenings.[60] If the at-risk population is being reached, screening in correctional settings is cost-effective compared with costs of untreated complications [61] and may decrease community syphilis rates.[60]

Although routine testing of high-risk patients may pick up undiagnosed infections, many private health care providers do not regularly test their high-risk patients. Currently the USPSTF recommends routine screening of people at risk for syphilis,[53] but in an analysis of medical claims data in California, less than 40 percent of patients reporting a high-risk sexual behavior were given a syphilis test.[62] In a national survey of physicians, less than 20 percent of providers who see men reported that they routinely screened for syphilis.[63] Continued monitoring of adherence to screening guidelines is needed to identify populations not receiving services, allowing for targeted interventions such as provider training [64] and public-private partnership development.[65]

PARTNER NOTIFICATION

Partner notification is an important part of any screening program. Since it takes an average of three weeks for an infected person to become infectious (that is, develop a lesion), partners of individuals who have tested positive need to be identified and informed to attempt to stop further transmission. Partner notification (contact tracing) began with the national control plan in the 1940s and still is widely implemented. In a 1999 survey of public STD programs in high-morbidity areas, 87 percent of health departments reported that they attempted to contact all patients with infectious syphilis to provide partner notification, and 89 percent of infectious patients were successfully contacted and interviewed about their sex partners.[66] Interestingly, it was not until the late 1990s that the cost and effectiveness of the pervasive intervention were evaluated.[67] In a three-arm randomized control trial in three U.S. cities, partner notification for infectious syphilis resulted in the location of about one partner per index case, with treatment (either after diagnosis or prophylactically) of about 0.65 contacts per index case. Costs ranged from $317 to $362 per partner treated. Based on a review of published studies, it is estimated that partner notification finds, on average, one new case of syphilis for every four to five partners interviewed.[68] Partner notification may be more useful in identifying "high-value" cases (that is, those likely to transmit syphilis based on stage of infection and number of partners). In a review of intervention methods, partner notification was estimated to identify 11 percent of high-value male cases compared with 47 percent identified by self-referral to STD clinics and 14 percent identified in jail screenings.[69] Although the empirical evidence for the effectiveness of partner notification may be limited, other aspects to the intervention, such as identification of community sexual networks and public health's "ethical obligation to warn the unsuspecting," are considered to be program benefits.[70] Additionally, as incidence decreases, the cost of finding a case by partner notification may become comparable to (or less than) the cost of finding a case via screening.

One of the challenges of partner notification is the inability to locate partners, often due to incomplete contacting information. In one study, more than 80 percent of partners named by the index case were not able to be located.[67] Recent increases in anonymous sex partners met via the Internet [71] have frustrated partner notification efforts, and public health departments responded with new approaches. Prompted by a syphilis outbreak linked to an Internet chat room, the health department in San Francisco contacted hundreds of users of the chat room via screen

names. Although the sample was small (seven index cases), contact tracing resulted in 5.9 sexual partners medically evaluated per index case, with more than 40 percent of all named partners tested.[72] Online partner notification seems acceptable to MSM. In a national survey of more than 1,800 MSM, 70 percent said that if they were infected with an STD, they would use online partner notification through a public health specialist.[73] Although this strategy may be useful, additional resources will be required to train partner notification staff and develop guidelines.

A second challenge to partner notification is that services may not be accepted by the infected individual. Confidential partner notification for syphilis is usually provided by a Disease Intervention Specialist (DIS) who is affiliated with a public health authority. Preference for partner-delivered notification may keep some individuals from cooperating with a DIS [74] and even if the infected individual does notify their partner(s), an important opportunity may be a missed for treatment if the partner does not seek care. Fear of partner reaction may discourage participation as notification may end relationships and result in emotional or physical violence.[75] Additionally, syphilis often is diagnosed outside of public health clinics, particularly for MSM. Patients may not be willing to come to public health clinics for interviews or may be difficult to contact. One way to alleviate the participation obstacle is to have a DIS placed in settings outside of public health clinics, such as HIV care facilities and gay men's health centers.[76] Enhanced training of partner notification staff, including supplemental interviewing techniques,[68] as well as ensuring staff have necessary tools (for example, the most current guidance, access to computers) may increase the effectiveness of the intervention.[77]

Provider Education

Provider awareness of early syphilis symptoms is important to ensure accurate diagnosis and treatment during the infectious period. Unfortunately, many private providers, especially in low-prevalence areas, do not recognize symptoms of early syphilis or may confuse them with other ulcerative diseases. This may translate into a delay in treatment as well as a delay in partner notification. In Arizona, only 55 percent of patients diagnosed with primary or secondary syphilis in a non-STD clinic were treated within a week compared with 93 percent of patients diagnosed at an STD clinic.[78] In the same study, less than 20 percent of non-STD clinic patients were interviewed for contact tracing within the week compared with 95 percent of STD clinic patients. In Australia, a review of all early symptomatic cases of early syphilis among MSM documented similar missed opportunities by general practitioners. In the sample of 123 men, 34 cases were first seen by a private provider who had made a referral to the public health center without treatment.[79]

Delays in treatment may be due to the challenges of diagnosing infectious syphilis with serologic tests. In a survey of infectious disease physicians, more than 80 percent did not have access to the darkfield confirmatory test for primary syphilis diagnosis. While the majority presumptively treated for syphilis based on lesion presence, some physicians reported that if the serologic test was negative, they would assume the lesion was not syphilitic.[80] With the low sensitivity of serologic tests for primary stage disease, these physicians could be misdiagnosing 20 to 30 percent of cases. Even if an accurate diagnosis is made, providers may not report cases of

syphilis. In a national survey of community-based physicians, only 50 percent of providers stated that they reported the patient's name to the health department after a syphilis diagnosis.[63] In HIV care settings, providers may be concerned about patient confidentiality or "getting their patient in trouble" by reporting an incident STD. Mandatory lab reporting can alleviate provider nonparticipation, but may result in reporting delays.

As part of the response to the current MSM epidemic, health departments in eight cities undertook targeted activities to decrease diagnosis, improve treatment, and report barriers among non-STD clinic providers in the early 2000s.[64] Tactics included activities to increase provider awareness of symptoms and reporting regulations. The New York Department of Health distributed more than 4,000 STD information packets, including treatment and case report guidelines, to incoming medical residents and fellows. In Los Angeles, information packets, which included posters and color photographs of symptoms of primary and secondary syphilis were sent to HIV care providers. To address delays in treatment, San Francisco STD nurses offered in-person training to private providers on how to administer treatment, as well as provided free benzathine penicillin upon request. To address barriers in reporting delays in New York, the health department offered providers the option of reporting early syphilis cases by phone instead of the traditional paper case report. Although anecdotal evidence suggests these interventions have increased provider knowledge and testing uptake, no empirical data are available to evaluate their impact or cost-effectiveness.

Public Education

For control programs to effect population disease incidence, the scope and scale must be sufficient. Public education campaigns can help increase the community-level impact of other activities of syphilis control by effectively targeting at-risk populations. For example, syphilis is infectious only during early stages when a lesion is present. Consequently, use of screening and treatment programs are effective in stopping the spread of disease only if patients present for care during an early stage. Many men do not know what lesions look like, do not know where to look (for example, in their mouth or anus in addition to their genitals), or think that when the sore goes away the infection is gone. Targeted education campaigns focusing on how to recognize signs and symptoms of syphilis may serve to minimize delays in seeking treatment after symptoms.

In addition to not recognizing syphilis symptoms, many people, particularly MSM, do not consider syphilis a problem and may not seek routine testing. In a qualitative study of syphilis beliefs among MSM in the United Kingdom, men discussed the rarity of the diseases: "anything else you pick up now, as I said like gonorrhea or chlamydia, it's sort of common. Not nice, but OK, lots of people get it. Syphilis is like, 'is that still going on out there?'"[81]

This is likely due in part to the success of elimination campaigns that have resulted in low- or no-morbidity areas. For example, in 2007, the entire state of Idaho reported one case of early syphilis and even in high-morbidity areas, such as Louisiana, only 533 cases were reported.[82] Compared with STDs such as chlamydia (3,722 cases in Idaho and 19,362 cases in Louisiana in 2007[83]), syphilis is a

rare disease. In some subpopulations such as MSM in urban areas, however, the disease is not rare and social marketing campaigns to raise awareness are needed. In 2003, the San Francisco Department of Public Health launched the Healthy Penis campaign.[84] Based on input from a community advisory board, the campaign focused on increasing testing through a branded message that was informative in addition to being humorous and sex-positive. Six months after the launch of the campaign, a convenience sample of men reporting male sex partners was interviewed (n = 244). Men who reported being aware of the campaign also were more likely to report being tested for syphilis and had increased syphilis knowledge, including knowing the disease symptoms.

A challenge of syphilis education campaigns is that they often overlap and sometimes contradict other STD prevention interventions. Widespread HIV prevention programs have resulted in prevention fatigue for some populations, including MSM,[85] and may account for increased sexual risk behaviors contributing to increases in syphilis.[86] Although some risk reduction messages apply to prevention of both diseases (for example, using condoms during anal sex), some messages may conflict. HIV risk reduction messages often use a risk continuum method and oral sex may be presented as a "safer" sexual activity for HIV transmission.[87] Oral sex, however, can be efficient for transmitting syphilis.[48] This conflicting message may contribute to increases in syphilis rates among MSM.

SURVEILLANCE

Syphilis has been routinely reported for national surveillance by state and city epidemiologists since 1941.[88] At the population level, case surveillance identifies trends, helps in fiscal planning by documenting burden of disease, and can be used to evaluate prevention and control programs. Risk profiles of new cases also can be used to identify individuals most likely to acquire the next syphilis infections. Knowing who is likely to become infected can help practitioners tailor risk reduction interventions and target screening campaigns to diagnosis and treat infectious cases, thus preventing further transmission. Unfortunately, uniformly collected information on syphilis morbidity reports is limited. Unlike the national HIV/AIDS case surveillance system that compiles risk behavior data in additional to demographics, national syphilis case reports were (until recently) were limited to age, gender, race or ethnicity, and geographic location (county or city). Without behavioral data, transmission of syphilis among MSM had to be estimated using a change in the male-to-female sex ratio of cases.[50] In 2005, gender of sex partner was added to case report data allowing for a more detailed risk profile.

Supplementing morbidity data with special studies can enhance disease surveillance and aid in control program design. The San Francisco Department of Public Health responded to an early syphilis epidemic in MSM with a multifaceted approach, which began with expanded risk factor surveillance.[89] Staff used data from DIS interviews with index cases and partners to supplement risk profiles, including information on venues for meeting sex partners and recreational drug use. They conducted surveys of patients seeking care in the city STD clinic and, after linking surveys to test results, identified risk factors for syphilis infection. In response to similar rising rates in Florida, the Florida Bureau of STD Prevention and Control

supplemented routine data with an ethnographic assessment of both MSM and their health care providers. Learning from the interviews that much of the at-risk population was not able to access care during traditional hours, the bureau subsequently expanded clinic hours in key locations. During the first six months of the program, one clinic provided care for 401 patients during extended clinic hours of which 94 percent were MSM, 18 had primary or secondary syphilis, and 33 had early latent syphilis.[90]

Surveillance is necessary for identification and management of outbreaks. As part of the National Plan for the Elimination of Syphilis in the United States, the CDC provides guidelines that outline seven key elements of outbreak response, including expanded surveillance data review and development of tailored interventions.[91] For areas with endemic syphilis, it is likely that the required infrastructure to detect changes in disease patterns and respond quickly is in place. For example, Los Angeles County has an established outbreak response plan that can be activated whenever an increase in cases is noted.[92] After an increase in cases among MSM in early 2000, the county worked with more than 140 outbreak response partners to launch a $450,000 media campaign, hold 80 community screening events, and start a prophylactic treatment program in the county jail. This level of outbreak response is unfeasible for areas of the country with low syphilis morbidity. One solution is for state and national partners to develop rapid response teams that can be deployed to local communities when indicated.[43] Activation of response still requires continual monitoring of syphilis surveillance to detect changes, as well as an understanding of local epidemiology to know whether changes are indicative of a shift in the disease prevalence.

OTHER TOOLS FOR PREVENTION AND CONTROL

Although the majority of current syphilis control efforts focus on Parran's original paradigm of case finding, treatment, partner notification, and education, other control methods, some of which are still under development, may be important to effective syphilis control in the future.

In the early 1970s, a syphilis vaccine was shown to be theoretically feasible in animal models, but the tested inoculation was long and invasive (60 injections over 37 weeks).[93] Little progress has been made since then. In 1998, however, the complete genome sequence of *T. pallidum* was determined, which may help progress vaccine development.[94] In December 2008, nearly $4.7 million was provided to the Washington Vaccine Alliance as part of the state's tobacco lawsuit settlement. In addition to supporting research for *E. coli* and herpes vaccines, the funds will support preclinical and Phase 1 clinical trials for a syphilis vaccine.[95]

As syphilis has an effective cure, one theoretical control option is mass treatment. Few efforts have been documented using this approach. During a syphilis outbreak among seasonal farmworkers in California in the mid-1970s, county health officials were unable to control the epidemic using contact tracing because the majority of men reported sex worker partners with no locating information. Officials launched a "prostitute testing and treatment program" to selectively mass treat all sex workers who were not allergic to penicillin. More than 500 women participated. Although health officials were unable to link the reduction directly to the mass

treatment program, in the next year, reported cases of infectious syphilis among seasonal farmworkers decreased.[96] Similar tactics were used in a cocaine-related outbreak in Pennsylvania in the late 1980s, and evaluators concluded that the targeted mass treatment was less costly and time-consuming than contact tracing.[97]

More recently, a sex trade–related outbreak in 2000 in British Columbia prompted mass treatment using azithromycin. Although reported infections decreased in the six months following mass treatment, the case rate rebounded and resulted in a higher than expected number of cases the following year.[98] Mathematical modelers used the British Columbia data to investigate why mass treatment may have failed to affect morbidity and concluded that mass treatment campaigns will be effective only if core transmitters are treated and the population is not mobile.[99]

An additional treatment control strategy is expedited partner therapy (EPT), which includes patient-delivered partner treatment (PDPT). EPT can reduce treatment barriers for partners of infected patients, as well as help serve to reduce the burden on health department staff. Studies have shown this method to be at least as effective as patient-delivered partner notification and referral for chlamydia and gonorrhea, which is less intensive than partner notification for syphilis. The CDC recommends that EPT be available and be considered as an additional control strategy.[100]

The usual treatment for syphilis is an injection of benzathine penicillin. Although some programs have attempted PDPT for syphilis control using oral treatment, it is not considered a viable option for syphilis control. In San Francisco, for example, in response an increase in unlocatable Internet partners, syphilis infected patients were given "partner packs," which contained a single 1 g dose of azithromycin.[89] Reports of treatment failure,[101] and a rapid increase in incidence of azithromycin-resistant syphilis,[39] led the health department to eliminate PDPT for syphilis control. At this time, PDPT is not a recommended control method for syphilis.[100]

A new *point-of-care syphilis test* is currently being evaluated and is under consideration for U.S. Food and Drug Administration approval. The test uses fingerstick blood specimens and rapidly detects both treponemal and nontreponemal antibodies. As many patients fail to return for their STD results,[102] a point-of-care test would allow patients to be tested, receive their results, and be treated if necessary on the same day. This type of screening tool may be especially useful in prenatal care settings, particularly in low-resource settings, because it eliminates costs associated with transportation of specimens and multiple visits, as well as reduces loss to follow-up.[103]

CONCLUSION

In 1931, at the sixtieth annual meeting of the American Public Health Association, William Munson, a district state health officer, stated that "the control and eradication of syphilis seems . . . to be one of the easiest [problems] we have in public health."[104] Dr. Munson bravely made this statement even before the availability of penicillin, believing that public health control efforts could eradicate the disease. More than 70 years later, control efforts have yet to eliminate syphilis, or chlamydia, or gonorrhea, even though screening tests and effective treatment are available for all three diseases. The continued spread of these curable diseases highlights what

was summarized succinctly by Allan Brandt in his 1987 text on the social history of STD: essentially, he said that there is no magic bullet for STD control,[105] and these challenges are only amplified with incurable viral diseases. Although barriers to STD control are formidable, there are some areas where STD rates are low, and some sexually transmitted infections are now seen rarely. Hence, STD control and eradication is possible, but not easy. More effective implementation of existing strategies, development of new testing technologies, and new vaccines are all needed.

REFERENCES

1. Fight the Enemy at Home! Emergence of Advertising in America: 1850–1920.
2. Barnouw E. *Media Marathon: A Twentieth-Century Memoir*. Durham, NC: Duke University Press; 1996.
3. Clark L. *VD Attack Plan*. United States: Walt Disney Productions; 1973:16 min.
4. Get Yourself Tested. http://www.gyt09.org. Accessed November 9, 2009.
5. Institutes of Medicine. *The Hidden Epidemic: Confronting Sexually Transmitted Diseases*. Washington, DC: Institute of Medicine; 1997.
6. Holmes KK, Stamm WE, Piot P, et al., eds. *Sexually Transmitted Diseases*. 4th ed. New York: McGraw Hill Medical; 2008.
7. Schmid G, Rowley J, Samuelson, et al. World Health Organization 2005 global estimates of the incidence and prevalence of sexually transmitted infections. In: World Health Organization, ed. *18th Meeting of the International Society for Sexually Transmitted Diseases Research*. London: World Health Organization; 2009.
8. Hall-Baker PA, Nieves E, Jajosky RA, et al. Summary of notifiable diseases: United States, 2007. *MMWR*. 2009;56(53):1–94.
9. Weinstock H, Berman S, Cates W Jr. Sexually transmitted diseases among American youth: incidence and prevalence estimates, 2000. *Perspect Sex Reprod Health*. 2004;36(1):6–10.
10. Forhan SE, Gottlieb SL, Sternberg MR, et al. Prevalence of sexually transmitted infections and bacterial vaginosis among female adolescents in the United States: data from the National Health and Nutritional Examination Survey, 2003–2004: National STD Prevention Conference; March 10–13, 2008; Chicago, IL.
11. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance, 2007*. Atlanta, GA: US Department of Health and Human Services; 2008.
12. Adimora AA, Schoenbach VJ. Social context, sexual networks, and racial disparities in rates of sexually transmitted infections. *J Infect Dis*. 2005;191(Suppl 1):S115–S122.
13. World Health Organization. *The Top 10 Causes of Death: Fact sheet No 310*. Geneva, Switzerland: World Health Organization; 2008.
14. Centers for Disease Control and Prevention, National Center for Health Statistics. *Deaths, Percent of Total Deaths, and Death Rates for the 15 Leading Causes of Death in 5-year Age Groups, by Race and Sex: United States, 2004*. Hyattsville, MD: Public Health Service; 2004.
15. Parkin DM, Bray F. The burden of HPV-related cancers. *Vaccine*. 2006;24(Suppl 3):S3, 11–25.
16. World Health Organization. *The Global Elimination of Congenital Syphilis: Rationale and Strategy for Action*. Geneva, Switzerland: World Health Organization; 2007.
17. Schaller UC, Klauss V. Is Crede's prophylaxis for ophthalmia neonatorum still valid? *Bull World Health Organ*. 2001;79(3):262–263.

18. Hutchinson AB, Farnham PG, Dean HD, et al. The economic burden of HIV in the United States in the era of highly active antiretroviral therapy: evidence of continuing racial and ethnic differences. *J Acquir Immune Defic Syndr*. 2006;43(4):451–457.
19. World Health Organization. *World Health Report: Changing History, Statistical Annex*. Geneva, Switzerland: World Health Organization; 2004.
20. Centers for Disease Control and Prevention. *VD Statistical Letter*. Atlanta, GA: US Department of Health and Human Services; 1977.
21. May RM, Anderson RM. Transmission dynamics of HIV infection. *Nature* 1987;326(6109):137–142.
22. Fee E. Veneral disease: the wages of sin? In: Peiss K, Simmons C, eds. *Passion and Power: Sexuality in History*. Philadelphia, PA: Temple University Press; 1989.
23. Brandt AM. The syphilis epidemic and its relation to AIDS. *Science*. 1988; 239(4838):375–380.
24. Chong J-R. Anti-Syphilis TV message finds few takers. *Los Angeles Times*. December 2, 2004.
25. Coates RA, Calzavara LM, Soskolne CL, et al. Validity of sexual histories in a prospective study of male sexual contacts of men with AIDS or an AIDS-related condition. *Am J Epidemiol*. 1988;128(4):719–728.
26. Peterman TA, Aral SO. Exact reproduction of survey instrument. *Am J Epidemiol*. 1989;129(6):1310–1311.
27. Harper KN, Ocampo PS, Steiner BM, et al. On the origin of the treponematoses: a phylogenetic approach. *PLoS Negl Trop Dis*. 2008;2(1):e148.
28. Parascandola J. *Sex, Sin, and Science*. Westport, CT: Praeger; 2008.
29. Schroeter AL, Turner RH, Lucas JB, Brown WJ. Therapy for incubating syphilis effectiveness of gonorrhea treatment. *JAMA*. 1971;218(5):711–713.
30. Shafii T, Radolf JD, Sanchez PJ, Schulz KF, Murphy FK. Congenital syphilis. In: Holmes KK, Stamm WE, Piot P, et al., eds. *Sexually Transmitted Diseases*. 4th ed. New York: McGraw Hill Medical; 2008:1577–1612.
31. Harrison LW. The Oslo study of untreated syphilis, review and commentary. *Br J Vener Dis*. 1956;32(2):70–78.
32. Sparling PF, Swartz MN, Musher DM, Healy BP. Clinical manifestations of syphilis. In: Holmes KK, Stamm WE, Piot P, et al., eds. *Sexually Transmitted Diseases*. 4th ed. New York: McGraw Hill Medical; 2008:661–684.
33. Larsen SA, Steiner BM, Rudolph AH. Laboratory diagnosis and interpretation of tests for syphilis. *Clin Microbiol Rev*. 1995;8(1):1–21.
34. Sartin JS, Perry HO. From mercury to malaria to penicillin: the history of the treatment of syphilis at the Mayo Clinic, 1916–1955. *J Am Acad Dermatol*. 1995;32(2 Pt 1):255–261.
35. Centers for Disease Control and Prevention. Update to CDC's sexually transmitted diseases treatment guidelines, 2006: fluoroquinolones no longer recommended for treatment of gonococcal infections. *MMWR*. 2007;56(14):332–336.
36. Workowski KA, Berman SM. Sexually transmitted diseases treatment guidelines, 2006. *MMWR*. 2006;55(RR-11):1–94.
37. Bai ZG, Yang KH, Liu YL, et al. Azithromycin vs. benzathine penicillin G for early syphilis: a meta-analysis of randomized clinical trials. *Int J STD AIDS*. 2008;19(4):217–221.
38. Riedner G, Rusizoka M, Todd J, et al. Single-dose azithromycin versus penicillin G benzathine for the treatment of early syphilis. *N Engl J Med*. 2005;353(12): 1236–1244.
39. Lukehart SA, Godornes C, Molini BJ, et al. Macrolide resistance in Treponema pallidum in the United States and Ireland. *N Engl J Med*. 2004;351(2):154–158.

40. Hook EW III, Marra CM. Acquired syphilis in adults. *N Engl J Med*. 1992; 326(16):1060–1069.
41. Peterman TA, Heffelfinger JD, Swint EB, Groseclose SL. The changing epidemiology of syphilis. *Sex Transm Dis*. 2005;32(10 Suppl):S4–S10.
42. Centers for Disease Control and Prevention. Primary and secondary syphilis: United States, 1981–1990. *MMWR*. 1991;40(19):314–315, 321–323.
43. Centers for Disease Control and Prevention. *The National Plan to Eliminate Syphilis from the United States*. Atlanta, GA: US Department of Health and Human Services; 1999.
44. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance, 2003*. Atlanta, GA: US Department of Health and Human Services; 2004.
45. Peterman TA, Furness BW. The resurgence of syphilis among men who have sex with men. *Curr Opin Infect Dis*. 2007;20(1):54–59.
46. Centers for Disease Control and Prevention. Resurgent bacterial sexually transmitted disease among men who have sex with men: King County, Washington, 1997–1999. *MMWR*. 1999;48(35):773–777.
47. Centers for Disease Control and Prevention. Primary and secondary syphilis among men who have sex with men: New York City, 2001. *MMWR*. 2002;51(38): 853–856.
48. Centers for Disease Control and Prevention. Transmission of primary and secondary syphilis by oral sex: Chicago, Illinois, 1998–2002. *MMWR*. 2004;53(41): 966–968.
49. Centers for Disease Control and Prevention. Outbreak of syphilis among men who have sex with men: Southern California, 2000. *MMWR*. 2001;50(7):117–120.
50. Heffelfinger JD, Swint EB, Berman SM, Weinstock HS. Trends in primary and secondary syphilis among men who have sex with men in the United States. *Am J Public Health*. 2007;97(6):1076–1083.
51. Parran T. *Shadow on the Land*. New York: Reynal and Hitchcock; 1937.
52. Gottlieb SL, Pope V, Sternberg MR, et al. Prevalence of syphilis seroreactivity in the United States: data from the National Health and Nutrition Examination Surveys 2001–2004. *Sex Transm Dis*. 2008;35(5):507–511.
53. US Preventative Services Task Force. *Screening for Syphilis Infection: Recommendation Statement*. Rockville, MD: Agency for Healthcare Research and Quality; 2004.
54. Hollier LM, Hill J, Sheffield JS, Wendel GD Jr. State laws regarding prenatal syphilis screening in the United States. *Am J Obstet Gynecol*. 2003;189(4): 1178–1183.
55. Ciesielski C, Kahn RH, Taylor M, Gallagher K, Prescott LJ, Arrowsmith S. Control of syphilis outbreaks in men who have sex with men: the role of screening in nonmedical settings. *Sex Transm Dis*. 2005;32(10 Suppl):S37–S42.
56. Levine DK, Scott KC, Klausner JD. Online syphilis testing: confidential and convenient. *Sex Transm Dis*. 2005;32(2):139–141.
57. Buchacz K, Greenberg A, Onorato I, Janssen R. Syphilis epidemics and human immunodeficiency virus incidence among men who have sex with men in the United States: implications for HIV prevention. *Sex Transm Dis*. 2005;32(10 Suppl):S73–S79.
58. Cohen CE, Winston A, Asboe D, et al. Increasing detection of asymptomatic syphilis in HIV patients. *Sex Transm Infect*. 2005;81(3):217–219.
59. Phipps W, Stanley H, Kohn R, Stansell J, Klausner JD. Syphilis, chlamydia, and gonorrhea screening in HIV-infected patients in primary care: San Francisco, California, 2003. *AIDS Patient Care STDS*. 2005;19(8):495–498.

60. Kahn RH, Scholl DT, Shane SM, Lemoine AL, Farley TA. Screening for syphilis in arrestees: usefulness for community-wide syphilis surveillance and control. *Sex Transm Dis*. 2002;29(3):150–156.
61. Silberstein GS, Coles FB, Greenberg A, Singer L, Voigt R. Effectiveness and cost-benefit of enhancements to a syphilis screening and treatment program at a county jail. *Sex Transm Dis*. 2000;27(9):508–517.
62. Tao G, Irwin KL. Receipt of HIV and STD testing services during routine general medical or gynecological examinations: variations by patient sexual risk behaviors. *Sex Transm Dis*. 2008;35(2):167–171.
63. St Lawrence JS, Montano DE, Kasprzyk D, Phillips WR, Armstrong K, Leichliter JS. STD screening, testing, case reporting, and clinical and partner notification practices: a national survey of US physicians. *Am J Public Health*. 2002; 92(11):1784–1788.
64. Taylor M, Prescott L, Brown J, et al. Activities to increase provider awareness of early syphilis in men who have sex with men in 8 cities: 2000–2004. *Sex Transm Dis*. 2005;32(10 Suppl):S24–S29.
65. Swain GR, Kowalewski SJ, Schubot DB. Reducing the incidence of congenital syphilis in Milwaukee: a public/private partnership. *Am J Public Health*. 1998; 88(7):1101–1102.
66. Golden MR, Hogben M, Handsfield HH, St Lawrence JS, Potterat JJ, Holmes KK. Partner notification for HIV and STD in the United States: low coverage for gonorrhea, chlamydial infection, and HIV. *Sex Transm Dis*. 2003;30(6): 490–496.
67. Peterman TA, Toomey KE, Dicker LW, Zaidi AA, Wroten JE, Carolina J. Partner notification for syphilis: a randomized, controlled trial of three approaches. *Sex Transm Dis*. 1997;24(9):511–518.
68. Brewer DD, Potterat JJ, Muth SQ, et al. Randomized trial of supplementary interviewing techniques to enhance recall of sexual partners in contact interviews. *Sex Transm Dis*. 2005;32(3):189–193.
69. Kahn RH, Peterman TA, Arno J, Coursey EJ, Berman SM. Identifying likely syphilis transmitters: implications for control and evaluation. *Sex Transm Dis*. 2006;33(10):630–635.
70. Potterat JJ. Contact tracing's price is not its value. *Sex Transm Dis*. 1997;24(9): 519–521.
71. Garofalo R, Herrick A, Mustanski BS, Donenberg GR. Tip of the iceberg: young men who have sex with men, the Internet, and HIV risk. *Am J Public Health*. 2007;97(6):1113–1117.
72. Klausner JD, Wolf W, Fischer-Ponce L, Zolt I, Katz MH. Tracing a syphilis outbreak through cyberspace. *JAMA*. 2000;284(4):447–449.
73. Mimiaga MJ, Fair AD, Tetu AM, et al. Acceptability of an internet-based partner notification system for sexually transmitted infection exposure among men who have sex with men. *Am J Public Health*. 2008;98(6):1009–1011.
74. Mimiaga MJ, Reisner SL, Tetu AM, et al. Partner notification after STD and HIV exposures and infections: knowledge, attitudes, and experiences of Massachusetts men who have sex with men. *Public Health Rep*. 2009;124(1):111–119.
75. Kissinger PJ, Niccolai LM, Magnus M, et al. Partner notification for HIV and syphilis: effects on sexual behaviors and relationship stability. *Sex Transm Dis* 2003;30(1):75–82.
76. Hogben M, Paffel J, Broussard D, et al. Syphilis partner notification with men who have sex with men: a review and commentary. *Sex Transm Dis* 2005; 32(10 Suppl):S43–S47.

77. Centers for Disease Control and Prevention. *The National Plan to Eliminate Syphilis from the United States*. Atlanta, GA: US Department of Health and Human Services; 2006.
78. Chen SY, Johnson M, Sunenshine R, England B, Komatsu K, Taylor M. Missed and delayed syphilis treatment and partner elicitation: a comparison between STD clinic and non-STD clinic patients. *Sex Transm Dis*. 2009;36(7):445–451.
79. Bissessor M, Fairley CK, De Guingand D, Bradshaw CS, Chen MY. Delay in the diagnosis of early syphilis among men who have sex with men: need for greater community and health provider education. *Int J STD AIDS*. 2009;20(1):52–53.
80. Dowell D, Polgreen PM, Beekmann SE, Workowski KA, Berman SM, Peterman TA. Management dilemmas in syphilis: a survey of infectious disease experts. *Clin Infect Dis*. 2009.doi:10.1086/644737.
81. Lambert NL, Imrie J, Fisher MJ, Phillips A, Watson R, Dean G. Making sense of syphilis: beliefs, behaviors and disclosure among gay men recently diagnosed with infectious syphilis and the implications for prevention. *Sex Health*. 2006;3(3):155–161.
82. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2007 Supplement, Syphilis Surveillance Report*. Atlanta, GA: US Department of Health and Human Services; 2009.
83. Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance 2007 Supplement, Chlamydia Prevalence Monitoring Project Annual Report 2007*. Atlanta, GA: US Department of Health and Human Services; 2009.
84. Montoya JA, Kent CK, Rotblatt H, McCright J, Kerndt PR, Klausner JD. Social marketing campaign significantly associated with increases in syphilis testing among gay and bisexual men in San Francisco. *Sex Transm Dis*. 2005;32(7):395–399.
85. Heavner KK, Shaw B, Battles H, McClamroch K, Tesoriero J, Birkhead GS. HIV prevention fatigue: men who have sex with men are tuning out HIV prevention messages: American Public Health Association 135th Annual Meeting and Expo; November 3–7, 2007; Washington, DC.
86. Ostrow DG, Silverberg MJ, Cook RL, et al. Prospective study of attitudinal and relationship predictors of sexual risk in the multicenter AIDS cohort study. *AIDS Behav*. 2008;12(1):127–138.
87. Varghese B, Maher JE, Peterman TA, Branson BM, Steketee RW. Reducing the risk of sexual HIV transmission: quantifying the per-act risk for HIV on the basis of choice of partner, sex act, and condom use. *Sex Transm Dis*. 2002;29(1):38–43.
88. Cases of syphilis and gonorrhea reported to the public health service by state and territorial health departments, fiscal years 1941–1951. *J Vener Dis Inf*. 1951;32(12):392–400.
89. Klausner JD, Kent CK, Wong W, McCright J, Katz MH. The public health response to epidemic syphilis: San Francisco, 1999–2004. *Sex Transm Dis*. 2005;32(10 Suppl):S11–S18.
90. Schmitt K, Bulecza S, George D, Burns TE, Jordahl L. Florida's multifaceted response for increases in syphilis among MSM: the Miami-Ft. Lauderdale initiative. *Sex Transm Dis*. 2005;32(10 Suppl):S19–S23.
91. Finelli L, Levine WC, Valentine J, St Louis ME. Syphilis outbreak assessment. *Sex Transm Dis*. 2001;28(3):131–135.
92. Chen JL, Kodagoda D, Lawrence AM, Kerndt PR. Rapid public health interventions in response to an outbreak of syphilis in Los Angeles. *Sex Transm Dis*. 2002;29(5):277–284.

93. Miller JN. Immunity in experimental syphilis: successful vaccination of rabbits with Treponema pallidum, Nichols strain, attenuated by irradiation. *J Immunol*. 1973;110(5):1206–1215.
94. Fraser CM, Norris SJ, Weinstock GM, et al. Complete genome sequence of Treponema pallidum, the syphilis spirochete. *Science*. 1998;281(5375):375–388.
95. Fred Hutchinson Cancer Research Center. Life sciences discovery fund awards $4.7 million to support new vaccine program. http://www.fhcrc.org/about/ne/news/2008/12/16/wava.html. Published December 16, 2008.
96. Jaffe HW, Rice DT, Voigt R, Fowler J, St John RK. Selective mass treatment in a venereal disease control program. *Am J Public Health*. 1979;69(11):1181–1182.
97. Hibbs JR, Gunn RA. Public health intervention in a cocaine-related syphilis outbreak. *Am J Public Health*. 1991;81(10):1259–1262.
98. Rekart ML, Patrick DM, Chakraborty B, et al. Targeted mass treatment for syphilis with oral azithromycin. *Lancet*. 2003;361(9354):313–314.
99. Pourbohloul B, Rekart ML, Brunham RC. Impact of mass treatment on syphilis transmission: a mathematical modeling approach. *Sex Transm Dis*. 2003;30(4):297–305.
100. Centers for Disease Control and Prevention. *Expedited Partner Therapy in the Management of Sexually Transmitted Diseases*. Atlanta, GA: US Department of Health and Human Services; 2006.
101. Mitchell SJ, Engelman J, Kent CK, Lukehart SA, Godornes C, Klausner JD. Azithromycin-resistant syphilis infection: San Francisco, California, 2000–2004. *Clin Infect Dis*. 2006;42(3):337–345.
102. Schwebke JR, Sadler R, Sutton JM, Hook EW III. Positive screening tests for gonorrhea and chlamydial infection fail to lead consistently to treatment of patients attending a sexually transmitted disease clinic. *Sex Transm Dis*. 1997;24(4):181–184.
103. Rydzak CE, Goldie SJ. Cost-effectiveness of rapid point-of-care prenatal syphilis screening in sub-Saharan Africa. *Sex Transm Dis*. 2008;35(9):775–784.
104. Munson WL. Practicability of epidemiological methods in the control of syphilis. *Am J Public Health Nations Health*. 1932;22(2):134–140.
105. Brandt AM. *No Magic Bullet: A Social History of Venereal Disease in the United States Since 1880*. New York: Oxford University Press; 1987.
106. Zanetti AR, Van Damme P, Shouval D. The global impact of vaccination against hepatitis B: a historical overview. *Vaccine*. 2008;26(49):6266–6273.
107. Centers for Disease Control and Prevention. *Disease Burden from Viral Hepatitis A, B, and C in the United States*. Atlanta, GA: Centers for Disease Control and Prevention; 2009.
108. World Health Organization. *Global Strategy for the Prevention and Control of Sexually Transmitted Infections: 2006–2015*. Geneva, Switzerland: World Health Organization; 2007.
109. Hall HI, Song R, Rhodes P, et al. Estimation of HIV incidence in the United States. *JAMA*. 2008;300(5):520–529.
110. Centers for Disease Control and Prevention. HIV prevalence estimates: United States, 2006. *MMWR*. 2008;57(39):1073–1076.

This page intentionally left blank

SECTION 4

TREATMENT AND PREVENTION OF MENTAL HEALTH ILLNESS

This page intentionally left blank

CHAPTER 16

Depression and Public Health: An Overview

Enrico Tedeschini, MD, Paolo Cassano, MD, PhD, and Maurizio Fava, MD

INTRODUCTION

Depressive disorders, especially major depressive disorders (MDDs), are prevalent conditions among the general population. Anxiety disorders, mood disorders, eating disorders, childhood disorders, cognitive disorders, schizophrenia, and other psychotic disorders represent the MDDs affecting millions of individuals. Significant suffering, high morbidity and mortality, and psychosocial functional impairment typically are associated with depressive disorders. Unfortunately, despite the availability of numerous effective treatments, many depressive disorders often are misdiagnosed or underdiagnosed and often are undertreated. Several factors contributing to the poor recognition of depression have been identified, ranging from the stigma of depression to the relative lack of systematic ascertainment of depressive symptoms by primary care physicians (PCPs).

The medical, public health, and economic significance of depression is considerable. Individuals who are depressed tend to have a higher incidence of chronic illnesses compared with the general population. Similarly, depressed patients tend to present with more comorbid conditions, including chronic medical diseases. This chapter considers the issue of the public health relevance of depressive disorders and focuses on the recognition, management, and treatment of depression, particularly in the primary care setting. Public health strategies adopted to face and prevent the depressive disorders' burden to the community also are reviewed.

DEPRESSION: CLINICAL FEATURES AND DIFFERENTIAL DIAGNOSIS

The Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) defines MDD as depressive mood or reduced interest and pleasure, lasting for at least two weeks, accompanied by the vegetative, cognitive, and psychomotor symptoms for a total of at least five symptoms altogether.[1] The accompanying symptoms include insomnia or hypersomnia, loss of energy (fatigue), loss or increase of appetite or weight, diminished ability to think or concentrate, psychomotor

agitation or retardation, feelings of worthlessness or excessive guilt, and recurrent thoughts of death or suicide. The persistence of this syndrome for at least two years is called chronic depression. Minor depression is mild syndrome lasting for at least two weeks and is diagnosed when the depressed mood or lack of interest and pleasure is associated with a total number not exceeding three of these symptoms.

Irritability is not included as a symptom of MDD in the DSM-IV, although it is commonly found in adults with MDD (especially females, younger individuals) and is associated with distinctive clinical features. Evidence shows that irritability results in lower functional status and poorer quality of life and is associated with a history of at least one suicide attempt.[2–4] This raises the question whether irritability should be included as a symptom of MDD in the revised DSM-V and International Classification of Diseases (ICD)-11 diagnostic systems.

Patients who suffer from depressive disorders typically present with a constellation of psychological, behavioral, and physical symptoms. Tables 16.1, 16.2, and 16.3 list some of the most common psychological (see table 16.1), behavioral (see table 16.2), and physical symptoms (see table 16.3) reported by patients with depressive disorders. Traditionally, the diagnostic approach to depressive disorders identified depressed mood and loss of interest and pleasure in most activities as key features of these conditions. The cluster of associated symptoms, the duration of the syndrome, as well as the degree of functional impairment are essential to distinguish depressive disorders from physiological mood variability. The continuum of depression from mild, short-lasting syndromes toward severe, chronic or recurrent, and disabling disorders has been well documented.[5, 6]

Table 16.1.
Unipolar Depression: Common Psychological Symptoms

- Depressed mood
- Irritability
- Anxiety and nervousness
- Reduced concentration
- Lack of interest and motivation
- Inability to enjoy things
- Lack of pleasure or anhedonia
- Reduced libido
- Hypersensitivity to rejection and criticism
- Perfectionism and obsessiveness
- Indecisiveness
- Pessimism and hopelessness
- Feelings of helplessness
- Cognitive distortions (for example, "I am unlovable")
- Preoccupation with oneself
- Low self-esteem
- Feelings of worthlessness
- Thoughts of death or suicide
- Thoughts of hurting other people

Source: Compiled by authors.

Table 16.2.
Unipolar Depression: Common Behavioral Symptoms

- Crying spells
- Interpersonal friction and confrontation
- Anger attacks or outbursts
- Avoidance of anxiety-provoking situations
- Reduced productivity
- Social withdrawal
- Avoidance of emotional and sexual intimacy
- Reduced leisure-time activities
- Development of rituals or compulsions
- Workaholic behaviors
- Substance use and abuse
- Self-sacrifice and victimization
- Self-cutting and mutilation
- Suicide attempts or gestures
- Violent or assaultive behaviors

Source: Compiled by authors.

Table 16.3.
Unipolar Depression: Common Physical Symptoms

- Fatigue
- Leaden feelings in arms or legs
- Sleeping too little (insomnia)
- Sleeping too much (hypersomnia)
- Decreased appetite
- Weight loss
- Increased appetite
- Weight gain
- Sexual arousal difficulties
- Erectile dysfunction
- Delayed orgasm or inability to achieve orgasm
- Pains and aches
- Headaches
- Muscle tension
- Gastrointestinal upset
- Heart palpitations
- Burning or tingling sensations

Source: Kendler KS, and Gardner CO. Boundaries of major depression: an evaluation of DSM-IV criteria. *Am J Psychiatry.* 1998;155:172–177.

The diagnosis of depressive disorders may be complicated by the fact that patients with certain medical diseases such as diabetes may present with physical symptoms resembling those of depression, for example, fatigue. More than 70 percent of patients with depression have somatic presentation, which are less likely to be recognized.[7, 8] In addition to being aware of clinical symptoms that may mask symptoms of depression, practitioners need to be aware of the societal stigma of

depression and of the reluctance of some patients to report psychological distress. Many patients tend to focus on their physical symptoms rather than on their mental state. This bias may vary in relation to the cultural and ethnical background of patients, but stigma may account for up to 45 percent of people failing to report emotional problems to their doctors.[9, 10] Depressive disorders are quite important from a public health perspective, as they are common and often significantly impair psychosocial functioning. Effective interventions are available if a diagnosis is made in a timely fashion.

THE EPIDEMIOLOGY OF DEPRESSION

Numerous epidemiological studies have been conducted to assess the incidence and prevalence of depression. The Epidemiological Catchment Area Study and the National Comorbidity Survey Study Replication (NCS-R) clearly showed that major depression is a prevalent disorder in the general population with 12-month prevalence ranging from 2.3 percent to 6.7 percent in the population.[11, 12] Lifetime prevalence of major depressive disorder in both the United States and in Western Europe is estimated to be between 13.4 percent and 16.6 percent in the general population.[13] A word of caution must be offered when comparing such statistics among nations. Cultural bias and differing attitudes toward mental illness often can influence the statistics. A World Health Organization (WHO) report estimates the 12-month prevalence of depression in developed countries ranges from 3.1 percent in Japan to 9.6 percent in the United States.[14] These statistics are most certainly an underestimate of the condition as those with minor depression are probably not included.

Morbidity, Disability, and Mortality

Major depressive episodes also have been characterized by a significant burden of subjective suffering, increased morbidity, and impaired social and work functioning. It is well known that depression contributes significantly to the burden of medical illness.[15, 16] Depression has a negative effect on physical functioning compared with chronic physical diseases, for example. A WHO world health survey found that depression, compared with chronic diseases such as angina, arthritis, asthma, and diabetes, produces the greatest decrement in health.[17] Among the acutely ill-hospitalized older population, those with more symptoms of depression tend to do poorly and are less likely to improve during and after hospitalization compared with those who do not have symptoms of depression.[18, 19]

A 2000 WHO report ranked depression as the fourth medical condition with the greatest disease burden worldwide as measured in Disability-Adjusted Life Years (DALYs), which express years of life lost to premature death, and years lived with a disability (YLDs) of specified severity and duration. It is estimated that major depression accounts for 4.5 percent of total DALYs and for 12.1 percent of total YLDs, and depressive disorders constitute a large proportion in the global burden of disease, both in developed and developing countries. By 2020, unipolar depression, for example, is projected to be the second leading cause of disability worldwide and the leading cause of disability in the United States.[20] Many studies of patients with depressive disorders reported some form of disability, especially in the month preceding the referral to a physician. The degree of disability almost always is found

to be directly related to severity of depression.[21] Although appropriate antidepressant therapy can improve the daily functioning and overall health of patients with depressive disorders, evidence shows that patients treated for depression still represent a population with significant disability as measured by ability to function in a work setting. Depression is directly linked to high levels of absenteeism, loss of productivity at work, and high workforce turnover.[21–25]

Mortality and major depressive episodes have been shown to be linked. That is, the greater the depression, the greater the likelihood of premature mortality.[26–28] In particular, depression is an important risk factor for suicide. Whereas in the general population, suicides account for about 0.9 percent of all deaths, among those with depressive disorders, the rate of attempted suicide is estimated to be 21 percent.[29] Other complications of depressive disorders are commonly encountered in clinical practice, including homicidal and aggressive behavior and drug addiction.[30, 31]

Depressive conditions result in substantial disability as well as direct and indirect costs. Depression is associated with reduced productivity both at paid and unpaid work and increased work loss. Greenberg et al. [32] estimated the annual societal costs of clinical depression at $83.1 billion (in 2000 dollars), while Kessler et al. [33] estimated that MDD resulted in 27.2 lost workdays per ill worker each year. Moreover, they showed that MDD affected the "work performance," which can be several times the costs of work loss time. Overall, numerous studies have shown that medical costs are higher among those suffering from depression compared with those who are not.[34–38]

Impact of Depression on Medical Comorbidity

The association of depression and medical comorbidity is particularly important. The prevalence of chronic medical conditions, such as heart problems, diabetes, and cancer, in depressed patients is high.[39] A bidirectional relationship exists between depression and medical illnesses, each having a negative impact on prognosis and treatment of the other condition.[40] Several studies have shown that depression significantly influences the course and outcome of concomitant medical diseases. For example, some degree of depression in patients hospitalized for coronary artery disease was shown to be associated with an increased risk of mortality as well as with continuing depression over at least the first year following hospital discharge.[41] Depression post–myocardial infarction (MI) occurs in as many as 27 percent of patients and has a negative impact on recovery.[42] Post-MI patients who suffer from depression have a four- to sixfold increase in mortality rate compared with those without depression.[43] A follow-up study showed that the post-MI mortality risk in depressed patients is still increased by 3.5-fold five years after the MI.[44]

Similar findings have been shown for stroke patients. Pooled data from large studies conducted around the world have shown prevalence rates for major depression of 19 to 23.3 percent among stroke patients.[45] In the large North East Melbourne Stroke Incidence Study, depression and anxiety were significant determinants of persistent handicap after a stroke.[46]

Among patients with diabetes, Anderson et al. [47] reported that the mean prevalence of depression among diabetic patients was 14 percent (ranging from 9 to 27 percent), and the odds of depression in the diabetes group were twice that of the

nondiabetes group. In addition, the occurrence of depression has been associated with a significantly higher risk of diabetes-specific complications, such as diabetic retinopathy, nephropathy, neuropathy, macrovascular complications, and sexual dysfunction.[48, 49] Depression symptom severity also is associated with poor diet and medication regimen adherence, functional impairment, and higher health care costs in primary care diabetic patients.[50, 51]

Depression is the most common psychiatric condition occurring in approximately one-quarter of those with advanced cancer.[52] In particular, a higher prevalence of depression has been described in patients who have more severe diseases and a higher burden of symptoms such as pain and fatigue.[53] Patients with cancer and comorbid depression are at higher risk of mortality as well as longer hospital stays.[54, 55] Depression in cancer patients not only complicates coping with cancer and adherence to medical treatment, but also may affect aspects of endocrine and immune function that plausibly affect resistance to tumor progression.[53]

Underrecognition and undertreatment of depression in elderly patients has been associated with increased medical utilization.[56] Among the elderly age 65 years or older, a correlation exists between depression and the risk of recurrent falls (odds ratio of 3.9). These data are of particular importance because falls in the elderly population are a well-recognized medical and public health problem.[57]

DEPRESSION AND PRIMARY CARE

Clinicians, and especially PCPs, should be aware that the prevalence of depressive disorders among their patient population is substantial and often overlooked.[58] Depression is second only to hypertension as the most common chronic condition encountered in general medical practice.[59] Yet, the rate of missed diagnoses of depression approaches 50 percent of cases.[60] A recent study found that only about half of all Americans with depression receive any treatment and just one in five get the care that conforms to guidelines from the American Psychiatric Association.[61] Furthermore, this study, using the Collaborative Psychiatric Epidemiology Surveys (CPES) data, found that within the United States, depression is a leading cause of disability among major ethnic and racial groups (Mexican Americans and African Americans, in particular). Thus, it is crucial to provide clinicians and especially PCPs with information concerning risk factors for depression.

Risk Factors of Depression in Primary Care

Four risk factors consistently have been associated with depression, and the level of evidence suggests that at least some of the association is indeed causal: gender, stressful life events, adverse childhood experiences, and certain personality traits. The National Comorbidity Study–Replication Study found that women had a significantly higher risk than men of mood disorders (odds ratio of 1.5).[62] A wide range of environmental adversities, such as job loss, marital difficulties, major health problems, and loss of close personal relationships usually are associated with an increased risk for the onset of depression. A range of difficulties in childhood, including physical and sexual abuse, poor parent-child relationships, and parental discord and divorce also have been shown to increase the risk for depression later in life.[63] A family history of depression is another risk factor, as depressive first-degree relatives increase by three- to fourfold the risk of being affected by major depressive disorders.[64]

Associated Features of Depression in Primary Care

PCPs should consider some associated features of depression, as the presence of such features may increase the likelihood of a diagnosis of current MDD. For example, a prior history of depression may be significant; a substantial proportion of patients have their first episode during childhood or adolescence and the risk of recurrence is greater than 50 percent after a first episode of major depression.[65] Numerous previous depressive episodes are an even stronger predictor of recurrence of depression, with 70 percent and 90 percent of patients having recurrences of depression after having experienced two and three episodes, respectively.[66]

Indication for Referral to Psychiatrists

PCPs always should inquire about suicidal thoughts because suicide is one of the most serious complications of depressive disorders. Typically, generic questions such as "have you been thinking lately that life is not worth living?" are appropriate to ask as well as "have you also been thinking that you would be better off dead?" and, finally, "have you considered suicide lately?" and "have you tried?"[67] In the event the patient reports suicidal thoughts or intent, referral to a specialist or to the local psychiatric emergency facility (when appropriate) is strongly recommended. History of mania (elevated mood, increased energy, and impulsivity), suggested perhaps by a history of uncharacteristic behaviors, buying sprees, and excessive risk-taking behavior, often reported by family members, should be an indication for referral to a psychiatric specialist. Psychotic features (hallucinations or delusions) and substance abuse also indicate the need of referral to a psychiatrist. Referral also is indicated in the case of resistant depression and whenever there is a danger that the patient might harm someone else.[68]

Management of Depression

There are two main modalities of depression care that have been shown to be efficacious and well tolerated: pharmacotherapy and psychotherapy. Yet, most Americans who meet diagnostic criteria for major depression are either untreated or undertreated. In particular, several studies indicate that depression is undertreated in primary care settings, as only 21.2 percent of depressed patients are prescribed appropriate antidepressant medications.[69, 70] Michoulon et al. [71] showed that even when PCPs are informed that one of their patients suffers from MDD, inadequate treatment or nontreatment were quite common occurrences. Furthermore, when a pharmacological treatment is prescribed, the medication may be inappropriate. One study found that approximately one-fifth of the patients who were prescribed treatment were prescribed the wrong medication—that is, an anxiolytic for major depression.[69] Selective serotonin-reuptake inhibitors (SSRIs) (Citalopram, Escitalopram, Fluoxetine, Fluvoxamine, Sertraline, and Paroxetine) generally are considered the first-line therapy of antidepressant pharmacotherapy. SSRIs cause fewer side effects (such as dry mouth, blurry vision, constipation, and lightheadedness), they are much safer in overdose, and their use is related to a lower rate of attrition because of intolerable side effects.[72–74] SSRIs have now become the first-line treatment for depression in the elderly because of their favorable side-effect profiles and low risk of complications after an overdose.[75–76] In elderly patients, SSRIs should be titrated slowly to recommended therapeutic doses and

should be used cautiously because other agents are known to have the potential for drug interactions.[77] In children and adolescents, however, a recent meta-analysis by Hetrick et al. [78] showed that there is limited evidence of the efficacy of SSRIs. Fluoxetine was the only SSRI for which there was consistent proof of its effectiveness in improving "response" and depressive disorder symptoms in children and adolescents compared with placebo.

A particular aspect of the management of depressed patients in primary care is related to the unhealthy lifestyles and behaviors associated with depression. For instance, depressed individuals are more likely than nondepressed subjects to engage in smoking, excessive alcohol intake, physical inactivity, and unhealthy eating habits.[79–80] Adherence is another issue that must be considered in treating the depressed patient. Evidence shows that almost half of those who receive an initial prescription for an antidepressant discontinue treatment within the first month.[81] It is important to acknowledge and be aware that medically ill depressed patients tend to be more resistant to treatment, respond slower or less well to antidepressants, and have higher rates of depressive relapse in the continuation phase.[82–85]

PUBLIC HEALTH POLICY AND DEPRESSION

The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression emphasizes three levels at which relative deficiencies occur with respect to the recognition and treatment of depressive disorders: patient, provider, and health care system levels.[70] With respect to patient and family, recognition of depressive disorders, attitude toward treatment and education programs, and media initiatives such as the National Screening Day for Depression typically are considered to be quite useful, as are Internet resources, but more research needs to be conducted to empirically test this theory. With respect to recognition and treatment of depressive disorders by PCPs, it is unclear whether screenings in the primary care setting should be adopted routinely for depression. Screening alone does not improve outcomes for patients with unrecognized depression.[86] Conversely, screening may be useful when it is combined with additional programs aimed at providing support, such as those involving case management and collaborative care models.[87, 88]

Mental health care systems sometimes create barriers to the delivery of optimal treatment. In fact, health maintenance organizations may limit the access to appropriate specialty care specialists and to better tolerated treatments. Policy makers have been struggling for years with issues of cost containment in health care and the need to provide adequate, effective, and well-tolerated treatment for depression. Further work certainly is needed at a health care system level to improve the delivery of care for this important and highly disabling condition.

CONCLUSION

Depression must be viewed as an important and often-overlooked disease that affects a large proportion of the population. Indeed, those suffering from some form of depression probably are higher than that reported in surveys and studies. Often

patients' reporting of their symptoms glosses over or ignores the underlying causes of their symptoms. Yet, undertreatment of depression often can complicate or worsen treatment of other chronic diseases. The cost of depression, both medical and economic, is substantial. Encouraging physicians, particularly PCPs, to be more attuned to the possibility of depression in their patients, especially the elderly and those with multiple chronic diseases, is extremely important. Safe, well-tolerated, and effective antidepressant treatments are available and, if used appropriately, undoubtedly would improve patient outcomes as well as contain health care costs.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
2. Fava GA, et al. Prodromal symptoms in primary major depressive disorder. *J Affect Disord*. 1990;19:149–152.
3. Perlis RH, et al. Prevalence and clinical correlates of irritability in major depressive disorder: a preliminary report from the Sequenced Treatment Alternatives to Relieve Depression study. *J Clin Psychiatry*. 2005;66:159–166.
4. Fava M, et al. The importance of irritability as a symptom of major depressive disorder: results from the National Comorbidity Survey Replication. *Mol Psychiatry*. 2009;Epub ahead of print.
5. Kendler KS, Gardner CO. Boundaries of major depression: an evaluation of DSM-IV criteria. *Am J Psychiatry*. 1998;155:172–177.
6. Judd LL, Akiskal HS. Delineating the longitudinal structure of depressive illness: beyond clinical subtypes and duration thresholds. *Pharmacopsychiatry*. 2000;33:3–7.
7. Bridges KW, Goldberg DP. Somatic presentation of DSM III psychiatric disorders in primary care. *J Psychosom Res*. 1985;29(6):563–569.
8. Kirmayer LJ, Robbins JM, Dworkind M, Yaffe MJ. Somatization and the recognition of depression and anxiety in primary care. *Am J Psychiatry*. 1993;150(5):734–741.
9. Cape J, McCulloch Y. Patients' reasons for not presenting emotional problems in general practice consultations. *Br J Gen Pract*. 1999;49(448):875–879.
10. Chowdhury AN, et al. Prominence of symptoms and level of stigma among depressed patients in Calcutta. *J Indian Med Assoc*. 2001;99(1):20–23.
11. Weissman MM, et al. Affective disorders. In: Robins LN, Regier DA, eds. *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*. New York: The Free Press; 1991.
12. Kessler RC, et al. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62(6):617–627.
13. European Study of the Epidemiology of Mental Disorders 2000 Investigators Project. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders project. *Acta Psychiatr Scand Suppl*. 2004;420:21–27.
14. Nandi A, Beard JR, Galea S. Epidemiologic heterogeneity of common mood and anxiety disorders over the lifecourse in the general population: a systematic review. *BMC Psychiatry*. 2009;9:31.
15. Berndt ER et al. Lost human capital from early-onset chronic depression. *Am J Psychiatry* 2000;157:940–947.

16. Greden JF. The burden of recurrent depression: causes, consequences, and future prospects. *J Clin Psychiatry*. 2001;62(Suppl 22):5–9.
17. Moussavi S, et al. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet*. 2007;370(9590):851–858.
18. Covinsky KE, et al. Relation between symptoms of depression and health status outcomes in acutely ill hospitalized older persons. *Ann Intern Med*. 1997;126(6):417–425.
19. Noël PH, et al. Depression and comorbid illness in elderly primary care patients: impact on multiple domains of health status and well-being. *Ann Fam Med*. 2004;2(6):555–562.
20. Ustün TB, et al. Global burden of depressive disorders in the year 2000. *Br J Psychiatry*. 2004;184:386–392.
21. European Study of the Epidemiology of Mental Disorders 2000 Investigators Project. Disability and quality of life impact of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders project. *Acta Psychiatr Scand Suppl*. 2004;(420):38–46.
22. Coulehan JL, et al. Treating depressed primary care patients improves their physical, mental, and social functioning. *Arch Intern Med*. 1997;157:1113–1120.
23. Katzelnick DJ, et al. Randomized trial of a depression management program in high utilizers of medical care. *Arch Fam Med*. 2000;9:345–351.
24. Wang PS, Simon G, Kessler RC. The economic burden of depression and the cost-effectiveness of treatment. *Int J Methods Psychiatr Res*. 2003;12(1):22–33.
25. Nutt DJ, et al. A Consensus statement on the benefit to the community of European Study of the Epidemiology of Mental Disorders survey data on depression and anxiety. *J Clin Psychiatry*. 2007;68(Suppl 2):42–48.
26. Zheng D, et al. Major depression and all-cause mortality among white adults in the United States. *Ann Epidemiol*. 1997;7(3):213–218.
27. Bruce ML, et al. Psychiatric status and 9-year mortality data in the New Haven Epidemiologic Catchment Area study. *Am J Psychiatry*. 1994;151:716–721.
28. Penninx BW, et al. Minor and major depression and the risk of death in older persons. *Arch Gen Psychiatry*. 1999;56(10):889–895.
29. Sartorius N. The economic and social burden of depression. *J Clin Psychiatry*. 2001;62(Suppl 15):8–11.
30. Marzuk PM, Tardiff K, Hirsch CS. The epidemiology of murder-suicide. *JAMA*. 1992;267(23):3179–3183.
31. Nunes EV, et al. The management of treatment resistance in depressed patients with substance use disorders. *Psychiatr Clin North Am*. 1996;19(2):311–327.
32. Greenberg PE, et al. The economic burden of depression in the United States: how did it change between 1990 and 2000? *J Clin Psychiatry*. 2003;64:1465–1475.
33. Kessler RC, et al. Prevalence and effects of mood disorders on work performance in a nationally representative sample of US workers. *Am J Psychiatry*. 2006;163:1561–1568.
34. Simon GE, Von Korff M, Barlow W. Healthcare costs of primary-care patients with recognized depression. *Arch Gen Psychiatry*. 1995;52:850–856.
35. Henk HJ, et al. Medical costs attributed to depression among patients with a history of high medical expenses in a health maintenance organization. *Arch Gen Psychiatry*. 1996;53:899–904.
36. Unutzer J, et al. Depressive symptoms and the cost of health services in HMO patients age 65 years and older: a 4-year prospective study. *JAMA*. 1997;277:1618–1623.
37. Druss BG, Rohrbaugh RM, Rosenheck RA. Depressive symptoms and health costs in older medical patients. *Am J Psychiatry*. 1999;156:477–479.

38. Thomas MR, et al. Prevalence of psychiatric disorders and costs of care among adult enrollees in a Medicaid HMO. *Psychiatr Serv*. 2005;56:1394–1401.
39. Stein MB, et al. Does co-morbid depressive illness magnify the impact of chronic physical illness? a population-based perspective. *Psychol Med*. 2006;36(5): 587–596.
40. Krishnan KR, et al. Comorbidity of depression with other medical diseases in the elderly. *Biol Psychiatry*. 2002;52(6):559–588.
41. Frasure-Smith N, Lespérance F. Depression and cardiac risk: present status and future directions. *Heart*. 2009;Epub ahead of print.
42. Rudisch B, Nemeroff CB. Epidemiology of comorbid coronary artery disease and depression. *Biol Psychiatry*. 2003;54(3):227–240.
43. Goodnick PJ, Hernandez M. Treatment of depression in comorbid medical illness. *Expert Opin Pharmacother*. 2000;1(7):1367–1384.
44. Lesperance F, et al. Five-year risk of cardiac mortality in relation to initial severity and one-year changes in depression symptoms after myocardial infarction. *Circulation*. 2002;105(9):1049–1053.
45. Robinson RG. Poststroke depression: prevalence, diagnosis, treatment, and disease progression. *Biol Psychiatry*. 2003;54(3):376–387.
46. Sturm JW, et al. Quality of life after stroke: the North East Melbourne Stroke Incidence Study. *Stroke*. 2004;35(10):2340–2345.
47. Anderson RJ, et al. The prevalence of comorbid depression in adults with diabetes: a meta-analysis. *Diabetes Care*. 2001;24(6):1069–1078.
48. de Groot M, et al. Association of depression and diabetes complications: a meta-analysis. *Psychosom Med*. 2001;63(4):619–630.
49. Ludman EJ, et al. Depression and diabetes symptom burden. *Gen Hosp Psychiatry*. 2004;26(6):430–436.
50. Ciechanowski PS, Katon WJ, Russo JE. Depression and diabetes: impact of depressive symptoms on adherence, function, and costs. *Arch Intern Med*. 2000;160(21): 3278–3285.
51. Katon W, et al. Behavioral and clinical factors associated with depression among individuals with diabetes. *Diabetes Care*. 2004;27(4):914–920.
52. Lloyd-Williams M, Dennis M, Taylor F. A prospective study to determine the association between physical symptoms and depression in patients with advanced cancer. *Palliat Med*. 2004;18(6):558–563.
53. Spiegel D, Giese-Davis J. Depression and cancer: mechanisms and disease progression. *Biol Psychiatry*. 2003;54(3):269–282.
54. Pirl WF, Roth AJ. Diagnosis and treatment of depression in cancer patients. *Oncology*. 1999;13(9):1293–1301.
55. Brown KW, et al. Psychological distress and cancer survival: a follow-up 10 years after diagnosis. *Psychosom Med*. 2003;65(4):636–643.
56. Reynolds CF, et al. Chronic depression in the elderly: approaches for prevention. *Drugs Aging*. 2001;18(7):507–514.
57. Gostynski M, et al. Dementia, depression and activity of daily living as risk factors for falls in elderly patients. *Soz Praventivmed*. 2001;46(2):123–130.
58. Lecrubier Y. Prescribing patterns for depression and anxiety worldwide. *J Clin Psychiatry*. 2001;62(Suppl 13):31–36.
59. Wells KB, et al. *Caring for Depression*. Cambridge, MA: Harvard University Press; 1996.
60. Lecrubier Y. Widespread underrecognition and undertreatment of anxiety and mood disorders: results from 3 European studies. *J Clin Psychiatry*. 2007;68(Suppl 2):36–41.
61. Gonzalez HM, et al. Depression care in the United States: too little for too few. *Arch Gen Psych*. 2010;67(37–46).

62. Fava M, et al. The importance of irritability as a symptom of major depressive disorder: results from the National Comorbidity Survey Replication. *Mol Psychiatry*. 2009;Epub ahead of print.
63. Fava M, Kendler K. Major depressive disorder. *Neuron*. 2000;28:335–341.
64. Husain MM, et al. Family history of depression and therapeutic outcome: findings from STAR*D. *J Clin Psychiatry*. 2009;70(2):185–195.
65. Fava GA, Park SK, Sonino N. Treatment of recurrent depression. *Expert Rev Neurother*. 2006;6(11):1735–1740.
66. Kupfer DJ. Long-term treatment of depression. *J Clin Psychiatry*. 1991;(Suppl 52): 28–34.
67. Miller MC, Paulsen RH. Suicide assessment in the primary care setting. In: Jacobs DG, ed. *The Harvard Medical School Guide to Suicide Assessment and Intervention*. San Francisco, CA: Jossey-Bass; 1999:520–539.
68. Whooley MA, Simon GE. Managing depression in medical outpatients. *N Engl J Med*. 2000; 343(26):1942–1950.
69. European Study of the Epidemiology of Mental Disorders 2000 Investigators Project. Psychotropic drug utilization in Europe: results from the European Study of the Epidemiology of Mental Disorders project. *Acta Psychiatr Scand Suppl*. 2004;(420):55–64.
70. Hirschfeld RM, et al. The National Depressive and Manic-Depressive Association consensus statement on the undertreatment of depression. *JAMA*. 1997;277(4): 333–340.
71. Mischoulon D, et al. Management of major depression in the primary care setting. *Psychother Psychosom*. 2001;70:103–107.
72. American Psychiatric Association. Practice guideline for the treatment of patients with major depressive disorder. *Am J Psychiatry*. 2000;157:1–45.
73. Bauer M, et al. World Federation of Societies of Biological Psychiatry guidelines for biological treatment of unipolar depressive disorders in primary care. *World J Biol Psychiatry*. 2007;8:67–104.
74. Thase ME, Kupfer DJ. Recent developments in the pharmacotherapy of mood disorders. *J Consult Clin Psychol*. 1996;64:646–659.
75. Alexopoulos GS, et al. Pharmacotherapy of depressive disorders in older patients: a summary of the expert consensus guidelines. *J Psychiatr Pract*. 2001;7:361–376.
76. Reynolds CF, et al. Maintenance treatment of major depression in old age. *N Engl J Med*. 2006;354(11):1130–1138.
77. Draper B, Berman K. Tolerability of selective serotonin reuptake inhibitors: issues relevant to the elderly. *Drugs Aging*. 2008;25(6):501–519.
78. Hetrick S, et al. Selective serotonin reuptake inhibitors for depressive disorders in children and adolescents. *Cochrane Database Syst Rev*. 2007;(3):CD004851.
79. Aneshensel C, Huba G. Depression, alcohol use, and smoking over one year: a four-wave longitudinal causal model. *J Abnorm Psychol*. 1983;92:1134–1150.
80. Stephens T. Physical activity and mental health in the United States and Canada: evidence from four population surveys. *Prev Med*. 1988;17:35–47.
81. Lesperance F, Frasure-Smith N. Depression in patients with cardiac disease: a practical. *J Psychosom Res*. 2000;48(4–5):379–391.
82. Simon GE, Von Korff M, Lin E. Clinical and functional outcomes of depression treatment in patients with and without chronic medical illness. *Psychol Med*. 2005;35(2):271–279.
83. Iosifescu DV, et al. Comorbid medical illness and relapse of major depressive disorder in the continuation phase of treatment. *Psychosomatics*. 2004;45:419–425.

84. Oslin DW, et al. Association between medical comorbidity and treatment outcomes in late-life depression. *J Am Geriatr Soc*. 2002;50(5):823–828.
85. Koike AK, Unutzer J, Wells KB. Improving the care for depression in patients with comorbid medical illness. *Am J Psychiatry*. 2002;159(10):1738–1745.
86. Callahan CM, et al. Improving treatment of late life depression in primary care: a randomized clinical trial. *J Am Geriatr Soc*. 1994;42:839–846.
87. Wells KB, et al. Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA*. 2000;283:212–220.
88. Katon W, et al. Stepped collaborative care for primary care patients with persistent symptoms of depression: a randomized trial. *Arch Gen Psychiatry*. 1999;56:1109–1115.

This page intentionally left blank

CHAPTER 17

A Public Health Approach to Preventing Suicide

Catherine W. Barber, MPA, and Matthew Miller, MD, MPH, ScD

Robert, a 20-year-old with a history of drug problems, moved back in with his parents after his girlfriend broke up with him. When he stopped going to work, his parents contacted a mental health center and urged him to see a counselor. He refused. He called his girlfriend to reconcile, but she wouldn't speak to him. He felt desperate and went to his father's gun cabinet where he knew he'd find a loaded gun . . .

The World Health Organization (WHO) estimates that 1 million people die from suicide a year, about the same number as die from war and homicide combined.[1] Suicide decedents come from all segments of society, occupations, and educational backgrounds, from rural areas and densely populated cities, during financial crises and bear markets, in wartime and in peace. The final fatal act itself encompasses a panoply of methods, the distribution of which differs across geography and time. Suicide decedents can be as different as an Indian farmer in his 30s who swallows a pesticide following a family argument, a young American who turns his gun on himself when cornered by police, a Hungarian woman with major depression who dies by multiple-drug ingestion, or an older man facing health problems and isolation who dies by hanging.

This chapter reviews the salient aspects of the epidemiology of suicide and the challenges posed by a purely clinical approach to its prevention, outlines a complementary public health approach to suicide prevention, and, using Robert's case as a springboard, discusses specific examples of population-based approaches to reducing suicide. The emphasis will be on suicide in the United States, particularly with respect to reducing a suicidal person's access to lethal means of suicide, but also includes information from other countries.

EPIDEMIOLOGY

Based on data from its member states, the WHO estimates that suicide ranks among the three leading causes of death among people ages 15- to 34-years-old

worldwide.[2] Suicide is the second leading cause of death among Americans 40 and under, and the 11th leading cause among Americans of all ages.[3] Male rates surpass female rates in most parts of the world at a ratio of approximately three-to-one or four-to-one, although the ratio is narrower in certain Asian countries, and parts of China have reported slightly higher rates among females.

The overall incidence of suicide among nations supplying data to WHO is 16 per 100,000, approximately 50 percent higher than the rate in the United States of 11 per 100,000 (see table 17.1). Worldwide, suicide rates generally increase with age, particularly among males. In terms of raw numbers, however, suicide is by and large a disease of people in their prime working years: 84 percent of suicide decedents in the United States are under the age of 65. Age-specific patterns vary greatly across nations and even within nations by sex and ethnicity. For example, in the United States, suicide rates peak after age 65 for white males but remain flat after age 50 for females; among Native Americas, suicide rates peak between 15 and 24.

A small number of studies have examined the extent to which official data may underestimate the incidence of suicide. Most have concluded that while suicides are undercounted, the *relative* ranking of nations by official suicide rate would change little if all suicides were captured.[4, 5] These studies, however, largely have been conducted in developed nations where estimates of underreporting are in the 20 to 40 percent range (most at the lower end of the range).[6, 7, 8, 9] Although suicide rates vary across industrial nations (most falling between 10 and 28 per 100,000), the range is far wider among developing nations supplying mortality data to WHO, where reported suicide rates range from near zero in certain northern African and Middle Eastern countries to 50 or more per 100,000 among certain Eastern bloc countries (see table 17.1).

The range even within a geographic region can be arresting. Former Soviet nations report suicide rates ranging from 0.6 to 40.2 per 100,000. In Central Asia, two directly adjacent countries reported suicide rates of 4.7 and 25.9 in the same year, and in South America, bordering nations reported rates of 4.6 and 25.9 in the same year. Rates that vary to that degree are plausible (due to factors such as cultural differences in the acceptability of suicide, the lethality of locally acceptable and available methods, and the availability of effective emergency care), but disparities of such a magnitude raise questions about the extent to which differential misclassification and underreporting contribute to observed differences. Ranking countries by their suicide rate is further complicated by the fact that many nations have no centralized mortality reporting system or do not report their data to WHO, including many countries on the African continent and in parts of the Pacific Island region.

METHODS OF SUICIDE

The leading methods used in suicide vary considerably across countries. In the United States, firearms account for more suicide deaths than all other methods combined (51 percent in 2006). Suffocations (22 percent), poisoning by drugs (14 percent), and poisoning by carbon monoxide and other gases (4 percent) are the other leading methods in U.S. suicide deaths. The distribution of methods used in nonfatal self-harm admitted to the hospital is strikingly different, with poisonings making up about 83 percent of cases (the vast majority of which are medication overdoses),

Table 17.1
Annual Rate of Suicide per 100,000 Population by Country and Sex, Arranged in Descending Order of Total Suicide Rate

Country/Economy	Males	Females	Total	Year
Belarus	63.3	10.3	35.1	2003
Lithuania	55.9	9.1	30.7	2008
Russia	53.9	9.5	30.1	2006
Kazakhstan	46.2	9.0	26.9	2007
Japan	35.8	13.7	24.4	2007
Guyana	33.8	11.6	22.9	2005
Ukraine	40.9	7.0	22.6	2005
Korea	29.6	14.1	21.9	2006
Sri Lanka	N/A	N/A	21.6	1996
Hungary[a]	37.1	8.6	21.5	2008
Latvia	34.1	7.7	19.9	2007
Serbia and Montenegro	28.4	11.1	19.5	2006
Belgium	27.2	9.5	18.2	1999
Finland	28.9	9.0	18.8	2007
Slovenia[a]	30.2	7.9	18.4	2007
Croatia	26.9	9.7	18.0	2006
Switzerland	23.5	11.7	17.5	2006
France	25.5	9.0	17.0	2006
Estonia[a]	29.1	6.2	16.5	2008
Moldova	28.0	4.3	15.7	2007
Austria	24.7	7.0	15.6	2006
Poland	26.8	4.4	15.2	2006
Hong Kong	19.3	11.5	15.2	2006
Uruguay	24.5	6.4	15.1	2001
People's Republic of China (selected areas)	13.0	14.8	13.9	1999
Sweden	19.5	7.1	13.3	2002
Seychelles	N/A	N/A	13.2	1998
Trinidad and Tobago	20.9	4.9	12.8	2000
Slovakia	22.3	3.4	12.6	2005
Romania	21.5	4.0	12.5	2004
New Zealand	18.9	6.3	12.4	2005
Cuba	18.6	6.2	12.4	2004
Suriname	17.8	6.4	12.1	2000
Czech Republic[a]	20.2	4.2	11.8	2008
Bosnia and Herzegovina	20.3	3.3	11.8	1991
Norway	15.7	7.4	11.6	2005
Canada	17.9	5.4	11.6	2005
Iceland	16.2	6.1	11.2	2005
Portugal	17.5	4.9	11.2	2003
United States	17.7	4.5	11.1	2005
Luxembourg	17.7	4.3	11.0	2005
Australia	17.1	4.7	10.9	2003
Denmark[a]	16.0	5.7	10.6	2006
India	12.2	9.1	10.6	1998

(Continued)

Table 17.1
(Continued)

Country/Economy	Males	Females	Total	Year
Chile	17.8	3.1	10.4	2003
Singapore	12.9	7.7	10.3	2006
Bulgaria[a]	15.3	4.7	9.5	2007
Netherlands	12.7	6.0	9.4	2004
Germany[a]	14.9	4.4	9.4	2007
Ireland[a]	14.5	4.2	9.3	2008
Kyrgyzstan	15.3	3.2	9.2	2005
Argentina	14.1	3.5	8.7	2003
Turkmenistan	13.8	3.5	8.6	1998
Mauritius	13.2	3.8	8.5	2005
Zimbabwe	10.6	5.2	7.9	1990
Thailand	12.0	3.8	7.8	2002
Saint Lucia	10.4	5.0	7.7	2002
Belize	13.4	1.6	7.6	2001
Ecuador	10.4	4.0	7.2	2005
Nicaragua	11.1	3.3	7.2	2005
El Salvador	10.3	3.5	6.9	2005
Republic of Macedonia	9.5	4.0	6.8	2003
United Kingdom[a]	10.4	3.2	6.8	2005
Costa Rica	10.6	1.9	6.3	2005
Panama	11.1	1.4	6.3	2003
Israel	10.4	2.1	6.2	2003
Puerto Rico	10.9	1.8	6.2	2002
Spain[a]	9.6	3	6.1	2007
Malta	7.0	4.9	6.0	2004
Colombia	8.9	2.6	5.7	1999
Uzbekistan	8.1	3.0	5.5	2003
Italy[a]	8.4	2.3	5.2	2007
Venezuela	8.4	1.8	5.1	2002
Cyprus[a]	7.0	1.7	4.3	2008
Brazil	6.8	1.9	4.3	2002
Mexico	7.0	1.4	4.1	2005
Albania	4.7	3.3	4.0	2003
Bahamas, The	6.0	1.3	3.6	2000
St. Vincent and the Grenadines	6.8	0	3.4	2003
Bahrain	4.9	0.5	3.1	1988
Paraguay	4.5	1.6	3.1	2003
Greece[a]	4.8	1.0	2.8	2008
Tajikistan	2.9	2.3	2.6	2001
Georgia	3.4	1.1	2.2	2001
Guatemala	3.4	0.9	2.1	2003
Philippines	2.5	1.7	2.1	1993
Kuwait	2.5	1.4	2.0	2002
Armenia	3.2	0.5	1.8	2003
Dominican Republic	2.9	0.6	1.8	2001
Azerbaijan	1.8	0.5	1.1	2002

Table 17.1
(Continued)

Country/Economy	Males	Females	Total	Year
Peru	1.1	0.6	0.9	2000
São Tomé and Príncipe	0	1.8	0.9	1987
Barbados	1.4	0	0.7	2001
Iran	0.3	0.1	0.2	1991
Jamaica	0.3	0	0.1	1990
Syria	0.2	0	0.1	1985
Egypt	0.1	0	0	1987
Antigua and Barbuda	0	0	0	1995
Haiti	0	0	0	2003
St. Kitts and Nevis	0	0	0	1995

Source: Unless otherwise noted, data are from the World Health Organization at http://www.who.int/mental_health/prevention/suicide/country_reports/en/index.html. Data are assembled on Wikkipedia at http://en.wikipedia.org/wiki/List_of_countries_by_suicide_rate#cite_note-19 and were verified and retrieved on January 22, 2010.

Note:

a. Data are from Eurostat, European Commission, at http://epp.eurostat.ec.europa.eu/tgm/refreshTableAction.do?tab=table&plugin=0&pcode=tps00122&language=en and retrieved on January 10, 2010.

sharp instruments accounting for 11 percent, and firearms and suffocations each accounting for about 1 percent.[10] The distribution of methods used in suicidal acts reflects the ready availability and cultural acceptability of the methods chosen. For example, in most of the developing world, pesticides are the leading method of suicide, whereas this method accounts for less than one-tenth of 1 percent of suicides in the United States.

Lethality of suicidal behavior is a function of several factors, including the intrinsic lethality of the method used (that is, its ability to transfer fatally damaging kinetic energy or interfere with vital metabolic functions) and, for some methods, the attempters' technical knowledge about how to use the method most effectively and the care they take to ensure a lethal outcome (for example, using some drugs rather than others, locking doors to prevent rescue).[11, 12, 13, 14, 15] One measure of the lethality of a given method used in suicidal acts is the case fatality ratio (that is, the proportion of suicidal acts that prove fatal). Case fatality ratios vary widely by method, with 85 to 90 percent of self-injuries with a firearm proving fatal, compared with 2 percent or less for overdoses and self-injuries with a sharp instrument.[16, 17, 18, 19] Even within the category of poisoning, case fatality ratios vary considerably, with pesticides and other agricultural chemicals on average proving more lethal than the psychotropic and analgesic medications often used in developed countries. For example, a study in Northern India reported case fatality at 59 percent for aluminum phosphide, a fumigant used to protect grain stores.[20]

For every adult who dies by suicide in a given year, more than 30 others attempt and survive, according to results from the U.S. National Survey on Drug Use and Health, which found that 0.5 percent of adults ages 18 and over report having made a suicide attempt in the past year.[21] The ratio is far higher among youth—about 1,400 attempts for every death—with some 7 percent of high school

students reporting an attempt in the past year, according to the Youth Risk Behavior Survey.[22]

THE SEARCH FOR THE UNDERLYING CAUSES OF SUICIDE

The effort to identify the underlying causes of suicide has, for the past four decades, focused largely on identifying mental illnesses that are disproportionately common among suicide decedents. A consistent finding across dozens of case-control, psychological autopsy studies is that close to 90 percent of decedents had a diagnosable mental health or substance use disorder,[23] with diagnoses of substance-related (40 percent) and affective (34 percent) disorders predominating in North American studies.[24] Psychotic disorders, although a risk factor for suicide, have been identified in less than 5 percent of suicides. Psychiatric comorbidity (having more than one psychiatric disorder) increases suicide risk, and a majority of decedents screen positive for more than one diagnosis in psychological autopsy studies. A history of previous suicide attempts repeatedly has been found to be one of the strongest individual-level predictors that a person will die by suicide (although the majority of completers do *not* have a history of previous attempts).[25] Additional well-established risk factors include a family history of suicide and psychopathology, history of childhood traumas, impulsive and aggressive personality traits, divorce, psychosocial crises (such as intimate partner discord; legal, financial, or employment problems), and certain types of media coverage of suicide.[26] Several epidemiologic studies at both the individual and population level have identified access to lethal means (particularly within the United States, access to firearms) as a strong, independent risk factor for death by suicide, imposing risk above and beyond psychiatric illness and acute stressors.

Risk factors for suicide attempt are much the same as those for completed suicide, with the exception of firearm access, which is a risk factor for completed suicide but not for attempts.[27, 28] Gay or lesbian sexual identity also has been established as a risk factor for attempts, but has not been established as a risk factor for deaths, given a lack of definitive data on sexual identity of decedents.[29]

CONTACT WITH PROVIDERS

Less than one-third (32 percent) of suicide victims have had contact with a mental health professional in the year preceding their death, according to studies done in a variety of developed nations, and only 19 percent had contact in the month preceding death.[30] Contact with primary care providers (although not necessarily for suicidal complaints) is far more common, with more than three-quarters (77 percent) of decedents having visited a professional in the year before death and 45 percent in the month before death.

Although providers may be treating people who go on to take their own lives soon thereafter, it is, unfortunately, far from clear how to take advantage of this potential opportunity. Providers face an empirically daunting challenge identifying the patients likely to be at highest risk for suicide: established risk factors for suicide have high sensitivity (up to 90 percent of suicide decedents screen positive for a diagnosable psychiatric or substance abuse disorder), but they also notoriously have low

specificity (that is, a high false-positive rate: the vast majority of people identified as having psychiatric or substance abuse disorders, or with a history of suicide attempts, will *not* die by suicide).[31, 32] Furthermore, the demographic risk factors for suicide attempt and suicide completion can be virtual polar opposites. For example, in the United States, young females of color are at highest risk for suicide *attempt*,[21, 33] whereas older, white males are at highest risk for completion.[3]

Another challenge to both mental health and primary care providers is their lack of training on how to manage suicidal risk among patients and the absence of a strong research evidence base from which to guide their care. A Web-based survey of master's-level social workers found that while nearly all had worked with a suicidal patient, the majority received fewer than two hours of instruction on suicide intervention.[34] Of perhaps greater concern, training of clinical psychologists and psychiatrists is also deficient. Only half of graduate students in psychology reported having been trained in managing suicidal patients, although 99 percent said they treated a suicidal patient while in training.[35] Only 27 percent of psychiatry residency programs surveyed offered a skills workshop in suicide intervention.[36] Although mental health care has shifted more to the primary care arena in the past decade, most program directors in internal medicine and pediatrics reported that training about suicide is inadequate.[37]

The science of rigorously evaluating the comparative efficacy and effectiveness of interventions for suicidality is still in its infancy. Some interventions, such as the use of dialectical behavioral therapy for suicidality among people with borderline personality disorder [38] and the use of lithium for people with bipolar disorder,[39] have strong evidence of effectiveness in reducing suicidal behavior. Other interventions remain inadequately understood. For example, little research has elucidated whether hospitalization is protective or harmful and, if protective (or harmful), for which types of patients. How to treat major depressive disorder in adolescents is not without controversy. Evidence indicates that remission rates are higher among adolescents who are treated with a combination of antidepressant (fluoxetine) therapy and cognitive behavioral therapy (CBT) than with a placebo or with CBT alone.[40] There also is a twofold increase in the risk of suicidal events associated with antidepressant therapy compared with placebo, according to both the U.S. Food and Drug Administration's (FDA's) analysis of clinical trial data and findings from the Treatment of Adolescent Depression Study.[41] Clinicians must weigh potential harms and benefits when making treatment decisions; however, as of this writing, empirically based guidelines have not been adequately developed.

In spite of these challenges, modest evidence indicates that training primary care providers to improve identification and care of patients with depression is associated with lowered suicide rates. Intensive training efforts in locations as disparate as Gotland, Sweden,[42] and Kiskunhalas, Hungary,[43] have been associated with a decline in suicide rates and an increase in antidepressant prescribing.

APPLYING A PUBLIC HEALTH APPROACH

Improving the clinical care of patients at risk of suicide is an important aspect of suicide prevention, but it is directed only to the individual who has presented for care, usually at a time when problems already have developed. A public health

approach defines prevention more broadly and employs a far wider range of strategies.

The hallmarks of a public health approach are as follows:

1. Data driven: Population-based data are used to better understand the epidemiology and etiology of suicide, to assist in designing interventions, and to test the impact of interventions on rates of suicidal behavior. The goal is to use data, not ideology, to drive program and policy decisions and to create surveillance systems (ongoing data systems) to track the problem.
2. Population based: Change must occur at the population level. The public health approach focuses on the entire community, not just on those who present for care.
3. Interdisciplinary and multilevel: Public health uses a variety of disciplines and approaches (for example, engineering, education, fiscal policy, treatment, regulation, and so on) aimed at potential victims, the social and physical environment, and the agent (suicide method).

The effort to reduce smoking in the United States illustrates the difference between a clinical and a public health approach to disease management. A clinical approach counsels patients not to smoke and uses nicotine patches, medications, and other techniques to treat those who wish to quit. The clinical approach may use data to test the effectiveness of the interventions, but the outcome of interest is the individual's health and smoking behavior. This approach is patient-specific and relies on the individual to seek care for his or her problem. The field of public health embraces the clinical approach, but also uses a variety of other strategies to prevent or reduce smoking, such as legislating restrictions on smoking in public places; issuing regulations to reduce allowable toxins in the product, to limit vending machines, and to restrict television advertising; increasing taxes to reduce demand; engaging youth in protests over certain tobacco company practices; and promoting media campaigns, school curricula, and National Smoke-Out Days.

A public health approach focuses on understanding the distribution of suicide within and between populations. Although we are interested in understanding why certain individuals in Lithuania, for example, die by suicide—and turn to psychiatric epidemiology for insights—we also are interested in understanding why Lithuania as a nation has a far higher rate of suicide than that of its Western European neighbors, or, indeed, why the rate in France is twice that in the United Kingdom—questions that suicidology, psychiatric epidemiology, and public health cannot yet answer fully.

Reasoning from individual-level studies and clinical care paradigms, one would expect national suicide rates to correlate well with national rates of mental illness and indices of economic and social disorder, or, at least more deterministically, with rates of suicidal ideation and suicide attempts. Reasoning thus, one would be surprised by the data. For example, in the United States, a state-level index for serious mental illness (derived from a household survey, the National Survey of Drug Use and Health) is neither correlated with state suicide rates nor hospital-treated self-harm rates.

This disjuncture between the prevalence of risk factors for suicide and rates of suicide at the population level is observed in developing nations among sites participating in the WHO's multisite intervention study on suicidal behaviors (SUPRE-MISS) (see table 17.2).[44] Communities in which suicidal thoughts or attempts are high are not always those in which suicide completions are high. Broadly speaking,

three population-level constructs interact to determine a community's suicide rate: (1) the underlying rate of distress among individuals in the community, (2) the individual and cultural acceptability of self-harm as a solution to that distress, and (3) the lethality of the suicidal act. Each of these three, in turn, is multidetermined. Underlying rates of distress can be influenced by economic conditions, health care systems, levels of social support, family cohesion, genetics, and so on. The acceptability of self-harm to the individual, and the individual's perception that options other than suicide will not resolve the underlying problem, in turn, are affected by cultural and religious values, health care systems, individual problem-solving skills, and so on. Finally, the lethality of suicidal acts is governed by the availability of emergency health care, the attempter's planning and technical knowledge, and, especially, the local availability and cultural acceptability of highly lethal methods of suicide.

When people seek to understand differences in suicide rates across nations, they often focus on differences in the first construct, and sometimes the second. Frequently, however, the answer lies in the third. For example, efforts to explain the far higher rate of suicide among women in Sri Lanka compared with women in the United Kingdom may point to women's presumed lower status or greater poverty in Sri Lanka. However, as suicide researcher Michael Eddleston has pointed out, the rate at which women in the United Kingdom and women in Sri Lanka are seen in the hospital for self-poisoning is about equal.[45] But the U.K. women typically swallow pills, with a case fatality proportion below 0.5 percent, while the Sri Lanka women swallow pesticides, with a case fatality above 7 percent, netting a 15-fold higher yield of suicide completions.

Understanding and preventing suicide at the population level, then, requires an approach that addresses the interaction between the individual and the environment across all three constructs: underlying distress, cultural acceptability of suicide, and lethality of attempts.

THE MOVEMENT TO PREVENT SUICIDE

Finland was the first nation to design a multidisciplinary, national public health strategy to prevent suicide.[46] Between 1950 and 1990, the male suicide rate increased 50 percent to a high of nearly 50 suicides per 100,000 Finnish men, one of the highest in Europe and well above the U.S. rate of 20 per 100,000 men. Rates of mental illness were not higher in Finland than in other countries. In 1987, the state-run Finnish health care system launched a 10-year prevention program rolled out in three phases: study (1987–1988), program planning (1989–1991), and implementation (1992–1997).

During the first phase, all 1,397 suicides that occurred in Finland in 1987 were examined using psychological autopsy techniques. Most decedents had suffered from depression, a personality disorder, alcoholism, schizophrenia, or another mental disorder. Fifty-six percent died on their first attempt. Those with previous attempts received little or no follow-up care after the attempt. Most decedents had experienced a recent negative event such as a divorce, layoff, arrest, or rift with their family. A majority—particularly among the third who were alcohol dependent—had signaled to others their suicidal feelings, and most had visited a doctor within the month before dying. Finally, many decedents were socially isolated, particularly older suicide victims.[47]

Table 17.2
Community Characteristics, Lifetime Suicidal Thoughts, Lifetime Suicide Attempts, and Mean Annual Suicide Rate in Sites Participating in WHO SUPRE-MISS Suicide Intervention Study

	WHO SUPRE-MISS Study Sites							
Characteristics	Hanoi Vietnam	Campinas Brazil	Karaj Iran	Chennai India	Durban S. Africa	Yuncheng China	Tallinn Estonia	Colombo Sri Lanka
Population	2.7 million	1 million	1.2 million	4.2 million	3 million	1.1 million	400,000	642,000
Population density (/km^2)	3,000	1,200	900	24,000	95	70	2,500	17,200
Main religious affiliation	None	Christian	Islam	Hindu	Christian	None	Christian	Buddhist
Unemployment (percent)	8	14	24	10	38	8	10	8
Psychiatrists (per 100k)	2	21	1	3	1	1	13	4
Suicidal thoughts[a]	8.9	18.6	14.1	2.6	25.4	18.5	12.4	7.3
Suicide attempt rate[a]	0.4	3.1	4.2	1.6	3.4	2.4	3.6	2.1
Suicide rate[b]	**1**	**4**	**6**	**17**	**17**	**23**	**34**	**46**

Source: Bertolote J, Fleischmann A, DeLeo D, et al. Suicide attempts, plans, and ideation in culturally diverse sites: the WHO SUPRE-MISS community survey. *Psychol Med.* 2005;35:1457–1465.

Note:

a. Lifetime self-reported serious suicidal thoughts and suicide attempt (community survey via interview).

b. Mean rate per 100,000 population over last 10 years for which data are available.

Based on their study findings, the program developed a multifaceted strategy. All health care providers, mental health clinicians, social workers, police, and religious groups were invited to participate. One hundred thousand agreed—about half of those invited. The goal was to make asking patients about suicide and depression as routine as checking for hypertension or diabetes. To aid in this effort, the campaign produced and widely distributed booklets that discussed how to recognize and respond to those at risk for suicide.

In addition to designing interventions aimed at providers, suicide awareness programs were put in place in schools and church groups (an easier goal to achieve in Finland than in more heterogeneous nations, because Finland has a national public school system and 96 percent membership in the Lutheran church). A coordinated effort was made to work with the media to reduce newspaper coverage that sensationalized or glamorized suicide and to promote coverage that destigmatized seeking help. The Finnish program also trained military personnel to identify and refer soldiers and potential soldiers at risk for suicide. Because of the universal military draft, this includes all 18-year-old Finnish males and, importantly, young men who are removed from military service, whose suicide rates are extraordinarily high. The program also worked with the provincial governments to increase their capacity to address suicide prevention locally.

A central component of the program was the explicit promotion of social support for those at risk for suicide. For example, support groups were organized and maintained for those suffering from depression and alcoholism, a step intended to counterbalance the Finnish cultural tendency toward social reserve and self-reliance.

The suicide rate has declined in the implementation and postimplementation period, and by 2005, male suicide rates had fallen to 28.1, a rate not seen since the 1950s. The extent to which the national program can be credited is unclear, however, because suicide rates have dropped in many Western Europe nations during this period.

Finland's national plan has played a seminal role. It served as the basis for WHO's 1996 publication *Prevention of Suicide: Guidelines for the Formulation and Implementation of National Strategies.*[48] By 2002, among 38 European nations responding to a WHO European region survey, 18 had initiated national suicide prevention initiatives and a number of nations in other parts of the world have begun to follow suit.[49] The U.S. published the *National Strategy for Suicide Prevention* in 2001,[50] the culmination of an effort led by the Surgeon General's Office, the Centers for Disease Control and Prevention's (CDC) injury prevention center, and advocacy groups comprised of people who had lost a loved one to suicide. By 2007, nearly every state in the United States had either published, or was in the process of publishing, a state plan to prevent suicide.[51]

These reports outline a multidisciplinary, multilevel plan to bring down suicide rates through better understanding of the problem (research and data systems), improving access to and quality of mental health and substance abuse treatment, reducing the stigma associated with receiving mental health care, reducing access to lethal means of suicide, increasing lay people's awareness of how to identify and help people at risk for suicide, and developing and evaluating interventions.

Within the United States, infrastructure has been created to support suicide prevention. Data collection systems were established to better characterize the suicide

problem and to track trends over time. In 2002, the CDC established the National Violent Death Reporting System, which collects ongoing, detailed data on suicides and homicides in participating states.[52] Data on self-reported suicidal feelings and attempts is collected in every state on an annual or semiannual basis among youths and adults. Federal funding for local youth suicide prevention efforts began in 2004 under the Garrett Lee Smith Memorial Act. The funding also supports the Suicide Prevention Resource Center, which serves as the official technical assistance center to states and to Garrett Lee Smith grantees in their prevention efforts and maintains a registry of evidence-based practices in suicide prevention. Specific initiatives to develop and test interventions for veterans and active duty military have been a recent area of focus.

CHANGING BEHAVIOR AND CHANGING THE LETHALITY OF BEHAVIOR: LESSONS FROM THE HIGHWAY SAFETY FIELD

At the broadest level, efforts to prevent suicide fall into two general categories:

- Reducing suicidal behavior (by improving access to and quality of mental health treatment and social support, destigmatizing help-seeking, reducing media glorification of suicide, preventing substance abuse, changing cultural values that implicitly embrace suicide, and so on)
- Reducing the *lethality* of suicidal behavior (by reducing a suicidal person's access to suicide methods most likely to prove lethal in an attempt and by improving emergency medical system response)

Most prevention efforts in the United States have focused on reducing suicidal feelings and behavior, an important and intuitively appealing target because most suicidal acts are motivated by mental anguish. Reducing the lethality of the behavior, however, also has an important role to play—a lesson learned from the traffic safety field. During the first half of the 20th century, efforts to reduce motor vehicle fatalities focused on educating people to drive more safely, with disappointing results. It was not until the field shifted its focus to effectively reengineering cars and highways that the traffic fatality rate began to decline in the second half of the 20th century. One of the most consequential and generalizable insights in the field of injury prevention is that it is usually easier and almost always more effective to reengineer the environment in a way that protects people from risky behavior than it is to change the behavior itself. Erecting a locked fence around a pool, for example, is more effective in preventing childhood drowning than educating children and parents about pool safety. It was only with a paradigm shift from educating drivers to reengineering the environment and lowering speed limits that motor vehicle mortality fell. Engineering solutions that proved most effective (like seat belts, air bags, antilock brakes, highway improvements, and crash-worthy car design) did so not by reducing the probability of a *crash* (the motor vehicle analogue of a suicide *attempt*) but by greatly reducing the odds that a crash proved fatal (the analogue of a suicide *death*).[53]

PREVENTING ROBERT'S SUICIDE

We will use Robert's case as a launching point to discuss some ways in which strategies from both categories—reducing suicidal behavior and reducing the lethality

of that behavior—have been applied to the problem of suicide, and we will focus particularly on the latter, which has received less attention in the United States.

Like many suicide decedents, Robert had begun to spiral downward, one negative event leading to the next: the substance abuse led to a relationship breakup, which in turn led to acute emotional distress, housing dislocation, interruption in role performance, and the likely prospect of imminent job loss. In the external environment, at least two protective factors were at play. Robert's parents appeared to be aware that he was in need of help and knew where to call, and a mental health agency was available to offer resources. A critical risk factor also was at play: the presence of guns in his home. These salient issues are examined in turn.

Substance Abuse Prevention

Preventing Robert's substance abuse problem in the first place may have averted his downward plunge. Population-based strategies to reduce alcohol use, for example, have been associated with declines in suicide. Policies in the former Soviet Union,[54] Belarus,[55] and Slovenia [56] that effectively reduced alcohol consumption have been associated with concomitant decreases in the national suicide rate. Gorbachev's major anti-alcohol campaign in the former Soviet Union in 1985, for example, was associated with an immediate plummet in the female suicide rate of 50 percent, comparing 1986 deaths with 1984 deaths.[57] Strategies that have been applied in the former Soviet Union and elsewhere include price increases on alcohol products, purchasing and production limits, greater enforcement of public drunkenness laws, workplace interventions to discourage drinking, and social role models. In the United States, increasing the minimum drinking age from 18 to 21 was associated with modest declines in the youth suicide rate at the state level.[58]

Once Robert developed a substance use problem, traditional treatment approaches were then needed to contend with his illness. Given the high rate at which people with a serious substance abuse problem report having attempted suicide—ranging from 37 to 47 percent according to studies in Australia [59] and Norway [60]—preventing and treating substance disorders is an important target of suicide prevention.

Destigmatizing Help-seeking

The fact that Robert, like many suicide decedents, did not view his situation as a mental health problem in need of treatment poses challenges to relying solely on a traditional treatment approach. In the mid-1990s, suicide was the second leading killer of Air Force personnel, responsible for 24 percent of all deaths. Three-quarters of Air Force suicides had a serious intimate partner problem at the time of death, many were facing criminal or disciplinary actions, and few had received any formal counseling. The U.S. Air Force launched a suicide prevention campaign in the late 1990s with these findings in mind.[61] A central tenet of the campaign was that suicide was not a medical problem but an Air Force problem. The culture in which suicides had been occurring was inimical to help-seeking behaviors that might have saved lives. Therefore, the campaign sought to reframe seeking help from a sign of weakness to a sign of strength. Air Force personnel at all levels were trained in suicide prevention and "buddy care" (peers looking out for peers). Supervisory officers were trained to recognize and respond to distress among their supervisees with

appropriate referrals and support. A variety of legal and logistical barriers to mental health treatment and social supports were removed.

As in Robert's case, many people at risk for suicide do not perceive themselves as having a mental health problem but instead are focused on the triggering events that preceded their suicidal crisis, such as a divorce or arrest. The Air Force took steps to make family advocacy services available before problems became acute, mandated supervisory support in the period following arrest or other disciplinary action, and provided trauma response teams following traumatic incidents. Suicides declined by about one-third in the postimplementation period. Fatal accidents, serious family violence, and homicide also declined.

Interventions in other settings have attempted to destigmatize help-seeking. For example, recognizing that police officers often feel uncomfortable using professional mental health services, some police departments also make chaplain services and peer support services available.[62] The "Athabaskan" Native American suicide prevention program reduced barriers to care by using peers and natural leaders, under the supervision of mental health professionals, to provide support and to bring services directly to the community, for example, at dances and outdoor events.[63]

Gatekeeper Training

A prominent component of many states' suicide prevention efforts is "gatekeeper training," an educational approach that trains informal helpers (for example, parents, supervisors, teachers, co-workers) to recognize risk, to ask the at-risk person whether he or she is thinking about suicide, and to respond appropriately with support and referrals. In Robert's case, the parents reached out for help from a mental health agency, and they learned a piece of information that enabled the case to turn out quite differently from the path Robert had been headed down:

> Robert, a 20-year-old with a history of drug problems, moved back in with his parents after his girlfriend broke up with him. When he stopped going to work, his parents contacted a mental health center and urged him to see a counselor. He refused. He called his girlfriend to reconcile, but she wouldn't speak to him. He felt desperate and went to his father's gun cabinet where he knew he'd find a loaded gun. . . . *But, when he got there, the guns were gone. Robert found a razor and cut his wrists. His parents found him an hour later and brought him to the hospital where he was treated and survived.*

REDUCING A SUICIDAL PERSON'S ACCESS TO LETHAL MEANS: THE LOW-HANGING FRUIT

In Robert's case, the mental health worker who answered his parent's call was trained to ask suicidal patients and their family members about guns at home. She urged the parents to remove the guns until the situation improved or, second best, to lock them securely. Robert survived because he substituted a method that was less lethal—as is nearly every method compared with a firearm.

Among methods commonly used in suicidal acts, there is a hierarchy. Firearms and jumps from very high places are highly lethal and neither affords the ambivalent attempter an opportunity for rescue or to back out once the attempt is initiated. Both are therefore promising targets for means restriction (that is, bridge barriers at popular

jumping sites). Hangings and carbon monoxide poisonings can also be highly lethal (case fatality ratios using hospital and mortality data are in the 40 to 80 percent range),[18, 19] but these methods afford a window of opportunity for backing out or rescue, in which case the attempter may not be injured or be seen in the hospital. A study of hanging suicides found that 50 percent were partial suspension, which means that the attempt can be interrupted by standing upright in the event of a change of heart.[64] Attempts by cutting and overdose are at the bottom of the hierarchy with case fatality below 5 percent. A survey of laypeople and pathologists in the United States found that laypeople consistently overrated the lethality of these two methods, a misperception that is probably responsible for saving many lives.[65]

Means restriction is one of the few low-hanging fruits in suicide prevention. Taking the guns from Robert's house to his uncle's house for safe-keeping was far more easily accomplished than treating his substance abuse, changing his attitudes toward counseling, or restructuring his response to acute rejection.

Truly a Life Saved?

Was Robert's a life truly saved or an eventual suicide only temporarily averted? After all, Robert was heading for a gun, and a suicide attempt is one of the strongest risk factors for suicide completion. In fact, Robert's prognosis is relatively good: 9 out of 10 people who make even a serious suicide attempt do not go on to die by suicide. A review of more than 90 studies that have followed suicide attempters over time found that, on average, 7 percent (ranging from 5 to 11 percent) eventually die by suicide, 23 percent reattempt nonfatally, and 70 percent have no further attempts.[66] Even studies that counted equivocal deaths as suicides, followed attempters for multiple decades, and focused on medically serious attempts (such as jumping in front of a train) found similarly low completion rates.

Crises, Impulsivity, and Deliberation

The relatively good long-term, suicide-free survival rate among prior attempters is consistent with the fact that suicidal crises often are short-lived. Although some suicides are deliberative and involve careful planning, many appear to have an impulsive component. A variety of studies have asked suicide attempters how long a period elapsed between deciding to attempt suicide and initiating the attempt: 24 percent reported less than five minutes, in a study of nearly lethal attempts in Texas.[67] Only 29 percent reported an interval greater than one hour. In an Australian study of emergency department–treated attempts, 40 percent reported less than five minutes.[68] A more recent study of people referred to a psychiatric hospital following an attempt found that 48 percent reported acting on their decision in 10 minutes or less.[69]

Robert's attempt was immediately preceded by his ex-girlfriend's rejection on the telephone. The National Violent Injury Statistics System (which was the pilot version of the National Violent Death Reporting System) gathered data from the death investigation reports that medical examiners, coroners, and police file after a suicide. For more than one-third of youth under 18 and one-fifth of adults under 65, the report mentioned that a crisis such as an arrest, breakup, or family argument occurred within 24 hours of the suicide. Suicides can be prevented if the suicidal

person makes it through these crises safely. Putting time or distance between the impulse to die and lethal suicide methods is one way to do that.

Of course, means restriction does not always work, particularly if the person substitutes another highly lethal method or if the person has a sustained impulse to die by suicide. Robert acted impulsively and angrily, like many young people. But if he had been more like Henry, an older person who prepared his will, set out his bank statements, researched the lethality of the suicide method he settled on, and deliberately waited until no one was nearby before attempting, means restriction may not have worked. Prevention strategies are targeted to specific populations. If a one-size-fits-all solution works, like a vaccine, public health will use it. But given the lack of such a solution, a public health approach works incrementally, chipping away at the problem, finding the strategies that are best suited to specific subpopulations.

Enough suicides have a strong impulsive component, especially among younger people, that restricting access to highly lethal methods has the potential to save more lives than perhaps any other approach we have. In the next section we briefly review some of the salient concepts and studies in support of this strong assertion.

Firearms as a Risk Factor: Case Control Studies

At least 12 individual-level, case control studies have examined the link between firearm access and suicide in the United States, often among different subpopulations. Each has found that a gun in the home is a risk factor for suicide. Five psychological autopsy studies were conducted among adolescents by David Brent and colleagues in the 1980s and 1990s. Adolescents who died by suicide were compared with demographically matched living adolescents using record reviews and interviews with an informant. Control groups varied across each study; some used suicidal psychiatric inpatients who had not died,[70] and others used nonsuicidal psychiatric inpatients,[71] adolescents with a history of substance abuse [72] or affective disorder,[73] or community-dwelling adolescents.[74] In logistic regression controlling for factors such as psychiatric diagnosis, suicidal intent score, and presence of a male in the home, the presence of firearms in the home significantly increased the risk of suicide (odds ratios ranging from 2–5). Odds ratios were highest among adolescents without a psychiatric diagnosis, particularly when the firearm was stored loaded. The Brent studies were all relatively small (30 to 70 per comparator group).

A larger case control study by Kellermann et al. ($n = 438$ matched pairs of cases and controls) examined people of all ages dying by suicide in the home in a rural and an urban county.[75] The presence of a gun was associated with increased suicide risk (adjusted odds ratio, 4.8; 95 percent confidence interval, 2.7–8.5) for both the gun owner and all other members of the household. Odds ratios were greatest for youths, those without a psychiatric disorder, and when guns were stored locked and unloaded.

A number of studies used mortality follow-back data assembled by the National Center for Health Statistics (NCHS). NCHS conducts interviews with a family member or other informants on a sample of death certificates to shed light on the etiology and characteristics of a number of fatal diseases and injury types. The studies found that a firearm in the home elevated the risk of overall suicide and firearm suicide, but not nonfirearm suicide, among adults of all ages,[76] for both males and

females,[77] and for both blacks and whites.[78] A psychological autopsy study conducted among adults ages 50 and over that compared suicide decedents with demographically matched community controls also confirmed that firearm access was a risk factor for suicide, particularly for men.[79] For both the young [80] and the old,[79] suicide risk is higher in cases in which guns are stored loaded and unlocked than when guns are locked away. Ecologic studies at the population level in the United States also have established that in states in which gun ownership is more common, suicide rates are higher, both overall and among firearm suicides but not among nonfirearm suicides.[81]

Maybe It's Not the Guns?

A committee appointed by the National Research Council issued a report in 2004 that concluded that while the association between firearm access and increased suicide risk was clear, less clear was *why* that association exists.[82] They conjectured that firearm access may serve as a proxy for an unmeasured third variable that drives the observed relationship. For example, since firearm ownership is more prevalent in rural areas, perhaps it is rural isolation that leads to an increase in mental health problems and suicidal feelings, and firearm access plays a minimal role.

A number of papers have since addressed this issue and have reported that those living in households with guns are no more likely than those living without guns to screen positive for a mental health problem. Two papers used data from the National Comorbidity Survey, an in-depth mental health survey that includes questions about suicidal thoughts and attempts. People from gun-owning households were no more likely than people from nongun households to report having felt suicidal, and were *less* likely to report having made a suicide attempt.[83, 84] A study of older patients visiting a primary care clinic found patients who have a gun in the home were no more or less likely to have suicidal ideation than were people who live in homes without guns.[85] Gun owners are not more suicidal; they simply are more likely to die if they become suicidal.

The Evidence for Means Restriction

An international panel charged with examining the evidence for suicide prevention strategies concluded in a 2005 paper published in the *Journal of the American Medical Association* that only two strategies currently had sufficient evidence of effectiveness: means restriction and training physicians in depression care.[86] Five points summarize the logic underlying means restriction:

- People who attempt suicide are often ambivalent.
- Many people act impulsively during a suicidal crisis.
- Suicidal crises often are fleeting.
- If a highly lethal method is not readily available and a less lethal method is used instead, the attempt is less likely to prove fatal.
- Fewer than 10 percent of people who make nonfatal attempts go on to die by suicide.

The first large-scale natural experiment that drew attention to the benefits of means restriction has been called the Coal Gas Story.[87, 88] During the 1950s,

nearly half of all suicides in the United Kingdom were attributed to poisonings with domestic gas used in ovens and heaters. With the discovery of a new, less toxic and less expensive source of gas in the North Sea, the carbon monoxide content of domestic gas fell over the next 15 years to near-zero levels. Following this detoxification, rates of suicide by gas fell, in effect to zero (for example, in 1960, there were 2,499 suicides by domestic gas, in 1977 there were eight). Moreover, and critically, overall rates of completed suicide dropped substantially for both men and women (by approximately 30 percent), saving thousands of lives over a 15-year period.

A similar experiment occurred in Sri Lanka, where pesticides are the leading method of suicide. From 1950 to 1995, the suicide rate increased eightfold. In 1995 and 1998, the government banned a number of the most highly toxic pesticides commonly used in agriculture. By 2005, the suicide rate had fallen by half, an astonishing achievement.[89] Secular trends in unemployment, alcohol misuse, divorce, pesticide use, and Sri Lanka's civil war did not appear to be associated with the decline. As suicidologist C. H. Cantor has written, "[T]he most dramatic reductions in suicide rates to date have occurred not as a result of improvements in well-being, but through reductions in availability of certain lethal methods of suicide."[90]

Challenges to Means Restriction in the United States

The pesticides banned in Sri Lanka possess several characteristics that make them good, if not ideal, targets for means restriction. First, they account for a large proportion of the total number of suicides in that country. Second, they frequently result in death when ingested. Third, the method most likely to be substituted for the banned pesticides—in this case, pesticides that were not banned—are less lethal. Fourth, pesticides are not cultural icons. Firearms in the United States share the first three characteristics, but decidedly not the fourth. Indeed the iconic status of firearms in the United States presents formidable, but not insuperable, challenges to adopting measures that would reduce their availability to people during a suicidal crisis.

A 2007 survey with the directors of statewide suicide prevention groups in the 50 U.S. states found that only 9 states had sponsored any programs aimed at reducing a suicidal person's access to firearms.[51] Why so few? Impediments to more broadly adopting means restriction include the assumption that all suicidal acts are far more deliberative and less ambivalent than many are, and the mistaken belief that means restriction refers exclusively to gun control legislation, a political "no-go" for young suicide prevention coalitions wary of partisan controversy.

What saved Robert, however, was not legislation. Rather, it was a counselor who routinely asks about firearm access when a client is at risk of suicide. Most providers do not. For example, a record review found that psychiatric residents at a psychiatric emergency department assessed firearm access in only 3 percent of pediatric patients.[91] Grossman et al. reported that although 80 percent of emergency nurses responding to a survey had recent experience with suicidal adolescents, only 28 percent provided education on means restriction to parents.[92] Another study interviewed parents or other caretakers whose adolescent had deliberately overdosed.[93] Only 12 percent of those with medications at home and none of those with firearms at home reported having received counseling from emergency department personnel about the importance of restricting the adolescent's access to lethal means. Those

who did receive such education were more likely than those who did not to restrict access, a finding that was replicated in another emergency department study.[94]

Making More Cases End Like Robert's

What would it take to ensure that the screening forms that school psychologists use when a youth is referred for possible suicidality include a gun access question? What about including advice about firearm storage in brochures given to family members of veterans facing acute emotional distress? Shouldn't emergency department personnel routinely assess gun access among suicidal patients? Changes of this sort require changes in thinking, policy, and training.

Some of the states that have sponsored means restriction work are training providers to make lethal means counseling a routine component of managing suicidal risk (for example, New Hampshire's Counseling on Access to Lethal Means [CALM] program). Because most people who die by suicide are not in the mental health system, training must extend beyond mental health clinicians to others who come into contact with suicidal people: medical providers, school personnel, employee assistance programs, domestic violence programs, the clergy, probation officers, divorce and defense attorneys, and police. Advocates, institutional directors, and professional associations can play a role in this work by adding lethal means counseling to their recommended protocols, spreading the word about this intervention to their memberships, and sponsoring CALM-type trainings. A source of information on this approach is available at the Means Matter Web site, www.meansmatter.org.

PESTICIDES: A KEY TARGET FOR MEANS RESTRICTION GLOBALLY

Thousands of lives each year, especially in the United States, could be saved by reducing a suicidal person's access to firearms. *Tens of thousands* could be saved by prohibiting the use of pesticides with high human toxicity. Pesticides are the leading method of suicide globally, claiming some 300,000 lives each year, or about 30 percent of all suicides.[95] The proportion is far higher in rural areas of Asia, the Pacific Islands, Africa, and Latin America, and at near-zero levels in developed nations. Case fatality rates vary widely across individual pesticides, ranging from below 5 percent to between 60 and 70 percent.[20] Many farmers keep pesticides in or near their homes; often they are sold in a concentrated form that is later diluted for agricultural use. Ingestion of even small quantities can cause death, but a long interval often occurs between ingestion and death. Interviews during this interval in one study revealed that many of the suicide attempts were impulsive. More than half of the people took the poison after less than 30 minutes of thought (often immediately following an argument), the choice of poison most often was governed by what was immediately available in the home, and those who died were about as likely as those who did not to believe the poison they used would result in death.[12]

When several of the most lethal pesticides were banned from agricultural use in Sri Lanka, for example, the suicide rate dropped by half. Agricultural output did not suffer, and the cost of production did not increase.[96] In Western Samoa, the rise and fall of suicides—but not suicide attempts—closely tracked the introduction and later banning of paraquat on the island.[97] *No other single act could so efficiently and immediately save lives with such little cost as restricting the use of pesticides*

with high human toxicity. Other approaches undertaken in China, Japan, India, and Sri Lanka to reduce access to, or lethality of, pesticides have included adding emetics or chemicals to reduce absorption if ingested, creating solid forms, restricting the sale of concentrated liquids, and installing locked boxes or boxes located near fields and away from homes to reduce the most impulsive acts.

Means Restriction and Other Suicide Methods

In many regions of the world, neither pesticides nor firearms are frequent causes of death, and means restriction work has centered on other methods. For example, in Hong Kong, where there has been an epidemic of charcoal-burning suicides, participating supermarkets kept charcoal in a locked area for sale only on request. In Australia, a cost-benefit analysis was undertaken to examine the use of an air quality monitoring device in motor vehicles to prevent both fatigue and suicide. The device shuts off the engine when toxic levels are detected and the car is not moving, and it alerts the driver and circulates fresh air when it detects low levels and the car is moving. In Singapore, where 90 percent of the population lives in high-rise apartment buildings, and jumping accounts for 70 percent of suicides, efforts are under way to place railings or plastic barriers along balconies and to encourage families of suicidal individuals to secure the windows and prevent access to the balconies. In Hong Kong and urban areas of the Republic of Korea, where jumps in front of subways are problem, the transit systems erected screen barriers in 2004 and 2005 to prevent access to the tracks. When a train arrives, the doors of the train align with the doors in the screen barriers, which then open to allow passengers to board or leave. In many communities, bridge barriers have been erected at suicide hot spots to prevent jumps.[98]

Two highly lethal methods of suicide pose challenges to a means restriction approach because they are widely accessible: suffocation and fire. Suicide by fire is infrequently used in most countries with a few exceptions such as in India and Iran. Fire seems unlikely to grow in acceptability as a suicide method. Suffocation (typically by hanging or ligature), however, is a leading method in many countries. Because the means for hanging cannot be physically controlled in anything other than institutional settings, an alternative is to reduce its acceptability, for example, by discouraging how-to Web sites that promote suffocation as painless and fast, by discouraging "the choking game" (a game played by adolescents for "kicks"), by requesting the media not to cover details about method when reporting individual suicide stories, and by emphasizing the painful and disfiguring aspects of hanging. Any population-level increase in the acceptability of hanging that would result in even a small proportion of attempters switching from medication overdose or cutting to hanging could dramatically increase the population's suicide rate.

Not a Panacea

Reducing access to highly lethal means is not a panacea. It does not always result in large or measurable decreases in the overall suicide rate. For example, restricting a low-lethality method or a method infrequently used may not make a detectable difference in the suicide rate, even if it does save some lives. It is also not intended as a stand-alone intervention because it does not address the underlying despair. At the population level, reducing access to a highly lethal and frequently used

method of suicide is likely to reduce suicide rates (absent any countervailing trends toward an increase overall in suicidal behavior or an increase in the availability or acceptability of another high-lethality method). But for any given individual, reducing access may or may not save his or her life. Reducing access buys time. Time allows the suicidal impulse to pass and crises to diminish in intensity, and it may help psychic wounds to heal on their own. Time also provides an opportunity to mobilize thoughtful treatment and social support to address the psychosocial disruptions that underlie the suicidal feelings.

DO NO HARM

A guiding principle in medicine is, "First, do no harm." The suicide prevention movement, too, has an obligation to assess any harm that may be caused by an intervention—whether direct harm that increases suicidality or distress, or indirect harm that channels precious resources to ineffective interventions. For example, many local suicide prevention initiatives have launched awareness campaigns. Increased media coverage of suicide may be positive if it encourages more people in distress to seek and find help [99]; however, not all coverage is good coverage. A number of studies have found that coverage that focuses on celebrity suicides, discusses a decedent's method in detail, and implicitly glorifies the act as a romantic solution to an intractable problem is associated with temporary increases in suicides.[100, 101]

With respect to means restriction, given the very large numbers of attempts made with pills and sharp instruments, any migration to higher-lethality methods could increase the suicide rate. For this reason, public education campaigns directed to the general public or to at-risk groups should *not* provide specific information about case fatality ratios for suicide methods lest it have the unintended consequence of encouraging some attempters to choose more lethal methods.

CONCLUSION

Although a global problem, the public health approach to suicide prevention is still in its infancy. The tenfold difference in suicide rates across nations and across demographic groups suggests that preventing suicide is not a quixotic goal. Public health strategies—such as changing cultural attitudes, increasing social support, improving access to high-quality treatment, and reducing access to lethal means—can reduce suicide, just as analogous strategies already have reduced the incidence of smoking and motor vehicle crashes. Surveillance systems are being established to track the problem. Many countries have developed national plans that outline comprehensive, multidisciplinary, and culturally appropriate strategies to address the problem. Survivor and advocacy groups have sprung up in many areas to advocate for funding. Research is under way to evaluate both treatment and population-based interdisciplinary approaches to suicide prevention.[102] Nations are collaborating and learning from one another under the leadership of the WHO, the International Association for Suicide Prevention, and other groups. Although funding for interventions and research still falls far short of levels commensurate with the magnitude of the suicide problem, intellectual and institutional support for a public health approach is growing, raising an empirically based hope that suicide incidence will fall in tandem.

REFERENCES

1. Krug EG, et al., eds. *World Report on Violence and Health*. Geneva, Switzerland: World Health Organization; 2002.
2. World Health Organization, Suicide Prevention. Introduction with country reports and charts. http://www.who.int/mental_health/prevention/suicide/country_reports/en/index.html. Accessed January 20, 2010.
3. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. Web-based Injury Statistics Query and Reporting System: leading causes of death report. http://webappa.cdc.gov/sasweb/ncipc/leadcaus.html. Accessed October 30, 2009.
4. Speechley M, Stavraky KM. The adequacy of suicide statistics for use in epidemiology and public health. *Can J Public Health*. 1991;82(1):38–42.
5. Phillips DP, Ruth TE. Adequacy of official suicide statistics for scientific research and public policy. *Suicide Life Threat Behav.* 1993;23(4):307–319.
6. Huusko R, Hirvonen J. The problem of determining the manner of death as suicide or accident in borderline cases. *Z Rechtsmed*. 1988;100(2–3):207–213.
7. Sampson HH, Rutty GN. Under-reporting of suicide in South Yorkshire: a retrospective study of suicide and open verdicts returned by HM Coroner, 1992–1997. *J Clin Forensic Med*. 1999;6(2):72–76.
8. Connolly J, Cullen A. Under-reporting of suicide in an Irish county. *Crisis*. 1995; 16(1):34–38.
9. Allebeck P, Allgulander C, Henningsohn L, Jakobsson SW. Causes of death in a cohort of 50,465 young men: validity of recorded suicide as underlying cause of death. *Scand J Soc Med.* 1991;19(4):242–247.
10. Children's Safety Network Economics and Data Analysis Resource Center, Suicide Prevention Resource Center. *United States Suicide Prevention Fact Sheet*. Newton, MA: Suicide Prevention Resource Center; 2008.
11. Brown GK, Henriques GR, Sosdjan D, Beck AT. Suicide intent and accurate expectations of lethality: predictors of medical lethality of suicide attempts. *J Consult Clin Psychol*. 2004;72(6):1170–1174.
12. Eddleston M, Karunaratne A, Weerakoon M, et al. Choice of poison for intentional self-poisoning in rural Sri Lanka. *Clin Toxicol.* 2006;44(3):283–286.
13. Hamdi E, Amin Y, Mattar T. Clinical correlates of intent in attempted suicide. *Acta Psychiatr Scand*. 1991;83(5):406–411.
14. Plutchik R, van Praag HM, Picard S, Conte HR, Korn M. Is there a relation between the seriousness of suicidal intent and the lethality of the suicide attempt? *Psychiatry Res*. 1988;27:71–79.
15. Swahn MH, Potter LB. Factors associated with the medical severity of suicide attempts in youths and young adults. *Suicide Life Threat Behav*. 2001;32:21–29.
16. Elnour AA, Harrison J. Lethality of suicide methods. *Inj Prev*. 2008;14(1):39–45.
17. Miller M, Azrael D, Hemenway D. The epidemiology of case fatality rates for suicide in the Northeast. *Ann Emerg Med*. 2004;723–730.
18. Spicer RS, Miller TR. Suicide acts in 8 states: incidence and case fatality rates by demographics and method. *Am J Public Health*. 2000;90(12):1885–1891.
19. Vyrostek S, Annest J, Ryan G. Surveillance for fatal and nonfatal injuries: United States, 2001. *MMWR*. 2004;53(SS07):1–57, Table 6.
20. Eddleston M. Patterns and problems of deliberate self-poisoning in the developing world. *QJM*. 2000;93(11):715–731.
21. Substance Abuse and Mental Health Services Administration, Office of Applied Studies. *The NSDUH Report: Suicidal Thoughts and Behaviors among Adults*.

Rockville, MD: Substance Abuse and Mental Health Services Administration; 2009.
22. Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion. Trends in the prevalence of suicide-related behaviors: National Youth Risk Behavior Survey, 1991–2007 factsheet. http://www.cdc.gov/HealthyYouth/yrbs/trends.htm. Accessed December 7, 2009.
23. Cavanagh JTO, Carson AJ, Sharpe M, Lawrie SM. Psychological autopsy studies of suicide: a systematic review. *Psychological Med*. 2003; 33:395–405.
24. Arsenault-Lapierre G, Kim C, Turecki G. Psychiatric diagnoses in 3275 suicides: a meta-analysis. *BMC Psychiatry*. 2004;4:37.
25. Gunnell D, Frankel S. Prevention of suicide: aspirations and evidence. *Br Med J*. 1994;308:1227–1233.
26. Hawton K, van Heeringen K. Suicide. *Lancet.* 2009;373(9672):1372–1381.
27. Ilgen MA, Zivin K, McCammon RJ, Valenstein M. Mental illness, previous suicidality, and access to guns in the United States. *Psychiatr Serv*. 2008;59(2):198–200.
28. Miller M, Barber C, Azrael D, Hemenway D, Molnar BE. Recent psychopathology, suicidal thoughts and suicide attempts in households with and without firearms: findings from the National Comorbidity Study Replication. *Inj Prev*. 2009;15(3):183–187.
29. McDaniel JS, Purcell D, D'Augelli AR. The relationship between sexual orientation and risk for suicide: research findings and future directions for research and prevention. *Suicide Life Threat Behav*. 2001;31(Suppl):84–105.
30. Luoma JB, Martin CE, Pearson JL. Contact with mental health and primary care providers before suicide: a review of the evidence. *Am J Psychiatry*. 2002;159(6): 909–916.
31. Goldstein RB, Black DW, Nasrallah A, Winokur G. The prediction of suicide: sensitivity, specificity, and predictive value of a multivariate model applied to suicide among 1906 patients with affective disorders. *Arch Gen Psychiatry*. 1991;48(5):418–422.
32. Pokorny AD. Suicide prediction revisited. *Suicide Life Threat Behav*. 1993;23(1):1–10.
33. The Substance Abuse and Mental Health Services Administration. Substance Abuse and Suicide Prevention: Evidence and Implications. A White Paper. www.samhsa.gov.
34. Feldman BN, Freedenthal S. Social work education in suicide intervention and prevention: an unmet need? *Suicide Life Threat Behav*. 2006;36:467–480.
35. Dexter-Mazza ET, Freeman KA. Graduate training and the treatment of suicidal clients: the students' perspective. *Suicide Life Threat Behav*. 2003;33(2):211–218.
36. Ellis TE, Dickey TO, Jones EC. Patient suicide in psychiatry residency programs: a national survey of training and postvention practices. *Acad Psychiatry*. 1998; 22:181–189.
37. Sudak D, Roy A, Sudak H, Lipschitz A, Maltsberger J, Hendin H. Deficiencies in suicide training in primary care specialties: a survey of training directors. *Acad Psychiatry*. 2007;31(5):345–359.
38. Linehan MM, Comtois KA, Murray AM, et al. Two-year randomized controlled trial and follow-up of dialectical behavior therapy vs therapy by experts for suicidal behaviors and borderline personality disorder. *Arch Gen Psychiatry*. 2006; 63(7):757–766.
39. Young AH, Hammond JM. Lithium in mood disorders: increasing evidence base, declining use? *Br J Psychiatry*. 2007;191:474–476.
40. Kennard B, Silva S, Vitiello B, et al. Remission and residual symptoms after short-term treatment in the Treatment of Adolescents with Depression Study. *J Am Acad Child Adolesc Psychiatry*. 2006;45(12):1404–1411.

41. March JS, Silva S, Petrycki S, et al. The Treatment for Adolescents with Depression Study: long-term effectiveness and safety outcomes. *Arch Gen Psychiatry*. 2007;64(10):1132–1143.
42. Rihmer Z, Rutz W, Pihlgren H. Depression and suicide on Gotland: an intensive study of all suicides before and after a depression-training program for general practitioners. *J Affect Dis*. 1995;35(4):147–152.
43. Szanto K, Kalmar S, Hendin H, Rihmer Z, Mann JJ. A suicide prevention program in a region with a very high suicide rate. *Arch Gen Psychiatry*. 2007;64(8):914–920.
44. Bertolote J, Fleischmann A, DeLeo D, et al. Suicide attempts, plans, and ideation in culturally diverse sites: the WHO SUPRE-MISS community survey. *Psychol Med.* 2005;35:1457–1465.
45. Edddleston M, Gunnell D, Karunaratne A et al. Epidemiology of intentional self-poisoning in rural Sri Lanka. *BJPsych.* 2005;187:583–584.
46. Wilson J. Finland pioneers international suicide prevention. *Ann Intern Med.* 2004; 140(10):853–856.
47. Upanne M, Hakanen J, Rautava M. *Can Suicide be Prevented? The Suicide Project in Finland 1992–1996: Goals, Implementation and Evaluation.* Helsinki, Finland: National Research Development Centre for Welfare and Health; 1999.
48. World Health Organization. *Prevention of Suicide: Guidelines for the Formulation and Implementation of National Strategies*. Geneva, Switzerland: World Health Organization; 1996.
49. World Health Organization Regional Office for Europe. *Suicide Prevention in Europe: The WHO European Monitoring Survey on National Suicide Prevention Programmes and Strategies.* Copenhagen, Denmark: World Health Organization; 2002.
50. US Department of Health and Human Services. *National Strategy for Suicide Prevention: Goals and Objectives for Action.* Rockville, MD: US Department of Health and Human Services; 2001.
51. Harvard Injury Control Research Center. *Key Findings from the Means Matter Survey Fact Sheet*. Boston, MA: Harvard Injury Control Research Center; 2008. http://www.hsph.harvard.edu/means-matter/basic-suicide-facts/state-information/index.html. Accessed December 8, 2009.
52. Paulozzi LJ, Mercy J, Frazier L Jr, Annest JL. CDC's National Violent Death Reporting System: background and methodology. *Inj Prev*. 2004;10(1):47–52.
53. Institute of Medicine. *Reducing the Burden of Injury*. Washington, DC: National Academy Press, 1999.
54. Wasserman D, Värnik A. Suicide-preventive effects of perestroika in the former USSR: the role of alcohol restriction. *Acta Pscyhiatr Scand*. 1988;98(Suppl 394):1–4.
55. Razvodovsky YE. Alcohol and suicide in Belarus. *Psychiatr Danub*. 2009;21(3): 290–296.
56. Pridemore WA, Snowden AJ. Reduction in suicide mortality following a new national alcohol policy in Slovenia: an interrupted time-series analysis. *Am J Public Health*. 2009;99(5):915–920.
57. Wasserman D, Värnik A, Eklund G. Female suicides and alcohol consumption during perestoika in the former USSR. *Acta Pscyhiatr Scand*. 1988;98(Suppl 394):26–33.
58. Birckmayer J, Hemenway D. Minimum-age drinking laws and youth suicide: 1970–1990. *Am J Public Health*. 1999;89(9):1365–1368.
59. Darke S, Ross J, Lynskey M, Teesson M. Attempted suicide among entrants to three treatment modalities for heroin dependence in the Australian Treatment Outcome Study: prevalence and risk factors. *Drug Alcohol Depend.* 2004;73(1):1–10.
60. Landheim AS, Bakken K, Vaglum P. What characterizes substance abusers who commit suicide attempts? factors related to Axis I disorders and patterns of

substance use disorders: a study of treatment-seeking substance abusers in Norway. *Eur Addict Res*. 2006;12(2):102–108.
61. Knox KL, Litts DA, Talcott GW, Feig JC, Caine ED. Risk of suicide and related adverse outcomes after exposure to a suicide prevention program in the US Air Force: cohort study. *Br Med J.* 2003;327(7428):1376.
62. International Association of Chiefs of Police, US Department of Justice, EEI Communications. *Preventing Law Enforcement Officer Suicide: A Compilation of Resource and Best Practices*. Alexandria, VA: International Association of Chiefs of Police; 2008. CD-ROM; NCJ 224436.
63. May P, Serna P, Hurt L, DeBruyn L. Outcome evaluation of a public health approach to suicide prevention in an American Indian tribal nation. *Am J Public Health*. 2005;95(7):1238–1244.
64. Bennewith O, Gunnell D, Kapur N, et al. Suicide by hanging: multicentre study based on coroners' records in England. *Br J Psychiatry.* 2005;186:260–261.
65. Rhyne CE, Templer DI, Brown LG, Peters NB. Dimensions of suicide: perceptions of lethality, time, and agony. *Suicide Life Threat Behav.* 1995;25(3): 373–380.
66. Owens D, Horrocks J, House A. Fatal and non-fatal repetition of self-harm: systematic review. *Br J Psychiatry.* 2002;181:193–199.
67. Simon TR, Swann AC, Powell KE, et al. Characteristics of impulsive suicide attempts and attempters. *Suicide Life Threat Behav.* 2001;32(Suppl):49–59.
68. Personal communication, Thomas Simon, March 15, 2005.
69. Deisenhammer EA, Ing CM, Strauss R, Kemmler G, Hinterhuber H, Weiss EM. The duration of the suicidal process: how much time is left for intervention between consideration and accomplishment of a suicide attempt? *J Clin Psychiatry*. 2009;70(1):19–24.
70. Brent DA, Perper JA, Goldstein CE, et al. Risk factors for adolescent suicide: a comparison of adolescent suicide victims with suicidal inpatients. *Arch Gen Psychiatry*. 1988;45:581–588.
71. Brent DA, Perper JA, Allman CJ, et al. The presence and accessibility of firearms in the homes of adolescent suicides, a case-control study. *JAMA*. 1991;266:2989–2995.
72. Bukstein OG, Brent DA, Perper JA, et al. Risk factors for completed suicide among adolescents with a lifetime history of substance abuse: a case-control study. *Acta Psychiatr Scand*. 1993;88:403–408.
73. Brent DA, Perper JA, Moritz G, et al. Suicide in affectively ill adolescents: a case-control study. *J Affect Dis.* 1994;31:193–202.
74. Brent DA, Perper JA, Moritz G, et al. Firearms and adolescent suicide, a community case-control study. *Am J Dis Child*. 1993;147:1066–1071.
75. Kellermann AL, Rivara FP, Somes G, et al. Suicide in the home in relation to gun ownership. *N Engl J Med.* 1992;327:467–472.
76. Wiebe DJ. Homicide and suicide risks associated with firearms in the home: a national case-control study. *Ann Emerg Med.* 2003;41(6):771–782.
77. Kung HC, Pearson JL, Liu X. Risk factors for male and female suicide decedents ages 15–64 in the US: results from the 1993 National Mortality Followback Survey. *Soc Psychiatry Psychiatr Epidemiol*. 2003;38(8):419–426.
78. Kung HC, Pearson JL, Wei R. Substance use, firearm availability, depressive symptoms, and mental health service utilization among white and African American suicide decedents aged 15 to 64 years. *Ann Epidemiol.* 2005;15(8): 614–621.
79. Conwell Y, Duberstein PR, Connor K, Eberly S, Cox C, Caine ED. Access to firearms and risk for suicide in middle-aged and older adults. *Am J Geriatr Psychiatry*. 2002;10(4):407–416.

80. Grossman DC, Mueller BA, Riedy C, et al. Gun storage practices and risk of youth suicide and unintentional firearm injuries. *JAMA.* 2005;293(6):707–714.
81. Miller M, Lippmann S, Azrael D, Hemenway D. Household firearm ownership and rates of suicide across the 50 United States. *J Trauma*. 2007;62(4):1029–34, 1034–1035.
82. National Research Council. *Firearms and Violence: A Critical Review*. Washington, DC: The National Academies Press; 2005.
83. Ilgen MA, Zivin K, McCammon RJ, Valenstein M. Mental illness, previous suicidality, and access to guns in the United States. *Psychiatr Serv.* 2008;59(2):198–200.
84. Miller M, Barber C, Azrael D, Hemenway D, Molnar BE. Recent psychopathology, suicidal thoughts and suicide attempts in households with and without firearms: findings from the National Comorbidity Study Replication. *Inj Prev.* 2009; 15(3):183–187.
85. Oslin DW, Zubritsky C, Brown G, Mullahy M, Puliafico A, Ten Have T. Managing suicide risk in late life: access to firearms as a public health risk. *Am J Geriatr Psychiatry.* 2004;12(1):30–36.
86. Mann JJ, Apter A, Bertolote J, et al. Suicide prevention strategies: a systematic review. *JAMA*. 2005;294(16):2064–2074.
87. Kreitman N. The coal gas story: United Kingdom suicide rates, 1960–1971. *Br J Prev Soc Med.* 1976;30(2):86–93.
88. Gunnell D, Middleton N, Frankel S. Method availability and the prevention of suicide: a reanalysis of secular trends in England and Wales 1950–1975. *Soc Psychiatry Psychiatr Epidemiol*. 2000;35:437–443.
89. Gunnell D, Fernando R, Hewagama M, Priyangika WD, Konradsen F, Eddleston M. The impact of pesticide regulations on suicide in Sri Lanka. *Int J Epidemiol.* 2007;36(6):1235–1242.
90. Cantor CH. Suicide in the western world. In: Hawton K, van Heeringen K, eds. *The International Handbook of Suicide and Attempted Suicide.* West Sussex, UK: John Wiley & Sons; 2000:22.
91. Giggie MA, Olvera RL, Joshi MN. Screening for risk factors associated with violence in pediatric patients presenting to a psychiatric emergency department. *J Psychiatr Pract*. 2007;13(4):246–252.
92. Grossman J, Dontes A, Kruesi M, Pennington J, Fendrich M. Emergency nurses' responses to a survey about means restriction: an adolescent suicide prevention strategy. *J Am Psychiatr Nurses Assoc*. 2003;9:77–85.
93. McManus BL, Kruesi M, Dontes AE, et al. Child and adolescent suicide attempts: an opportunity for emergency departments to provide injury prevention education. *Am J Emerg Med*. 1997;15:357–360.
94. Kruesi MJ, Grossman J, Pennington JM, et al. Suicide and violence prevention: parent education in the emergency department. *J Am Acad Child Adolesc Psychiatry*. 1999;38:250–255.
95. Gunnell D, Eddleston M, Phillips M, Konradsen F. The global distribution of fatal pesticide self-poisoning: systematic review. *BMC Public Health.* 2007;7:357.
96. Manuweera G, Eddleston M, Egodage S, Buckley NA. Do targeted bans of insecticides to prevent deaths from self-poisoning result in reduced agricultural output? *Environ Health Perspect.* 2008;116(4):492–495.
97. Bowles JR. Suicide in Western Samoa: an example of a suicide prevention program in a developing country. In: Diekstra R, Gulbinat R, De Leo D, Kienhorst I, eds. *Preventive Strategies on Suicide*. Leiden, The Netherlands: Brill Academic Publishers; 1995.

98. Phillips M, Bertolote J, Beautrais A, et al. Reducing access in Asia to lethal means of self-harm. In: Hendin H, et al., eds. *Suicide and Suicide Prevention in Asia.* Geneva, Switzerland: World Health Organization; 2008.
99. Pirkis J, Blood RW, Beautrais A, Burgess P, Skehans J. Media guidelines on the reporting of suicide. *Crisis*. 2006;27(2):82–87.
100. Sudak HS, Sudak DM. The media and suicide. *Acad Psychiatry*. 2005;29(5):495–499.
101. Stack S. Media coverage as a risk factor in suicide. *J Epidemiol Community Health*. 2003;57(4):238–240.
102. Hegerl U, Wittenburg L, Arensman E, et al. Optimizing suicide prevention programs and their implementation in Europe: an evidence-based multi-level approach. *BMC Public Health.* 2009;9:428–436.

This page intentionally left blank

CHAPTER 18

Substance Use across the Life Span

Kenneth W. Griffin, PhD, MPH

Substance use and abuse are important public health problems across the life span from early adolescence to late adulthood. There is great diversity in the patterns of alcohol, tobacco, and other drug use over the life course, with some individuals abstaining from use throughout their lives and others facing ongoing battles with substance abuse and dependence. Although no pattern of substance use over the life course can describe the experience of all individuals, several national epidemiological studies show a consistent pattern of substance use across the life span from a population perspective. This general pattern includes initiation or experimental use during adolescence; an increase in the frequency, quantity, and types of substances used over the adolescent and young adult years; a peak in use during young adulthood; and a gradual decline through the remainder of the adult years.

Much of the published research on substance use and abuse has focused on adolescents and young adults because the use of alcohol, tobacco, and other drugs typically begins during adolescence and peaks in young adulthood. Like many public health problems that stem from behavioral factors, progress in addressing youth substance use has been achieved through advances in knowledge about the epidemiology, etiology, and psychosocial theory regarding the behavior. Documenting the epidemiologic patterns of substance use among young people has increased our understanding of the scope of the problem and the demographic subgroups at highest risk. Identification of the most salient risk and protective factors has shed light on the etiology of youth substance use. The organization of risk and protective factors into broader psychosocial models has provided theoretical conceptualizations of the problem that hint at possible intervention strategies. These comprehensive conceptual models have been used to guide the development of

Preparation of this chapter was supported in part by a grant from the National Institute on Drug Abuse (DA023890). Correspondence regarding this chapter should be sent to Kenneth W. Griffin, Department of Public Health, Division of Prevention & Health Behavior, Weill Cornell Medical College, 402 East 67th Street, Concourse 1, New York, NY 10065. Electronic mail may be sent to kgriffin@med.cornell.edu.

theory-based preventive interventions for schools, families, and communities. Through rigorous randomized trials, several of these programs have been found to be effective in reducing or preventing youth substance use.[1]

Examining the problem of youth substance use in a developmental framework has been critical to the development, focus, and timing of effective preventive interventions. During adolescence and young adulthood, there is substantial developmental change involving biological, psychological, occupational, and social role transitions. A range of new goals, developmental tasks, and life-span milestones occur, from the physical changes of puberty to the adoption of new adult roles, such as partner or parent. Adolescents increasingly strive for independence and self-definition. They make increasingly independent decisions about their own behavior during a time of life when they are highly susceptible to societal messages, media portrayals, adult and celebrity role models, and peer influences that promote substance use as a way to appear mature, rebellious, or independent.

Some young people, particularly those who are less successful in conventional pursuits, such as academics or sports, may turn to substance use as a way to define oneself as rebellious, to bond with deviant peers, or to attract attention from parents, peers, and other adults. Even if objectively these are not socially desirable goals, engaging in substance use to achieve them may be seen as functional from the perspective of the adolescent. Later, during the transition to young adulthood, substance use typically peaks during a time of new freedoms and relatively few responsibilities. Substance use then typically begins to decline in subsequent years as young adults adopt new adult responsibilities related to career, relationships, or parenting.

The literature on substance use during mid- to late adulthood has received less attention overall, and rarely is examined from a developmental perspective. This may be because adulthood is commonly thought of as developmentally static—the "end result" of adolescent growth and development. This, however, is an inaccurate view. Developmental psychologists and other experts in life course research emphasize the importance of various trajectories and transitions in development that occur over the entire life span.[2] A *trajectory* is a developmental pathway that occurs over the long term during the life span in a distinct aspect of life such as education, career, or parenthood. Trajectories include a series of *transitions*, which are discrete life events or changes in status that occur over short time spans, such as starting or leaving school, entering or leaving the first job, getting married, or becoming a parent. Thus, trajectories are long-term patterns of stability and change in which transitions occur from time to time giving them form and meaning.[2] These transitions or life events may be related to substance use in several ways. First, transitions can cause stress and test an individual's ability to adapt and self-regulate. In response to stress, an individual may use previously learned response patterns or adopt new ones. Some responses to stress may be adaptive, although many are maladaptive, such as engaging in substance use as a way to self-medicate subjective feelings of distress. Major transitions and life events can expose an individual to new people and situations, and these affiliations can contribute to increases or decreases in substance use.

Examining substance use in the context of trajectories and transitions across the life span, combined with what we know about patterns, risk, and protective factors, and psychosocial theory for different age-groups, may help us to gain a

more complete understanding of substance use at different stages of life. This may help inform the development and testing of effective preventive and treatment interventions across the life course.

GOALS OF THE CHAPTER

This chapter reviews the epidemiology of alcohol, tobacco, illicit drug use, and the nonmedical use of prescription and over-the-counter (OTC) medications across the life span. The chapter then reviews recent research linking patterns of substance use with developmental trajectories, transitions, goals, and milestones among adolescents, young adults, middle age, and older adults, including the elderly. Findings are presented in the context of how they may inform and guide prevention efforts. The focus is primarily on substance *use* rather than *abuse,* because substance use is more prevalent than abuse in the general population and therefore has a greater public health impact. A goal of the chapter is to examine the extent to which substance use can be thought of as a developmental phenomenon not only among young people, but throughout the life course. The implications of a life span developmental perspective to guide substance use prevention efforts are discussed.

EPIDEMIOLOGICAL DATA ON SUBSTANCE USE IN THE UNITED STATES

The National Survey on Drug Use and Health (NSDUH, formerly called the National Household Survey on Drug Use) is perhaps the most comprehensive ongoing epidemiologic study of substance use and abuse in the United States. NSDUH, sponsored by the Substance Abuse and Mental Health Services Administration,[3] has monitored the nature and extent of substance use among those ages 12 and older in the United States since 1971. NSDUH is now conducted annually to collect data on the magnitude and patterns of alcohol, tobacco, and other forms of substance use and abuse, including illicit drug use and the nonmedical use of prescription and OTC medications. Each year, NSDUH collects data from about 70,000 participants who randomly are drawn from the civilian, noninstitutionalized population of the United States. The NSDUH data are broken down in detail by age and other demographic variables, so it is well suited for comparing substance use rates among different age-groups in the United States A limitation, however, is that NSDUH is cross-sectional, and participants are not followed over time, so it is best thought of as a current "snapshot" of substance use. Because it is cross-sectional, it can mask cohort effects, or different rates of substance use among subgroups from different birth cohorts. Furthermore, issues related to the temporal sequencing and causal associations among substances cannot be examined in cross-sectional studies.

In the latest NSDUH survey,[3] the overall prevalence rate for past month alcohol use among Americans age 12 or older was 51 percent; 23 percent reported binge drinking (defined as five or more drinks on the same occasion) in the past month; and 7 percent reported heavy drinking (five or more episodes of binge drinking in the past month). The overall prevalence rate for lifetime cigarette smoking among those ages 12 and older was 65 percent, 29 percent for past-year smoking, and 24 percent for smoking in the past month. The overall prevalence rate for any illicit

drug use was 46 percent for lifetime use, 14 percent for past-year use, and 8 percent for past month use. Overall, the NSDUH data demonstrate that the use of legal substances (tobacco and alcohol) is much more prevalent than the use of illegal substances. Importantly, substantial differences exist in each of these behaviors according to the age category of respondents; these findings are reviewed below.

SUBSTANCE USE DURING ADOLESCENCE

Experimentation with substances in many cases begins during adolescence, usually in the context of one's peer group, and involves substances that are readily available, such as cigarettes, alcohol, and inhalants. Over the course of adolescence, prevalence rates typically escalate and some youth experiment with new drugs, such as marijuana, hallucinogens, cocaine, and the nonmedical use of prescription drugs or OTC medications.

Tobacco and Alcohol Use

To show how smoking prevalence differs across various age-groups of NSDUH respondents, Figure 18.1 shows rates of cigarette smoking among participants from 12 years of age to 65 and beyond. Before age 15, past month smoking is relatively rare, with prevalence rates under 10 percent. These rates increase rapidly in the next several age categories, however, with prevalence tripling to just under 30 percent at age 18. Furthermore, by the time adolescents reach age 18, the majority (52 percent) report having tried smoking in their lifetimes. When examining past month use of tobacco products among 12- to 17-year-olds as a group, 10 percent report cigarette use, 4 percent report cigar use, and 2 percent report smokeless tobacco use.

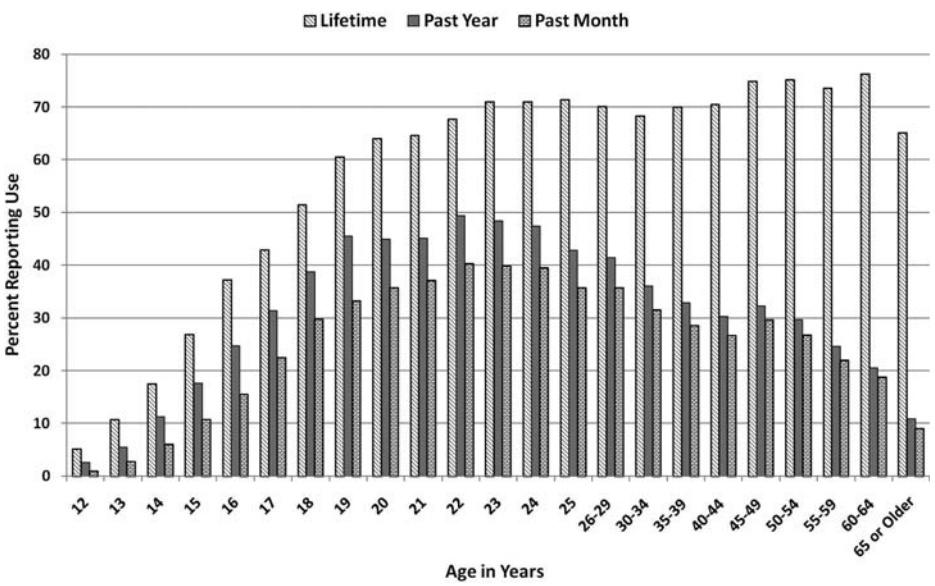

Figure 18.1 Cigarette Use in Lifetime, Past Year, and Past Month by Age Category in 2007. (*Source:* National Survey on Drug Use & Health (2008).)

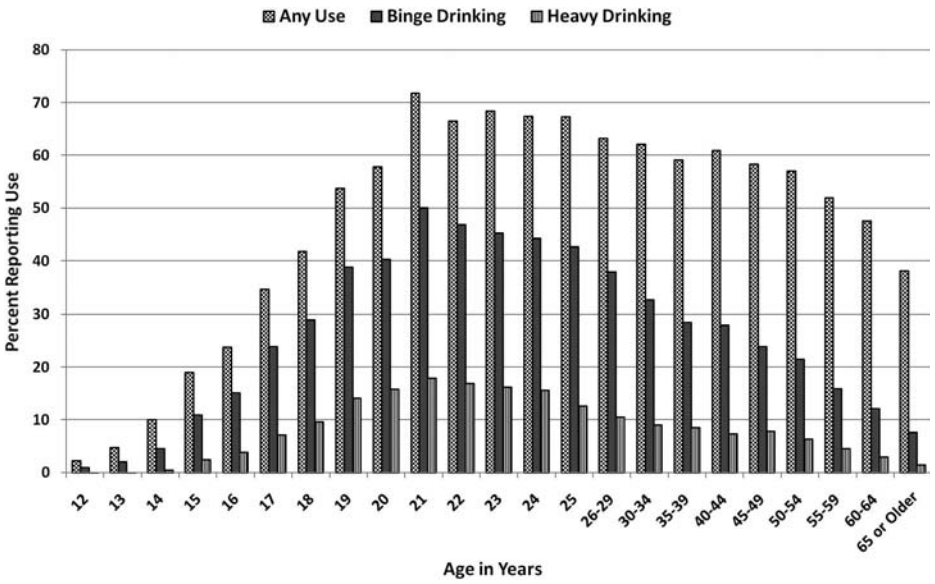

Figure 18.2 Alcohol Use, Binge Alcohol Use, and Heavy Alcohol Use in the Past Month by Age Category in 2007. (*Source:* National Survey on Drug Use & Health (2008).)

Similar to smoking, a dramatic increase is evident when comparing drinking rates among those in their early teens compared with those who are 18 years old. Figure 18.2 shows prevalence rates for alcohol use in the past month among NSDUH respondents by age. Less than 3 percent of 12-year-olds report having used alcohol in the past month, whereas 42 percent of 18-year-olds report this behavior. Rates of binge drinking in the past month are less than 1 percent for 12-year-olds, and increase for each subsequent age category, up to 29 percent among 18-year-olds. Heavy drinking is negligible before age 15, but 10 percent of 18-year-olds report five or more episodes of binge drinking in the past month. When examining past-month alcohol use among 12- to 17-year-olds as a group, 16 percent report any drinking, 10 percent report binge drinking, and 2 percent report heavy alcohol use.

Illicit Drug Use

Figure 18.3 shows prevalence rates for illicit drug use in the past month among NSDUH respondents by age. Rates of any illicit drug use in the past month increase sevenfold between those age 12 (3 percent) and those age 18 (21 percent). Among 12- or 13-year-olds, roughly 1 percent of participants reported using inhalants and 1 percent used marijuana in the past month. Among 14- or 15-year-olds, almost 6 percent used marijuana and 1 percent used inhalants in the past month. Marijuana also was the most commonly used illicit drug in the past month among 16- or 17-year-olds, at 13 percent, followed by hallucinogens, inhalants, and cocaine, each at about 1 percent. As shown in figure 18.4, more than 12 percent of 12- to 17-year-olds reported marijuana use in the past year, and 8 percent reported the use of other illicit drugs, the most prevalent being inhalants (4 percent) and hallucinogens (2 percent).

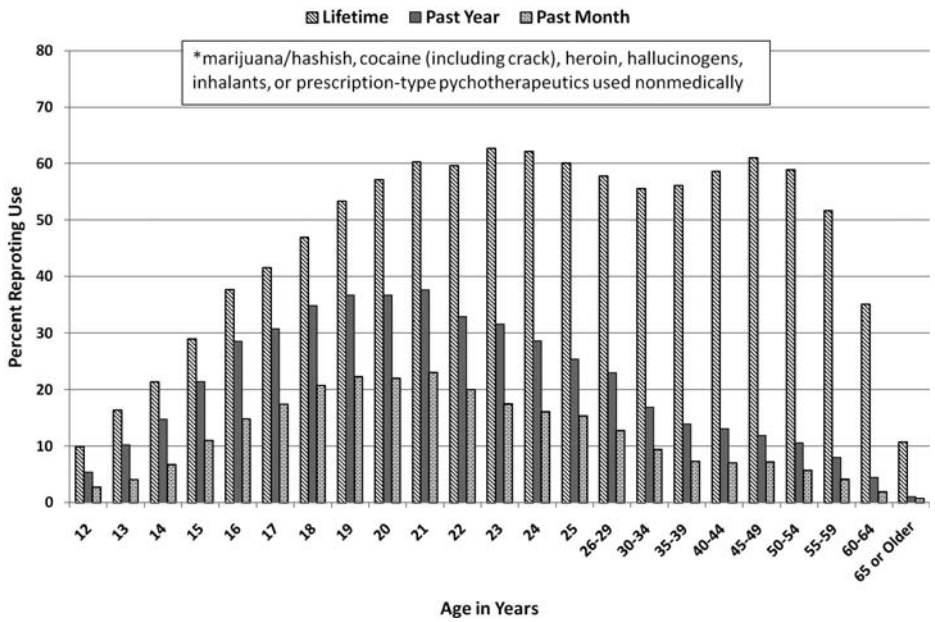

Figure 18.3 Any Illicit Drug Use by Age Category in 2007. (*Source:* National Survey on Drug Use & Health (2008).)

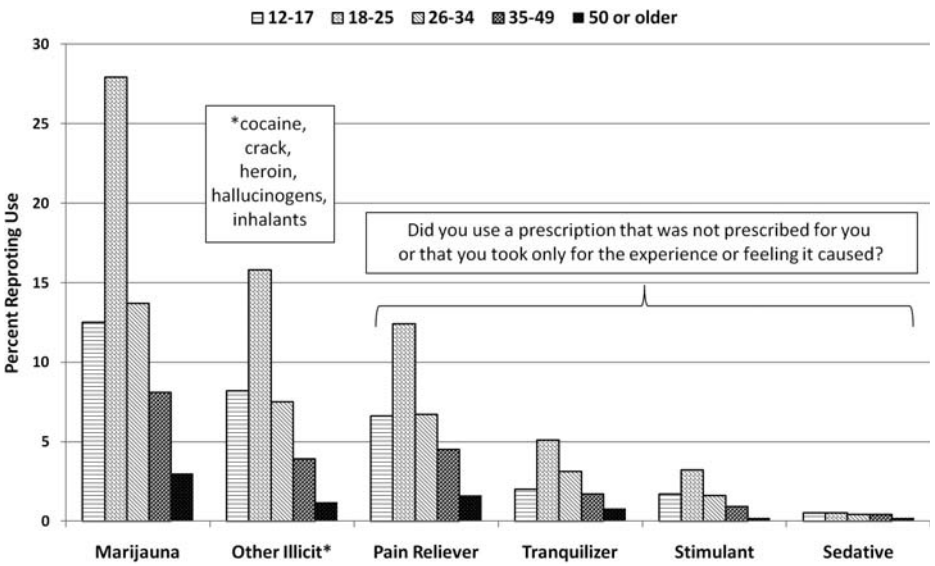

Figure 18.4 Annual Use of Illicit Drugs and Non-Medical Pill Use by Age Category in 2007. (*Source:* National Survey on Drug Use & Health (2008).)

Nonmedical Use of Prescription and Over-the-Counter Medications

Nonmedical pill use (NPU) was defined in the NSDUH as the abuse of prescription psychotherapeutics or prescription stimulants. Participants were asked if they had taken a list of psychoactive medications (specific pain killers, tranquilizers, stimulants, and sedatives) "prescribed for someone else or for the experience or feeling it caused." Almost 11 percent of 12- to 17-year-olds reported NPU in the past year, the most prevalent being pain relievers (7 percent) and tranquilizers (2 percent). The misuse of OTC medications is a growing problem among young people.[4] Approximately 4 percent of youth age 12 to 17 reported ever using an OTC cough and cold medication to get high, and nearly 2 percent had done so in the past year. The active ingredient in many cough suppressants, dextromethorphan (DXM), is found in more than 140 OTC cough and cold medications. DXM generally is safe when taken as recommended, but when taken in large doses, it can produce hallucinations or dissociative, "out-of-body" experiences. Unlike most other substances, where prevalence rates are typically higher for boys compared to girls, more teenage girls (4.3 percent lifetime, 2.3 percent past year) reported misusing OTC cough and cold medications than teenage boys (3.0 percent lifetime, 1.5 percent past year).

DEVELOPMENTAL FACTORS DURING ADOLESCENCE

To understand the dramatic increases in substance use among teenagers, it is important to consider that adolescence is marked by changes in many areas of a young person's life. Along with the biological and physical changes of puberty, there is significant development in social, emotional, and cognitive functioning.[5] Puberty involves changes in physical appearance, greater hormonal activity, development of secondary sex characteristics, and maturation of areas of the brain, such as the prefrontal cortex, responsible for advanced reasoning skills. During this time, youth experience significant development in capacities for moral reasoning, perspective taking, and problem solving. A number of important new developmental goals pertain to identity formation and individuation, a desire for greater autonomy from parents, an increase in the importance of peers, new feelings of sexual attraction, and a heightened awareness of gender roles. Furthermore, many new social roles emerge in the context of school (for example, athlete, scholar, and so on), employment (for example, first part-time job outside the home), and interpersonal relationships. Different life trajectories and key transitions during adolescence can have important implications for substance use, some of which are reviewed below.

Puberty

Several studies have examined the relationship between pubertal timing and substance use. Much of the research has focused on early maturing girls, because girls typically reach puberty before boys. Studies have shown that reaching puberty before one's peers is associated with substance use, particularly among girls, and that early maturing girls escalate in substance use at a faster rate than their peers.[6, 7, 8] A number of hypotheses have been advanced to explain the relationship between early pubertal timing and substance use. These include the possibility that (1) early maturing teenagers experience peer rejection, distress, or low self-esteem

because their bodies are changing at a time when their peers are not; (2) they are monitored less carefully by parents given their more mature physical appearance; or (3) they are more likely to seek out (or be sought out by) older peers and romantic partners that exert pressure to engage in substance use as well as sexual activity. Some studies have found that the effects of early pubertal timing on substance use persist after controlling for psychological distress,[9, 10] findings that are inconsistent with the first hypothesis. Currently, evidence is limited for these hypotheses, and further research is needed to better understand the mechanisms involved. Nevertheless, early maturation should be regarded as a risk factor for early onset and escalation of substance use.

Educational Transitions

During the transitions into middle school and high school, many young people experience difficulty adjusting to the change when moving from a smaller, nurturing school environment to a larger, more demanding, and challenging one.[11] School transitions can be accompanied by decreases in grade point averages, school attendance, and school bonding as well as increased psychological distress.[12, 13] However, surprisingly, little empirical support for the hypothesis indicates that school transitions uniquely increase substance use, independent from the increase in risk that typically occurs during these years of adolescence. One study compared the transition from eighth to ninth grades among a group of students that changed schools with a group that did not change schools,[14] a specific naturalistic design feature that few similar studies have incorporated. Findings indicated that several academic and nonacademic outcomes (including substance use) got significantly worse but did so equally among both groups. One interpretation of this finding is that certain transitional years of adolescence are associated with decreases in psychosocial adjustment and that this may be independent of school transitions. On the other hand, students entering new middle or high school environments are likely to be exposed to a larger and more diverse set of peers, their own age and older, who model substance use, have positive attitudes about use, and increase the availability of illicit drugs and alcohol. More research is needed to identify the characteristics of students that are most influenced by these negative social forces when entering a new school environment.

First Employment

A majority of high school students in the United States work outside the home. Although there can be important benefits to part-time work for students, such as learning to handle new responsibilities and gaining valuable work experience, employment may provide increased opportunities to engage in substance use. Young employees may have the opportunity to experiment with substances during unsupervised work-related social gatherings, may be exposed to workplace norms that may support use, drugs may be available from older co-workers, and the wages that young workers earn may be spent on drugs.[15, 16] In fact, research has shown that work outside the home is associated with higher rates of cigarette use, alcohol use, illicit drug use, and heavy substance use among secondary school students.[17, 18] Rates of cigarette smoking among employed adolescents are positively correlated

with the number of hours worked [19] and the amount of spending money available.[20] It appears likely that the norms, attitudes, and behaviors regarding substance use among co-workers, along with levels of workplace supervision and parental monitoring, are important factors in determining whether part-time work will adversely affect a student's risk for substance use.

SUBSTANCE USE DURING YOUNG ADULTHOOD

Use of alcohol, tobacco, and other substances typically reaches a peak during late adolescence and early adulthood, as young people begin to experience some of the new freedoms of young adulthood, such as living independently from parents. For most people, substance use begins to decline in the mid- to late 20s, with a more rapid decline occurring in the later years of adulthood. This decline typically occurs concurrently with some of the new responsibilities of adulthood, such as career and family.

Tobacco and Alcohol Use

Prevalence rates for cigarette smoking in 2007 among NSDUH respondents escalate rapidly for each subsequent age category during adolescence (see figure 18.1). At age 22, past-month smoking peaks at 40 percent and past-year smoking peaks at 49 percent. Smoking for subsequent age categories declines gradually, but prevalence rates for past-month smoking remain above 30 percent through the 30 to 34 age-group. When examining past-month use of tobacco products among 18- to 25-year-olds as a group, 36 percent report cigarette use, 12 percent report cigar use, and 5 percent report smokeless tobacco use. Rates of cigarette smoking in young adults have decreased in recent years, but cigar smoking has stayed steady or increased in prevalence. Rates of cigar use peak at age 18, several years earlier than cigarette smoking, and cigar use is three times as likely among young adult males than females.[21] The highest rates of alcohol use in the 2007 NSDUH data were seen in 21-year-olds (see figure 18.2). At 21, use of any alcohol in the past month was highest at 72 percent of respondents, binge drinking was highest at 50 percent, and heavy drinking was highest at 18 percent. When examining past-month alcohol use among 18- to 25-year-olds, 61 percent report any drinking, 42 percent report binge drinking, and 15 percent report heavy alcohol use.

Illicit Drug Use

Rates for past year use of any illicit drug peaked at 37 percent among 21-year-olds, and past month illicit drug use peaked at 23 percent (see figure 18.3). Thus, at age 21 and 22, after turning "legal" for alcohol use, prevalence rates reached their highest levels of any age-group in the 2007 NSDUH data not only for alcohol and tobacco, but also for illicit drug use. Marijuana is by far the most commonly used illicit drug in the United States across all age categories, as shown in figure 18.4. Almost 28 percent of 18- to 25-year-olds reported using marijuana in the past year, and 16 percent reported the use of other illicit substances, with cocaine being the most prevalent at 7 percent, followed by hallucinogens at 6 percent. Among hallucinogens, Ecstasy (MDMA) is most prevalent among 18- to 25-year-olds at 3.7 percent

in the past year. The second most prevalent hallucinogen at 1.7 percent prevalence is a new hallucinogen, the herb Salvia (*Salvia divinorum*), now used more commonly than Lysergic acid diethylamide (LSD) (1.3 percent) among young adults. Salvia is an herb common to Mexico that is becoming increasingly popular as a legal hallucinogen in the United States. Its use as a recreational drug involves chewing fresh leaves or smoking dried leaves, and salvia produces psychedelic like effects, feelings of detachment, and an altered sense of external reality and the self.[22] One study found that 4.4 percent of students at a southwest U.S. college reported using salvia in the past year.[23] Currently, salvia is being sold on several Internet Web sites, and is sometimes referred to as "Sally D." The U.S. Drug Enforcement Agency lists salvia as a "drug of concern" and is considering classifying it as a Schedule I drug, like LSD or marijuana.

Nonmedical Use of Prescription and Over-the-Counter Medications

NPU is the second most prevalent type of illicit drug use among young adults after marijuana. In NSDUH, annual rates of NPU peak at age 18 to 25 (21 percent), the most prevalent being pain relievers (12 percent), tranquilizers (5 percent), and stimulants (3 percent), as shown in figure 18.4. Furthermore, approximately 7 percent of persons age 18 to 25 reported ever using an OTC cough and cold medication to get high, and nearly 2 percent had done so in the past year.[4] Among stimulants, the nonmedical use of certain prescription amphetamines is fairly common in this age-group, especially among college students. One of these, Adderall, is prescribed to treat attention deficit hyperactivity disorder (ADHD). In the NSDUH data, 6.4 percent of full-time college students ages 18 to 22 engaged in the nonmedical use of Adderall, which was more than twice the rate of 3 percent among noncollege students of the same age.[24] A prescription methamphetamine (Desoxyn) is also used to treat ADHD. Both Adderall and Desoxyn have high abuse potential. Although Desoxyn is produced legally as a prescription methamphetamine, most methamphetamine in the United States is produced in illegal "meth labs" where individuals "cook" the ingredients, which include OTC medications. Methamphetamine abuse leads to devastating problems for individual users and their communities.[25] Regular meth users suffer from several psychological, medical, and social problems such as aggression, memory loss, paranoid and psychotic behavior, heart damage, serious dental problems, malnutrition, and homelessness. Widespread use within a community can lead to increased crime, unemployment, child neglect or abuse, and other social problems. Together, prescription and illegally produced methamphetamines report being ever used by 6 percent of 18- to 25-year-olds in the NSDUH data. The lifetime prevalence rate of methamphetamine use for those over 25 is also 6 percent. Past-year use of methamphetamines was reported by 1.2 percent of 18- to 25-year-olds, and 0.4 percent of those over 25.

DEVELOPMENTAL FACTORS DURING YOUNG ADULTHOOD

When making the transition from adolescence to young adulthood, a number of social role transitions typically occur in the contexts of education, occupation, close relationships, living arrangements, and family. Initially, young people may experience new freedoms, such as living independently from parents or attending college

away from home. At the same time, they may find themselves with relatively few responsibilities, particularly among the increasing number of young people that delay marriage and family. Many of these new freedoms in young adulthood are associated with increases in the use of certain substances. Eventually, however, most young people adopt new roles such as full-time employee, spouse, partner, or parent that come with a host of new responsibilities. These role transitions are typically associated with a decrease in substance use. This pattern of new freedoms followed by new responsibilities may be the driving forces responsible for peak levels of substance use in the early 20s, followed by a gradual decline in the mid- to late 20s that continues into subsequent years of adulthood.

Post–High School

A longitudinal study followed several cohorts of high school seniors from the *Monitoring the Future* study for several years after they graduated from high school.[26, 27] Findings indicated that patterns of substance use differed depending on occupational or educational pursuits after graduation. Young adults who pursued full-time employment after high school were found to increase their daily smoking but decrease rates of heavy drinking and marijuana use, compared with their rates as seniors in high school. Those who became full-time homemakers after graduation decreased their alcohol, marijuana, and cocaine use, but much of this effect appears to be due to their marital and parental status.[26] Young people who join the military after high school graduation increased their levels of smoking and alcohol use, but rates of marijuana and cocaine use dropped dramatically upon enlisting in the military, likely due to the imposition of drug testing before enlistment and continuing random testing after enlistment.

There were a number of differences in substance use patterns among those who attended college after high school, compared with those who began to work full time or pursued other activities.[26] Both cocaine use and cigarette smoking were lower among college students compared with nonstudents; in fact, smoking increased two to three times as much in the years after high school among nonstudents compared with full-time students. College students were found to use marijuana use at about the same rate as nonstudents in the years after high school.

The most notable difference between college students and nonstudents was in alcohol use, which is widely recognized as a serious problem among college students. Young people who attended college surpassed levels of alcohol use compared with nonstudents during the college years.[26] Several studies have shown that the first year of college, and in particular the first semester, is a high-risk period for binge drinking and heavy drinking.[28, 29, 30] Many first-year college students spend a substantial amount of time socializing in order to establish new friendships and a social network. First-year college students attend dorm and fraternity or sorority parties in a greater proportion than upperclassmen,[29] and alcohol and drug use is highly prevalent and socially approved in these environments.

Drinking among college students often occurs in highly predictable patterns. Throughout the first year of college, much of the drinking that occurs is binge or heavy episodic drinking, mostly on Thursday, Friday, and Saturday nights.[28] College sporting events such as high-profile college football games (for example,

"homecoming") often lead to heavy drinking, comparable with other well-known drinking days such as New Year's Eve, St. Patrick's Day, and Halloween.[31] First-year college students who go on spring break trips are likely to increase their alcohol consumption and heavy drinking not only on the trip itself, but also for weeks following the trip.[30] A recent study found that 21st birthday celebrations are often a time of extreme drinking among students.[32] The study found that four out of five students turning 21 drank alcohol to celebrate, and half of these exceeded their prior maximum number of drinks. Furthermore, 12 percent of participants in the study engaged in a dangerous practice called "21 for 21" or "21 Run" in which students attempt to consume 21 drinks as part of the celebration.

While some students entering college expect to engage in alcohol use and have positive attitudes about doing so, for others, changes in alcohol use behavior are accompanied by changes in attitudes about alcohol and substance use. College students who engage in regular drinking and other types of substance use report a decline in perceived risks and less disapproval of substance use.[27] Many college students come to view drinking as a rite of passage that is integral to the student experience and feel entitled to drink irresponsibly.[33]

Post-College

Despite the substantial increase in alcohol use in college students, college attendance is actually a protective factor for alcohol use and abuse in the years after college. The long-term risk of heavy drinking and alcohol dependence is lower among college graduates relative to high school dropouts and those who do not attend college.[34, 35] Just as the new freedoms at the beginning of young adulthood often are associated with increased substance use, the new responsibilities that typically occur a few years into young adulthood are associated with decreased substance use. As young adults begin full-time employment, enter into marriage or committed relationships, or start a family, substance use typically decreases as these responsibilities become central to their lives and identities.[26, 27]

Relationship Status

Getting married is a major life event that has pervasive effects on many aspects of young adult life, including drug use behavior. The longitudinal follow-up study of high school seniors from the *Monitoring the Future* study found that marriage was associated with a decrease in use of alcohol, cigarettes, marijuana, and cocaine.[26] Other researchers have found that entering marriage is associated with decreases in heavy alcohol use.[36, 37] Marriage is associated with a decrease in substance use in part because it ushers in a change in social and recreational activities and attitudes about alcohol and drugs.[27] Compared with those who remain single, married people reported fewer evenings out for fun and recreation, getting together less often with friends, and rarely going out to bars, nightclubs, and parties. Married people were found to form stronger negative attitudes about substance use and were more likely to disapprove of drug use and see it as risky.

In the *Monitoring the Future* cohort study, getting engaged to be married also was associated with reductions in substance use, perhaps because engagement take people out of the singles scene, but these effects were not as strong as marriage

effects.[27] Unlike marriage and engagement, cohabitation with a romantic partner was not related to changes in substance use. Cohabitation often reflects an embrace of nontraditional ideas and alternative lifestyle choices, and the findings reflected this. Cohabitating couples reported less religious behavior and values, lower perceptions of risk, and less disapproval of substance use, compared with those who marry or become engaged.[27]

After marriage, discrepancies in the levels of husbands and wives' substance use can predict decreased marital satisfaction and divorce.[38, 39] To the extent that people choose partners for marriage that are similar in backgrounds, values, and behaviors, an ongoing discrepancy in substance use behavior may suggest a less-than-ideal level of compatibility, and other facets of the marriage may suffer. This is particularly true when one partner engages in more serious levels of substance use than the other partner.[40] Controlling for substance use before marriage, marital discord, and religiosity, one study found that alcohol intoxication was an independent predictor of later divorce, although marijuana use was not.[41] Although substance use can predict divorce, it is also true that divorce often triggers an increase in substance use, as has been found in samples of divorced young adults.[26]

Pregnancy and Parenthood Effects

Substance use during pregnancy can lead to a number of adverse pregnancy outcomes. Alcohol use during pregnancy can lead to fetal alcohol syndrome. Smoking during pregnancy increases the risk of spontaneous abortions, stillbirth, preterm birth, and fetal growth restriction, and can contribute to long-term health effects on infants, including neurodevelopmental disorders and cancers.[42] Thus, it is not surprising that pregnancy has wide-ranging effects on substance use behavior among women. Smoking prevalence among pregnant women in the United States has decreased dramatically in recent decades. It is estimated that 40 percent of pregnant women in the United States smoked in 1967, and this number decreased to 12 percent in 2000. Much of this decrease, however, likely is due to lower levels of smoking initiation rather than increased smoking cessation before or during pregnancy.[42] Alcohol use among pregnant women has decreased as well in recent years.[43]

SUBSTANCE USE DURING MID- TO LATE ADULTHOOD

After peak levels of substance use in young adulthood, the prevalence rates of almost all forms of substance use decrease beginning in the late 20s and continue to decline as adults get older.

Tobacco and Alcohol Use

The decrease in alcohol and tobacco use during mid- and late adulthood is gradual. Prevalence rates for past month smoking in the NSDUH data hover between 35 percent and 25 percent for age-groups between 25 and 54 years, before dropping below 25 percent among 55- to 59-year-olds and below 10 percent for those 65 and older (see figure 18.1). Past-month binge drinking prevalence rates remain above 30 percent through the mid-30s age-group, drops below 20 percent in the late-50s age-group, and below 10 percent for those 65 and older (see figure 18.2). Prevalence

rates for past-month heavy drinking and any illicit drug use drop below 10 percent with the early 30s age-group and gradually decline for older age categories, to less than 2 percent for those 65 and older.

Illicit Drug Use

A notable gap exists in empirical research on illicit substance use and abuse among older adults (greater than or equal to 50 years of age) and the elderly (greater than or equal to 65 years of age). The lack of research may reflect the traditional view that prevalence rates of illicit drug use are likely to be quite low in these groups. More research has been conducted on alcohol and prescription drug abuse, which tends to be more prevalent than the use or abuse of illicit substances, especially among those age 65 and older.[44] Because of a number of factors, however, it is widely anticipated that substance use and abuse, including the use of various illicit drugs, will increase dramatically among the elderly over the next several decades. The elderly are the fastest growing age cohort in the United States, projected to more than double in size between 2000 and 2030.[45] This increase will be a function of longer life expectancies and the fact that the very large baby boom generation will begin to pass age 65 in the next few years.[46] The baby boom cohort (born between 1946 and 1964) grew up in the 1960s and 1970s and tends to have more lenient attitudes about recreational or casual drug use than previous cohorts. Thus, as the numbers of elderly persons are projected to increase dramatically, the proportion of individuals who use illicit drugs is anticipated to be much higher than in previous cohorts.

Nonmedical Use of Prescription Medications

The abuse of prescription medications among the elderly is likely to grow in prevalence in future decades. A literature review on psychoactive drug abuse in older adults [44] found that one in four older adults is prescribed psychoactive drugs with abuse potential (for example, pain killers, sedatives, tranquilizers, stimulants). An increase in prescription drug abuse is expected because of such factors as the overall increase in the number of psychoactive prescription drugs available, the growing number of prescriptions written by physicians for an increasingly large aging population with chronic conditions, and the increase in demand for prescription medications due to direct-to-consumer marketing by the pharmaceutical companies, among other factors. Generally, prescription drug abuse is more common among elderly women than men, and alcohol abuse is more common among elderly men than women. Furthermore, women over 65 are prescribed and consume more psychoactive medications and are more likely to be long-time users, compared with men over 65. The problem of prescription drug abuse in likely to continue to affect elderly women at a disproportionate rate.

DEVELOPMENTAL FACTORS DURING MID- TO LATE ADULTHOOD

Because of gaps in research on substance use and abuse among older adults and the elderly, and even less research focusing on how developmental factors in this age-group might affect substance use, few conclusions can be drawn definitively. Findings from existing research, however, may provide some directions for future work.

Life Events

During the years of middle-age to late life, major life events are numerous and diverse and can include children leaving home, divorce and remarriage, job loss and retirement, death or illness of spouse or other family members, and the onset or exacerbation of chronic illnesses. These events potentially may be related to substance use. The vast majority of existing research in this area has focused on alcohol use and abuse. In a six-year panel study of more than 7,700 adults between the ages of 51 and 61 at baseline, researchers examined how life events were associated with alcohol consumption.[47] In general, findings indicated that, for most participants, alcohol use did not change substantially, and when changes were observed, they primarily were decreases in use. Several life events related to occupation, social relationships, and health, however, were associated with changes in substance use. The onset of retirement was associated with increased drinking and persisted over time. Loss of a spouse was also associated with an increase in drinking, although the effect was limited to about two years' duration. Whether these increases were attributable to having more time available, increased social isolation, attempts to self-medicate, or some other mechanism could not be determined. As one might expect, new health problems were associated with a short-term reduction in alcohol use.

In studying alcohol use among the elderly, some investigators have made a distinction between those who have been using or abusing alcohol for many years and have reached 65, and a late-onset group that begins to drink later in life. Some evidence indicates that the late-onset group is more likely to initiate or escalate drinking in response to major life changes.[48, 49] Because of the lack of rigorous, longitudinal research on these issues, however, little empirical evidence to document the prevalence of this phenomenon or to address the specific etiologic mechanisms involved.

Physiological Changes

Substance use and abuse in older adults and the elderly are affected by the physiological changes that occur with aging. These changes can include a reduction in cardiac output, decreased efficiencies in the functioning of major bodily organs (for example, liver, kidneys, and gastrointestinal system), and other physiological changes that result in slower metabolism, absorption, and excretion of medications. Combined with the loss of body mass and use of multiple medications in the elderly, alcohol and drugs (illicit, prescribed, and OTC) tend to stay in the body longer at increased concentrations. These factors can prolong and magnify their effects and increase the risk of negative side effects and adverse reactions.[50] A variety of changes in brain structure and function also occur among elderly persons. Some of these, such as changes in neurotransmitter systems and areas of the brain responsible for reward sensitivity, may augment the risks of substance use among the elderly and complicate the diagnosis and treatment of substance abuse.[51]

CONCLUSION

This chapter has reviewed the epidemiology of substance use from a life-span development perspective. We have shown that from a population perspective, a

normative typical course of substance use is well documented in national epidemiological data sets. This involves onset and escalation of substance use during adolescence, a peak in use in young adulthood, and a gradual decrease in mid- to late adulthood. A similar pattern is consistently observed for alcohol, cigarette smoking, marijuana use, and most other forms of illicit drug use. Of course, for any set of individuals, one will see great diversity in substance use patterns and trajectories over the life course. This variation has been the focus of much recent etiologic work on substance use.[52, 53]

Whereas factors leading to *substance abuse* such as personality disorder, behavioral disinhibition, and affect dysregulation [54] typically are chronic and difficult to change, many of the developmental linkages to *substance use* are predictable in time and more amenable to intervention. It is clear that several developmental transitions, goals, and milestones are associated with an increased risk of substance use, such as early maturation among girls, starting a first part-time job, first semester of college, 21st birthday, marriage and parenthood, divorce and retirement. The well-established epidemiologic patterns of substance use across the life span, combined with the observed links to age-related tasks and transitions, provide compelling evidence that substance use is a developmental phenomenon. This in turn has important implications for prevention efforts.

Substance use during early adolescence often is tied to developmental goals, such as identity formation, separation from parents, bonding with peers, and a desire to "fit in." Prevention programs for early adolescents should be skills focused, helping young people deal with the internal psychological and external social pressures from peers and the media that contribute to substance use experimentation. Substance use during the transition to young adulthood is frequently linked to transitions and discrete milestones, and this knowledge may provide a basis for prevention efforts. For example, windows of risk for excessive drinking among college students include orientation at the beginning of the first school year, fraternity and sorority parties, homecoming celebrations and holidays, school breaks including spring break, turning "legal" at age 21, and graduation and other end-of-school-year celebrations. Knowledge of these high-risk events can help in the development of event-specific prevention activities. Intervention strategies to address specific high-risk periods can be developed at the individual, group, institution, community, and societal levels. They can be designed to address knowledge, change attitudes, correct perceived norms about use and abuse, reduce access to substances, eliminate marketing of alcohol and tobacco, and increase enforcement activities. For example, a recent study [55] describes attempts made by a University of Connecticut task force to reduce problem drinking during its annual on-campus "spring weekend" celebration. A broad range of prevention strategies were implemented, such as establishing fun, alcohol-free activities to compete with street parties, media campaigns on campus to reduce positive drinking expectancies, implementing a more visible police presence and driving while intoxicated checkpoints, and training nursing students to identify signs of alcohol poisoning and to monitor residence halls. These combined activities were successful in reducing the number of students seen by medical personnel during the intervention period.

A number of demographic shifts occurring now and in the coming decades will present new challenges. First, the transition to young adulthood is changing

compared with previous generations. Increasingly, the years between adolescence and adulthood are extended in time with young adults transitioning in and out of school, relationships, jobs, and living arrangements. Young people have more opportunities to take on different roles and explore various career and lifestyle options and have more opportunities for self-expression. The implications for substance use are not yet clear; however, the extended transition into young adulthood lengthens the period of time during which substance use is typically at its peak.

A recent and worrisome situation concerns the use of heroin among those in their teens and early 20s from middle- or upper-middle-class suburban families. These young adults are part of a new wave of heroin abuse heretofore not seen. During the first six months of 2009, for example, Nassau County (suburban New York City) counted 25 deaths from heroin overdoses, which was more than from homicide and drunken driving combined.[56] The drug is cheap and accessible and being used by younger kids than ever before. The number of individuals entering detoxification centers and withdrawal programs has soared in Nassau County and in Westchester County (also a New York City suburban county) and, in all probability, other counties are experiencing similar spikes in heroin abuse.

Another upcoming challenge will be the anticipated increase in substance use problems among the elderly, as this population doubles in size over the next few decades. By taking into account how age-related developmental factors can affect substance use, we may be better able to address these and other new prevention challenges in the future.

REFERENCES

1. Griffin KW, Botvin GJ. Preventing substance abuse among children and adolescence. In: Ries, RK, DA Fiellen, SC Miller, R Saitz, eds. *Principles of Addiction Medicine*. 4th ed. Baltimore, MD: Lippincott Williams & Wilkins; 2009:1375–1382.
2. Elder GH. Perspectives in the life course. In: GH Elder Jr, ed. *Life Course Dynamics*. Ithaca, NY: Cornell University Press; 1985:23–49.
3. Substance Abuse and Mental Health Services Administration. *Results from the 2007 National Survey on Drug Use and Health: National Findings*. Rockville, MD: US Department of Health and Human Services; 2008. Office of Applied Studies, NSDUH Series H-34, DHHS Publication No. SMA 08-4343.
4. Substance Abuse and Mental Health Services Administration. *The NSDUH Report: Misuse of Over-the-Counter Cough and Cold Medications Among Persons Aged 12 to 25*. Rockville, MD: Office of Applied Studies; 2008.
5. Steinberg L, Morris AS. Adolescent development. *Ann Rev Psychol*. 2001;52:83–110.
6. Dick DM, Rose RJ, Viken RJ, Kaprio J. Pubertal timing and substance use: associations between and within families across late adolescence. *Devel Psychol*. 2000;36:180–189.
7. Lanza ST, Collins LM. Pubertal timing and the onset of substance use in females during early adolescence. *Prev Sci*. 2002;3:69–82.
8. Stice E, Presnell K, Bearman SK. Relation of early menarche to depression, eating disorders, substance abuse, and comorbid psychopathology among adolescent girls. *Devel Psychol*. 2001;37:608–619.
9. Tschann JM, Adler NE, Irwin CE, Millstein SG, Turner RA, Kegeles SM. Initiation of substance use in early adolescence: the roles of pubertal timing and emotional distress. *Health Psychol*. 1994;13:326–333.

10. Wiesner M, Ittel A. Relations of pubertal timing and depressive symptoms to substance use in early adolescence. *J Early Adolesc*. 2002;22:5–23.
11. Eccles JS, Midgley C. Changes in academic motivation and self-perception during early adolescence. In: Montemayor R, Adams GR, Gulotta TP, eds. *From Childhood to Adolescence: A Transitional Period*. Newbury Park, CA: Sage Publications; 1991.
12. Chung HH, Elias M, Schneider K. Patterns of individual adjustment changes during middle school transition. *J School Psychol*. 1998;36:83–101.
13. Isakson K, Jarvis P. The adjustment of adolescents during the transition into high school: a short-term longitudinal study. *J Youth Adolesc*. 1999;28:1–26.
14. Weiss CC, Bearman PS. Fresh starts: reinvestigating the effects of the transition to high school on student outcomes. *Am J Educ*. 2007;113:395–421.
15. Arnett JJ. *Emerging Adulthood: The Winding Road from the Late Teens through the Twenties*. New York: Oxford University Press; 2004.
16. Wright JP, Cullen FT, Williams N. The embeddedness of adolescent employment and participation in delinquency: a life course perspective. *West Criminol Rev*. 2002;4:1–19.
17. Paschall MJ, Flewelling RL, Russell T. Why is work intensity associated with heavy alcohol use among adolescents? *J Adolesc Health*. 2004;34:79–87.
18. Wu L, Schlenger W, Galvin D. The relationship between employment and substance use among students aged 12 to 17. *J Adolesc Health*. 2003;32:5–15.
19. Ramchand R, Ialongo NS, Chilcoat HD. The effect of working for pay on adolescent tobacco use. *Am J Public Health*. 2007;97:2056–2062.
20. Zhang B, Cartmill C, Ferrence R. The role of spending money and drinking alcohol in adolescent smoking. *Addiction*. 2008;103:310–319.
21. Substance Abuse and Mental Health Services Administration. *The NSDUH Report: Cigar Use among Young Adults Aged 18 to 25*. Rockville, MD: Office of Applied Studies; 2009.
22. González D, Riba J, Bouso JC, Gómez-Jarabo G, Barbanoj MJ. Pattern of use and subjective effects of Salvia divinorum among recreational users. *Drug Alcohol Depend*. 2006;85:157–162.
23. Lange JE, Reed MB, Croff JM, Clapp JD. College student use of Salvia divinorum. *Drug Alcohol Depend*. 2008;94:263–266.
24. Substance Abuse and Mental Health Services Administration. *The NSDUH Report: Nonmedical Use of Adderall® among Full-Time College Students*. Rockville, MD: Office of Applied Studies; 2009.
25. National Institute on Drug Abuse. *Research Report: Methamphetamine Abuse and Addiction*. Bethesda, MD: National Institute on Drug Abuse; 2006. NIH Publication No.06-4210.
26. Bachman JG, Wadsworth KN, O'Malley PM, Johnston LD, Schulenberg JE. *Smoking, Drinking, and Drug Use in Young Adulthood: The Impacts of New Freedoms and New Responsibilities*. Mahwah, NJ: Lawrence Erlbaum Associates; 1997.
27. Bachman JG, O'Malley PM, Schulenberg JE, Johnston LD, Bryant AL, Merline AC. *The Decline of Substance Use in Young Adulthood: Changes in Social Activities, Roles, and Beliefs*. Mahwah, NJ: Lawrence Erlbaum Associates; 2002.
28. Del Boca FK, Darkes J, Greenbaum PE, Goldman MS. Up close and personal: temporal variability in the drinking of individual college students during their first year. *J Consult Clin Psychol*. 2004;72:155–164.
29. Harford TC, Wechsler H, Seibring M. Attendance and alcohol use at parties and bars in college: a national survey of current drinkers. *J Stud Alcohol*. 2002;63:726–728.
30. Lee CM, Maggs JL, Rankin LA. Spring break trips as a risk factor for heavy alcohol use among first-year college students. *J Stud Alcohol*. 2006;67:911–916.

31. Neal DJ, Fromme K. Hook 'em horns and heavy drinking: alcohol use and collegiate sports. *Addict Behav*. 2007;32:2681–2693.
32. Rutledge PC, Park A, Sher KJ. 21st birthday drinking: extremely extreme. *J Consult Clin Psychol*. 2008;76:511–516.
33. Crawford LA, Novak KB. Alcohol abuse as a rite of passage: the effect of beliefs about alcohol and the college experience on undergraduates' drinking behaviors. *J Drug Educ*. 2006;36:193–212.
34. Harford TC, Yi HY, Hilton ME. Alcohol abuse and dependence in college and non-college samples: a ten-year prospective follow-up in a national survey. *J Stud Alcohol*. 2006;67:803–809.
35. Lanza ST, Collins LM. A mixture model of discontinuous development in heavy drinking from ages 18 to 30: the role of college enrollment. *J Stud Alcohol*. 2006;67:552–561.
36. Curran PJ, Muthen BO, Harford TC. The influence of changes in marital status on developmental trajectories of alcohol use in young adults. *J Stud Alcohol*. 1998;59:647–658.
37. Miller-Tutzauer C, Leonard KE, Windle M. Marriage and alcohol use: a longitudinal study of "maturing out." *J Studies Alcohol*. 1991;52:434–440.
38. Homish GG, Leonard KE. The drinking partnership and marital satisfaction: the longitudinal influence of discrepant drinking. *J Consult Clin Psychol*. 2007;75:43–51.
39. Homish GG, Leonard KE, Cornelius JR. Illicit drug use and marital satisfaction. *Addict Behav*. 2008;33:279–291.
40. Mudar P, Leonard KE, Soltysinski K. Discrepant substance use and marital functioning in newlywed couples. *J Consult Clin Psychol*. 2001;69:130–134.
41. Collins RL, Ellickson PL, Klein DJ. The role of substance use in young adult divorce. *Addiction*. 2007;102:786–794.
42. Cnattingius N. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob Res*. 2004;6: S125–S140.
43. Centers for Disease Control and Prevention. Alcohol use among women of childbearing age-United States, 1991–1999. *MMWR*. 2002;51:273–276.
44. Simoni-Wastila L, Yang HK. Psychoactive drug abuse in older adults. *Am J Geriatric Pharmacother*. 2006;4:380–394.
45. Administration on Aging. A profile of older Americans, 2008. Washington, DC: Department of Health and Human Services. http://www.aoa.gov/AoARoot/Aging_Statistics/Profile/2008/4.aspx. Accessed August 20, 2009.
46. Colliver JD, Compton WM, Gfroerer JC, Condon T. Projecting drug use among aging baby boomers in 2020. *Ann Epidemiol*. 2006;16:257–265.
47. Perreira KM, Sloan FA. Life events and alcohol consumption among mature adults: a longitudinal analysis. *J Stud Alcohol*. 2001;62:501–508.
48. Abrams S. Inebriated elders: the problem of substance abuse among the elderly. *Elder Law J*. 2001;9:229–241.
49. Lay K, King LJ. Late onset of prescription drug abuse or dependence among older adults: implications for treatment. *Adv Soc Work*. 2007;8:264–274.
50. Benshoff JJ, Harrawood LK. Substance abuse and the elderly: unique issues and concerns. *J Rehabilitation*. 2003;69:43–48.
51. Downling GJ, Weiss SR, Condon TP. Drugs of abuse and the aging brain. *Neuropsychopharm*. 2008;33:209–218.
52. Oesterle S, Hill KG, Hawkins JD, Guo J, Catalano RF, Abbott RD. Adolescent heavy episodic drinking trajectories and health in young adulthood. *J Stud Alcohol*. 2004;65:204–212.

53. Schulenberg JE, Merline AC, Johnston LD, O'Malley PM, Bachman JG, Laetz VB. Trajectories of marijuana use during the transition to adulthood: the big picture based on national panel data. *J Drug Issues*. 2005; 35:255–279.
54. Glantz MD, Conway KP, Colliver JD. Drug abuse heterogeneity and the search for subtypes. In: Sloboda Z, ed. *Epidemiol Drug Abuse*. New York: Springer; 2005: 15–27.
55. Neighbors C, Walters ST, Lee CM, et al. Event-specific prevention: addressing college student drinking during known windows of risk. *Addict Behav*. 2007;32: 2667–2680.
56. Suozzi T. Young and suburban, and falling for heroin. *New York Times*. September 27, 2009:WE 1, 10.

About the Editor

MADELON L. FINKEL, Ph.D. is Professor of Clinical Public Health and Director of the Office of Global Health Education at the Weill Cornell Medical College in New York City. She also is Course Director in the Department of Public Health, coordinating the Epidemiology/Biostatistics, Evidence-Based Medicine, and Health Care Systems courses as well as the Public Health Clerkship. Dr. Finkel's interests focus on women's health issues. Research projects include a cervical cancer screening program in rural India, reduction of preterm birth in Brazil, and better management of hypertension in Peru. She has served as consultant to numerous organizations including law firms, pharmaceutical companies, and health care organizations. She also serves on several Boards of organizations whose mission is to promote public health. Her book, *Understanding the Mammography Controversy: Science, Politics, and Breast Cancer Screening* was published by Praeger 2005. Praeger also published her book, *Truth, Lies, and Public Health: How We Are Affected When Science and Politics Collide* in 2007.

This page intentionally left blank

About the Contributors

Brandon Aden, MD, MPH, is an Empire Clinical Research Investigator Program (ECRIP) Fellow at Weill Cornell Medical College in New York City. Dr. Aden's research interests include clinical epidemiology of hepatitis C, treatment and prevention with a specific interest in acute hepatitis C infection, and hepatitis C infection among persons on methadone maintenance therapy and among active intravenous drug users.

Jason C. Baker, MD, is an assistant professor of medicine (endocrinology) at Weill Cornell Medical College in New York City. His research interests include disease management through education and lifestyle interventions, prevention of Type 1 diabetes, and the impact of diabetes on international health.

Catherine W. Barber, MPA, directs public health practice activities at the Harvard School of Public Health's Injury Control Research Center. She is interested in developing surveillance systems for intentional injury, training professionals and lay people on how to use data from these systems to improve public policy, and bridging the gap between research and practice to reduce suicide and violence. She cofounded the National Center for Suicide Prevention Training.

Richard Beasley, MD, MBChB, FRACP, DM, FAAAAI, FRCP, DSc, is a physician at Capital & Coast District Health Board and director of the Medical Research Institute of New Zealand. He is an adjunct professor at the Universities of Otago and Canterbury, Victoria University of Wellington, New Zealand, and visiting professor at the University of Southampton in the United Kingdom. His research interests are in the fields of respiratory medicine and pharmacovigilance.

Marianne Berwick, PhD, MPH, is professor and chief, Division of Epidemiology, associate director of the Cancer Research and Treatment Center (CRTC), and coleader of the Population Health Sciences, Cancer Health Disparities, and Cancer Control Program at the University of New Mexico School of Medicine, Albuquerque, New Mexico.

Victoria S. Blinder, MD, MPH, MS, is an assistant attending physician in the Health Outcomes Research Group, Department of Epidemiology and Biostatistics, and Breast Cancer Medicine Service, Department of Medicine, at Memorial Sloan-

Kettering Cancer Center in New York City. She is a medical oncologist specializing in breast cancer.

William B. Borden, MD, MBA, is an assistant professor of medicine and public health at the Weill Cornell Medical College in New York City and the Nanette Laitman Clinical Scholar in the department of public health. He is a practicing preventive cardiologist and health policy researcher. His current areas of research interest are in the real-world implementation of appropriate medical therapy prior to coronary interventions and in the socioeconomic equity of proposed hospital value-based purchasing programs.

Abraham Bornstein, MD, is formerly a clinical assistant professor of medicine in public health at the Weill Cornell Medical College in New York City where he also held faculty positions at the Weill Cornell Divisions of Pediatric Cardiology, Cardiothoracic Surgery, and Cardiovascular Pathophysiology as well as in the Department of Medicine at Columbia University College of Physicians and Surgeons. He is a board certified internist and practicing cardiologist.

Paolo Cassano, MD, PhD, is a psychiatrist and researcher at Massachusetts General Hospital (MGH) in Boston, Massachusetts, as well as a full-time attending at North Suffolk Mental Health Association in East Boston and Chelsea. He also is a member of the Depression Clinical and Research Program at MGH where he studies depression and resistance to treatment as a co-investigator in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) trial.

Bernard M. Y. Cheung, MD, PhD, FRCP, is affiliated with the School of Clinical and Experimental Medicine, College of Medical and Dental Services, of the University of Birmingham, United Kingdom, and professor in the Department of Medicine at the University of Hong Kong.

Rubin S. Cooper, MD, was professor of clinical pediatrics and the Starr Foundation-David Wallace Chair of Clinical Pediatric Cardiology at the Komansky Center for Children's Health at Weill Cornell Medical College. He is presently chief of pediatric cardiology at the Schneider Children's Hospital on Long Island, New York. His research interest includes coronary artery disease surveillance and the noninvasive function in the pediatric population.

Sandra Demars, MD, graduated from Weill Cornell Medical College in New York City. She is presently a Resident in Internal Medicine in Boston, MA.

Elena B. Elkin, PhD, is an assistant attending outcomes research scientist Health Outcomes Research Group, Department of Epidemiology and Biostatistics Memorial Sloan-Kettering Cancer Center in New York City. Her area of expertise is in cancer health services and outcomes, decision analysis, and cost-effectiveness analysis. Her research addresses screening and treatment decisions in breast cancer, prostate cancer, and colorectal cancer.

Esther Erdei, PhD, is a member of the Division of Epidemiology and Biostatistics at the University of New Mexico, School of Medicine, Albuquerque, New Mexico.

Maurizio Fava, MD, is executive vice chair of the Department of Psychiatry, director of the Depression Clinical and Research Program, and executive director of the Clinical Trials Network and Institute all at the Massachusetts General Hospital in

Boston, Massachusetts. He also is professor of psychiatry at the Harvard Medical School.

Kristina Flores, PhD, is a member of the Division of Epidemiology and Biostatistics at the University of New Mexico, School of Medicine, Albuquerque, New Mexico.

Melissa Gonzales, PhD, is a member of the Division of Epidemiology and Biostatistics at the University of New Mexico, School of Medicine in Albuquerque, New Mexico.

Kenneth W. Griffin, PhD, MPH, is professor of public health at Weill Cornell Medical College in New York City. His research interests focus on the etiology and prevention of substance abuse and HIV risk behaviors among adolescents and young adults.

Helina Kassahun, MD, is a clinical cardiology fellow at New York-Presbyterian Hospital in New York City.

Michael W. Kattan, PhD, MBA, is chairman, Department of Quantitative Health Sciences at the Cleveland Clinic. He also is professor of medicine at the Cleveland Clinic and professor of epidemiology and biostatistics in the School of Medicine at Case Western Reserve University, Cleveland, Ohio. His research focuses on medical decision-making and medical informatics, with a focus on urologic oncology.

Namory Keita, MD, is head of the Department of Obstetrics and Gynecology at the Donka Teaching Hospital in Conakry Guinea and coordinator of the French Speaking Training Center in Gynecologic Cancer Prevention in Conakry Guinea. He is involved in population-based prospective studies evaluating alternative approaches for the early detection and control of cervical cancer.

Stephen Lyman, PhD, is an associate professor of public health at Weill Cornell Medical College and director of epidemiology and biostatistics at the Hospital for Special Surgery in New York City. He conducts policy-relevant orthopedic medical device comparative effectiveness research.

Madhu Mazumdar, PhD, is professor of biostatistics in public health and chief, Division of Biostatistics and Epidemiology, in the Department of Public Health at Weill Cornell Medical College in New York City.

Matthew Miller, MD, MPH, ScD, is associate director of the Harvard Injury Control Research Center and is a physician with training in internal medicine, medical oncology, medical ethics, injury prevention, epidemiology, and health policy. Dr. Miller has been teaching a course on a public health approach to suicide prevention at the Harvard School of Public Health, Cambridge, MA.

Paulo Naud, MD, PhD, is professor of obstetrics and gynecology at the Federal University of Rio Grande do Sul in Brazil. He is coordinator of a capacitating course on screening, diagnostic and prevention of cervical cancer for physicians and nurses in Rio Grande do Sul State.

Twalib A. Ngoma, MD, is executive director of the Ocean Road Cancer Institute in Dar es Salaam, Tanzania and is head of the International Network for Cancer Treatment and Research in Tanzania.

Carvell T. Nguyen, MD, is chief residentat at the Clickman Urological and Kidney Institute and the Cleveland Clinic Foundation in Cleveland, Ohio.

Joseph T. Nguyen, MPH, is a clinical research methodologist at the Hospital for Special Surgery in New York City. He primarily performs research in device and procedure evaluation in sports medicine.

Kwok Leung Ong, MPhil, is currently a University of Hong Kong Postgraduate Fellow in the Department of Medicine and the Research Centre of Heart, Brain, Hormone, and Healthy Ageing of the University of Hong Kong. His research interests include the epidemiology and genetics of diabetes, hypertension, and the metabolic syndrome.

Thomas A. Peterman, MD, PhD, MSc, is chief, Field Epidemiology Unit, Epidemiology and Surveillance Branch, Divison of Sexually Transmitted Disease Prevention, at the Centers for Disease Control and Prevention in Atlanta, Georgia. His research has focused on infectious diseases, especially HIV, and sexually transmitted diseases.

Rengaswamy Sankaranarayanan, MD, is head of the early detection and prevention and screening group at the International Agency for Research on Cancer in Lyon France. He is involved in a number of population-based, prospective randomized studies evaluating various approaches for the early detection and control of breast, cervical, and oral cancers.

Philippa Shirtcliffe, MD, MBChB, FRACP, is a physician at Capital and Coast District Health Board and Senior Research Fellow at the Medical Research Institute of New Zealand. Her research interests are in the fields of respiratory medicine.

Enrico Tedeschini, MD, is presently a nonclinical research fellow within the Massachusetts General Hospital Department of Psychiatry, Boston, Massachusetts.

Somanathan Thara, MD, is an associate professor of Pathology at the Regional Cancer Center in Trivandrum, India. She is a major investigator in randomized clinical trials of breast, cervical, and oral cancer screening in India.

Salina Torres, PhD, is a member of the Division of Epidemiology and Biostatistics at the University of New Mexico, School of Medicine in Albuquerque, New Mexico.

Elizabeth A. Torrone, MSPH, PhD, is a member of the Epidemic Intelligence Service of the Centers for Disease Control and Prevention, Division of Sexually Transmitted Disease Prevention in Atlanta, Georgia.

Lyndal Trevena, MBBS (Hons), MPhilPH, PhD, is associate professor of public health and associate dean (international) and director of the Office for Global Health at the University of Sydney (Australia) School of Medicine.

Index

This page intentionally left blank

Public Health in the 21st Century

This page intentionally left blank

Public Health in the 21st Century

Volume 3: Current Issues in Public Health Policy

Madelon L. Finkel, Editor

Foreword by David J. Skorton, MD

AN IMPRINT OF ABC-CLIO, LLC
Santa Barbara, California • Denver, Colorado • Oxford, England

Library of Congress Cataloging-in-Publication Data

Public health in the 21st century / Madelon L. Finkel, editor ; foreword by David J. Skorton.
p. ; cm.
Public health in the twenty-first century
Includes bibliographical references and index.
ISBN 978-0-313-37546-0 (hard copy : set : alk. paper) — ISBN 978-0-313-37548-4 (hard copy : vol. 1 : alk. paper) — ISBN 978-0-313-37550-7 (hard copy : vol. 2 : alk. paper) — ISBN 978-0-313-37552-1 (hard copy : vol. 3 : alk. paper) — ISBN 978-0-313-37547-7 (ebook : set) — ISBN 978-0-313-37549-1 (ebook : vol. 1) — ISBN 978-0-313-37551-4 (ebook : vol. 2) — ISBN 978-0-313-37553-8 (ebook : vol. 3)
1. Public health. 2. World health. 3. Health policy.
I. Finkel, Madelon Lubin, 1949– II. Title: Public health in the twenty-first century.
[DNLM: 1. Public Health. 2. Disease Management. 3. Health Policy. 4. World Health. WA 100]
RA424.8.P83 2011
362.1—dc22 2010039064

ISBN: 978-0-313-37546-0
EISBN: 978-0-313-37547-7

15 14 13 12 11 1 2 3 4 5

This book is also available on the World Wide Web as an eBook.
Visit www.abc-clio.com for details.

Praeger
An Imprint of ABC-CLIO, LLC

ABC-CLIO, LLC
130 Cremona Drive, P.O. Box 1911
Santa Barbara, California 93116-1911

This book is printed on acid-free paper ♾

Manufactured in the United States of America

To my husband Arnold,
whose ideas, insight, and, most of all,
support and love are so important to me.

This page intentionally left blank

Contents

Foreword

As we enter the second decade of the 21st century, myriad issues compete for the world's attention, from the continuing stresses of the deep and widespread recession, to personal and national security, to climate change. But no issue looms larger than that of public health. Within this general rubric fall multiple issues critical to the individual, the country, and the world. Infectious diseases, including malaria, multiple drug–resistant tuberculosis, pandemic influenza, and HIV/AIDS pose even greater threats because of enormously increased international travel. In the developing world, the traditional diseases of poverty—communicable disease, especially infectious diarrhea and other waterborne diseases, malnutrition, and inadequate maternal and child health care—and displacement and violence, which are the sequelae of political instability, increasingly are being joined by the ailments of excess such as obesity, diabetes, and atherosclerotic cardiovascular disease, making the burden infinitely greater. Long overdue recognition of the worsening plight of women, particularly but not only in resource-poor environments, compounds the ongoing dilemmas of maternal-child health. Even within rich societies, such as the United States, shocking health disparities stubbornly continue.

Despite these daunting challenges, the tools of public health are more robust than ever. In addition to the traditional tools of medicine and the social sciences, the use of molecular genetics techniques and advanced statistical analysis presents new opportunities for the student and practitioner of public health. This comprehensive work on public health thus appears at a most opportune time.

Including a carefully assembled combination of original work and important recent literature and covering a huge sweep of relevant problems, *Public Health in the 21st Century* succeeds admirably in bringing together much of the broad field into one work that should find its place as a reference for public health workers and academics as well as policy makers and those in the private sector, whether health care providers, insurers, or drug or device manufacturers. Dr. Madelon Finkel, an experienced and recognized expert in several aspects of public health and, importantly, in the pedagogy of public health, has assembled a most impressive group of writers on a huge variety of public health topics, covering everything from global population health, to special needs cohorts, to

health care policy, to the often-ignored topic of public health teaching strategies and tactics. Readers from across the spectrum of public health concerns will find thought-provoking material of great value.

I commend Professor Finkel and her many colleagues on bringing to fruition a work that undoubtedly will receive wide use.

David J. Skorton, MD
President, Cornell University

Acknowledgments

This three-volume set could not have been produced without the contributions of the authors who so generously took the time to research and write their respective chapter. Most of the authors are my friends and colleagues who gladly agreed to accept my invitation to be included in the effort. I thank each of the authors for their time, effort, and especially their friendship.

My editorial assistant, Sophia Day, was tremendously helpful in organizing the huge volume of material and keeping track of missing information. Editorial reviews of many of the chapters were graciously and professionally done by Dr. Rebecca Finkel, an author and former editor who also happens to be my accomplished daughter. Technical computer work and support was provided by Jean Policard of the Department of Public Health at Weill Cornell Medical College. His assistance was invaluable to me.

Many thanks to my editor, Debbie Carvalko of Praeger, who invited me to write this multivolume text. Her support and faith in my being able to deliver the goods on time was reassuring.

This page intentionally left blank

Introduction

Madelon L. Finkel, PhD

Compiling topics for inclusion in a multivolume text on public health at first seemed like a simple task. Because so many significant advances have been made in disease prevention and health promotion, and so many public health initiatives have been put in place over the years to improve health and well-being, deciding which topics to select proved more difficult than anticipated. Which ones should be included? Which ones are the most relevant, the most important to highlight? Narrowing the focus, but being as comprehensive and inclusive as possible, seemed the most prudent way to proceed. And, therein lay the problem. How was I to select from such a wide array of public health issues to produce a comprehensive text on current public health topics? In an effort to be both comprehensive and inclusive, I endeavored to select as many important and timely subjects as possible for these three volumes. For fear of overwhelming the readers with chapters on every conceivable public health issue, a careful selection was made to highlight topics that represent and reflect the field of public health's breadth and scope. As such, the three volumes include chapters on topics reflecting advances and progress in knowledge and practice as well as challenges that remain. Naturally, many more topics could have been included. The essays selected for inclusion, many written specifically for this multivolume set and others reprinted from the published literature, represent a broad overview of important public health issues in the 21st century.

Charles-Edward A. Winslow, a bacteriologist and professor of public health at the Yale School of Medicine from 1915 to 1945, proposed a definition of public health as

> the science and art of preventing disease, prolonging life, and promoting physical health and efficiency through organized community efforts for the sanitation of the environment, the control of community infections, the education of the individual in principles of personal hygiene, [and] the organization of medical and nursing service for the early diagnosis and preventive treatment of disease.[1]

To a large extent, his definition has not been changed or amended over the ensuing decades. Public health's focus, then and now, is to safeguard the public's health and to handle threats to public health.

Public health has its roots in antiquity. It has long been recognized that polluted water and air, inadequate waste disposal, overcrowding and a concomitant lack of hygiene, and lifestyle behavior contributed to the spread of disease. Moving away from the miasma theory of disease, which argues that most diseases are caused by miasma (Greek for "pollution," that is, a noxious form of "bad air"), to the germ theory of disease was an important step in disease prevention. The 19th century witnessed so many discoveries and advances in the fields of medicine and public health. Essentially, pre-20th-century efforts focused on the eradication of infectious diseases and improvements in hygiene and sanitation, which led to a dramatic increase in average life expectancy. For example, the science of epidemiology probably dates from Dr. John Snow's identification of polluted public water wells as the source of the 1854 cholera outbreak in London. Hungarian physician Ignaz Semmelweis, for example, successfully reduced infant mortality at a Vienna hospital by instituting a disinfection procedure. His findings were published in 1850, but his work was ill received by his colleagues, who unwisely discontinued the procedure. Disinfection did not become widely practiced until British surgeon Joseph Lister "discovered" antiseptics in 1865, helped significantly by the work of the French chemist and microbiologist Louis Pasteur and German physician and bacteriologist Robert Koch.

The early 20th century expanded the scope and complexity of public health concerns. High rates of infant mortality led to the establishment of maternal and child health programs that emphasized nutrition. The disgraceful state of the food processing industry was notably depicted in Upton Sinclair's book, *The Jungle*. The book dealt with conditions in the U.S. meat-packing industry, causing a public uproar that partly contributed to the passage of the Pure Food and Drug Act and the Meat Inspection Act in 1906. High rates of occupational injuries and occupational-related diseases led to programs for industrial hygiene and occupational health, but it was not until 1970 that the U.S. Occupational Safety and Health Administration (OSHA) was created by Congress under the Occupational Safety and Health Act. Its mission is to prevent work-related injuries, illnesses, and occupational fatality by issuing and enforcing rules called standards for workplace safety and health.

These and other public health efforts contributed substantially to a dramatic decrease in mortality. From 1900 to 1940, for example, mortality rates in the U.S. fell by 40 percent, and life expectancy at birth increased from 47 years to 63 years.[2] No other period in American history showed such a dramatic decline in overall death rates. Nearly all of this decrease can be accounted for by reductions in infectious diseases, which in 1900 accounted for 44 percent of deaths. Contributing to the decrease in infectious diseases was the implementation of clean water technologies, one of the most important public health interventions of the early 20th century. At the turn of the century, waterborne diseases accounted for one-quarter of reported infectious disease deaths in urban areas. By 1936, less than 20 percent of deaths were due to infectious diseases. Perhaps the greatest public health feat was the worldwide eradication of smallpox, a

highly contagious, serious viral disease that was a worldwide scourge. The last case was recorded in 1977 in Somalia, and the eradication was certified by the World Health Organization (WHO) in 1979.

By the mid- to late-20th century, achievements saw a shift in focus from acute infectious diseases to the treatment and prevention of the growing burden of noninfectious, chronic diseases. English physician-researcher Sir Richard Doll and English epidemiologist and statistician Sir A. Bradford Hill pioneered the randomized clinical trial, and together were the first to demonstrate the connection between cigarette smoking and lung cancer. The focus on individual behaviors and risk factors (for example, antismoking campaigns) was an important step in addressing the *social determinants of disease*. The *new* public health sought to address the burden of chronic disease in a more comprehensive way by focusing on the effects of disease on vulnerable populations (for example, the elderly, the young, and the disabled), how health status differs among population groups (health inequalities), and how health care systems are organized and financed. Indeed, the challenges facing modern public health in the 21st century must be broad and inclusive, and focus on improvement in population health through the reduction of preventable diseases, both communicable and noncommunicable.

Looking back over the last century, public health is credited with adding 30 years to the life expectancy of people in the United States over the course of the 20th century; 25 years of this gain are attributable to advances in public health.[3] The Centers for Disease Control and Prevention (CDC) cataloged 10 of what it considered to be the most notable public health achievements based on the opportunity for prevention and the impact on death, illness, and disability.[4] These include (not ranked by order of importance) the following:

- Vaccination programs (as a result of widespread vaccine use, many of the infectious diseases that once killed so many have been almost eliminated);
- Fluoridation of drinking water (fluoride was first added to the public water system in 1945; tooth decay and tooth loss has declined substantially as a result);
- Occupational safety policies (since 1980, the rate of fatal occupational injuries has decreased by 40 percent);
- Access to safe, improved family planning and contraceptive services;
- Control of infectious diseases as a result of antibiotics, clean water, and improved sanitation;
- Food safety (safer and more healthful foods can be attributed to decreased microbial contamination and increased nutritional;
- Recognition of tobacco use as a health hazard;
- Motor vehicle safety (safety belts, child safety seats, motorcycle helmets, and engineering improvements in both vehicles and highways have helped reduce fatal motor vehicle accidents);
- Decline in deaths from coronary heart disease and stroke (lifestyle modifications and pharmaceuticals have led to a decline in deaths for these diseases); and
- Healthier mothers and babies as a result of better hygiene, prenatal health care and nutrition.

The 21st century presents new challenges. Largely preventable infectious diseases, such as tuberculosis (TB), polio, measles, and cholera continue to

plague millions of people around the globe, especially children. HIV/AIDS, which appeared on the scene in the mid-1980s, continues to be a major public health problem, although antiretroviral medications have done wonders in terms of extending life. Malaria, multidrug-resistant TB, and global outbreaks of viral diseases, most recently the H1N1 swine flu pandemic of 2009, continue to challenge public health efforts. That being said, chronic diseases such as diabetes and heart disease are now prevalent around the world. Obesity is not just a problem of the wealthy nations, as the increase in adult and childhood obesity in the developing world threatens to jeopardize progress. Statistics compiled by WHO show that chronic diseases are the largest cause of death in the world today and that global prevalence of all the leading chronic diseases is increasing, with the majority occurring in developing countries. Cardiovascular disease is already the leading cause of mortality in the developing world.[5] The increased burden of chronic diseases in countries that also have a high burden of infectious diseases is creating both a tremendous economic and public health strain. Furthermore, the recognition that health is affected by many factors, including genetics, economics, ethnicity and race, and geography, has necessitated a shift in focus in thinking. Public health in the 21st century must address these health inequalities to reduce the incidence of disease and improve health and well-being. Malnutrition, poverty, lack of access to health care, and so forth threaten to undermine the progress made in disease control and prevention.

In summary, over the past 150 years, much of the focus has been on disease control, understanding sources of contagion, and implementing programs to prevent the spread of disease. As scientific knowledge grew, public health's purview expanded to include maternal and child health care, health education, nutrition, aging of the population, the recognition of the role of behavioral factors in determining health, the impact of violence (domestic, civil, and international), health care disparities, and globalization. Indeed, increased globalization and technological advances have contributed to a worldwide economic, political, and social interdependence. In 1945, the United Nations Conference in San Francisco unanimously approved the establishment of a new, autonomous international health organization, the WHO, which came into being on April 7, 1948. The WHO was established as a specialized agency of the United Nations to serve as a coordinating authority on international public health issues.

Despite the progress made in improving the health, so much still remains to be done, not just in the United States, but also globally. In 2000, for example, 11.1 million children under the age of 5 died from preventable diseases such as diarrhea and acute respiratory infection.[6] These and other primarily preventable diseases kill more people each year than conflicts alone. Worldwide, poverty is one of the most significant causes of preventable mortality. Gender inequality persists and perhaps in some areas of the world actually has increased. Population growth remains a serious concern as the world's population has surged to 6.7 billion, most of the increase occurring within the last century. Environmental degradation and climate concerns have significant health implications. Each has a dimension that necessarily involves public health. As such, public health must be looked at in a global context if it is to be successful in fulfilling its mandate. Microbes have no boundaries, and we have seen over and

over again, localized outbreaks can quickly spread to national epidemics, and even worldwide pandemics.

Global health refers to health problems that transcend national borders and are of such magnitude that they have a global political, social, and economic impact. Assessing and measuring the impact of globalization on population health status should not be done in a vacuum; a global public health perspective needs to be integrated into health, social, and economic policies and programs to be effective. Reducing social and economic deprivation, reducing health inequalities, and improving health status go hand-in-hand. Domestic and international entities whose function and purpose is promulgating public health policy need to work together to achieve common goals. Collectively, progress can be made; individually, the effect is more muted. At their summit in 2000, heads of state of the G8 countries went on record as recognizing health as a global challenge and acknowledging that health is the "key to prosperity" and that "poor health drives poverty."[7] Following up on this challenge, G77 heads of state from 130 developing countries also expressed support for working toward the reduction of disease worldwide.[8] The motives and intent are laudable, but a decade later, the world still finds itself grappling with disease control and health prevention issues. Indeed, a WHO assessment of the capacity of 185 countries to prevent, conduct surveillance, and control disease showed that while health ministries had a high level of awareness of the issues, they had little or no allocation of significant resources to address the problems.[9]

Global nongovernmental organizations (NGOs) have played and continue to play a critical role in building capacity and sustainable development in specific areas of the world, although often the focus is narrow (for example, tobacco, TB, malaria, diet and nutrition, and so forth). Foundations, such as the Bill and Melinda Gates Foundation, have provided extraordinary sums of money and manpower to address pressing global health issues. Their importance and impact cannot be denied or ignored. These private investments in global health far exceed government assistance. The pharmaceutical industry also has the potential for being an effective player in the global health arena, but the industry is constantly criticized for not taking a greater role in the access to life-saving drugs, particularly in developing nations. The arguments are plentiful, ranging from focusing research and development efforts on health issues for rich countries to pricing drugs at unaffordable levels. That being said, despite the challenges that the pharmaceutical industry faces, it has been involved in a number of global health initiatives. Often, this means that companies are donating drugs, cutting prices, and developing partnerships with local governments and NGOs.

The World Bank and the United Nations play a major role in setting priorities for global health. The World Bank recognizes the negative effect of the increasing burden of disease, especially on the poor. Billions of dollars have been provided to countries for disease prevention. These efforts are crucially important as most developing countries have inadequate financing, lack of manpower, and poor infrastructure. Numerous UN organizations are specifically designed to focus on the global burden of disease. For example, the United Nations Population Fund, the United Nations Children's Fund, and many other UN agencies and organizations focus on providing assistance to the poorest

countries in an effort to "make a difference." The Millennium Development Goals (MDGs), also, are an agreed-upon set of goals that were developed in response to the world's main development challenges. They were drawn from the actions and targets contained in the Millennium Declaration that was adopted by 189 nations and signed by 147 heads of state and governments during the UN Millennium Summit held in September 2000. The MDGs are targeted to promote poverty reduction, education, maternal health, and gender equality, and to combat child mortality, AIDS, and other diseases. Poor countries pledged to govern better and invest in their people through health care and education. Rich countries pledged to support them, through aid, debt relief, and fairer trade. The MDGs represent a global partnership that has grown from the commitments and targets established at the world summits of the 1990s. The eight goals are to be achieved by 2015.[10]

Many stakeholders, public and private, are working toward similar goals, but it seems at times as if progress has taken one step forward and two steps back. While tremendous progress has been made over the past century, substantial challenges remain. Capacity development for prevention, treatment, and research remains weak; global economic factors impede progress; and the need for health systems change (delivery, financing, organization, and insurance coverage) remains unmet. The three volumes in this text were formulated to address public health issues from a national and global perspective. Volume 1 focuses on global population health issues, while volume 2 presents chapters on various aspects of determinants of health and disease, and volume 3 examines current public health policy issues, including ethics and human rights, public health education, and challenges we face as we enter the second decade of the 21st century. I made a concerted effort to include authors from around the world. Colleagues from Africa, Australia, Canada, China, Europe, India, Latin America, and New Zealand are well represented in this multivolume text. Their perspective and insight add a global dimension to the set.

VOLUME 1

Section 1 of volume 1 focuses on global population health issues. In their chapter on the global burden of disease, Kishore and Michelow carefully review the salient features of the global burden of disease, including its distribution and changing patterns over time. If current trends continue, diseases such as diarrhea, AIDS, TB, and malaria will become less important causes of morbidity and mortality as heart disease, cancer, diabetes, and traffic accidents increase in prevalence. Although the "burden" of a disease can be defined in a variety of senses, the consensus definition, particularly from the WHO, is a fairly specific one. The global burden of disease (GBD) as defined by the WHO is a comprehensive regional and global assessment of mortality and disability from 136 diseases and injuries and 19 risk factors. While useful, the thinking was that a better measurement of the GBD was needed, one that integrated morbidity, mortality, incidence, and prevalence into a single common metric that can be compared across time, space, and interventions. A new metric, the Disability-Adjusted Life Year (DALY), is a summary measure of population health, measured in units of time (years), combining estimations of both fatal and nonfatal health outcomes

(morbidity and mortality) to provide an estimate of the number of years of fully healthy life lost by an individual with a particular illness or condition. When DALYs are used to estimate the GBD, communicable diseases displace noncommunicable diseases as leading drivers of illness. The authors discuss the explosion of noncommunicable chronic diseases worldwide and the existing burden of communicable diseases, the combination of which poses a significant threat to the public's health. The challenge we face is how to best deal with the double burden of disease.

China and India together account for 37 percent of the world's population, about 6.8 billion. In 2025, India will surpass China in total population. India thus will have the distinction of being the world's most populous country. China's fertility rate is decreasing, whereas India's continues to increase. The United States is the third most populous country. Bongaarts focuses on population projections to the year 2050 for the world and major regions, and then identifies the demographic factors responsible for continued expansion of human numbers. Discussion focuses on policy options for slowing population growth in the developing world, where the growth continues to surpass that in the industrial world. Four main demographic factors contribute to future population growth: continued high fertility, declining mortality, young age structure, and migration. Bongaarts concludes by noting that the unprecedented pace at which the world's population has grown over recent decades has had an adverse impact on social and economic development, on health care, and on the environment. Despite substantial and partially successful efforts to reduce growth in the less developed countries, this expansion of human members is expected to continue at a rapid pace over the next decades with nearly all of this growth occurring in Africa, Asia, and Latin America. He advocates for three key strategies to reduce this growth rate: strengthen family planning programs to provide women with the knowledge and means to regulate their fertility; emphasize "human development," in particular education, gender equality, and child health; and encourage delays in subsequent childbearing.

The effect of urbanization on the public's health is discussed by Galea and Vlahov. The authors focus on the substantial change from how most of the world's population lives, reflecting on how the characteristics of the urban environment affect population health. The key factors affecting health in cities are considered within three broad themes: the physical environment, the social environment, and access to health and social services.

Continuing the theme of adverse effects of rapid population growth and urbanization, Brown and DeGaetano present a scholarly piece on the consequences of climate change on health. Concerns about recent changes in global climates and possible future trends on the health of the world's population are now considered important policy topics. With the election of Barack Obama, who has pledged a new era of leadership and responsibility to reduce the serious negative effects of climate degradation, the United States resumed its leading role in combating climate change and the adverse effects thereof. The United States is the world's largest source of cumulative emissions in the atmosphere, and as such, needs to lead the way for other nations to make a serious effort on climate change. Brown and DeGaetano make the case that climatic changes have, and will continue to have, direct negative health effects from altered weather patterns, but state

also that the indirect effects on agriculture and wider population systems are important factors for the GBD. Global warming (that is, melting of Arctic ice), extreme weather (for example, heat waves, cold spells), flooding (for example, Hurricane Katrina), erosion of ocean coastlines (that is, a result of extreme and heavy precipitation), and drought (for example, dust bowls) are leading to a disruption of food production and to disease. The authors note that exposure to infectious diseases has altered because of changes in temperature, humidity, rainfall, and sea-level rise. Specifically, some evidence of changes in the distribution of mosquito, tick, and bird vectors has been attributed to climate change. Mosquitoes, for example, can transmit diseases, such as malaria, dengue, yellow fever, and Japanese encephalitis, but their sensitivity to weather conditions can inhibit or enhance their efficacy as a vector. Malaria is spread by mosquitoes, which are inhibited from transmitting the parasite in cooler temperatures. Air quality and pollutants are affected by the weather and climate, and can cause negative health effects; the incidence of asthma has soared over the past decades. The authors caution that vulnerability to climate change will depend on responses to prevention, adaptation, adaptive capacity, mitigation, and future advances in disease control. It is clear that doing nothing will only make the situation worse than it already is.

The issue of global health and nutrition is a complex interplay of many factors ranging from politics to economics to food production policies to environmental degradation. Food is a basic human need. With roughly 1 billion humans suffering from overnutrition and a similar number unable to find enough food to subsist, no one seriously disagrees about the urgency of world hunger. One in six individuals does not get enough food to be healthy and to lead an active life. Hunger and malnutrition adversely affect physical and mental development; indeed, one might argue that hunger and malnutrition are leading risks to the health and well-being of individuals worldwide. Davison presents a comprehensive overview of the salient issues and focuses in particular on the interdependence of nutrition, economic development, and health. His discussion of the topic includes an assessment of the MDGs drafted to address the issue of alleviating hunger and malnutrition and a brief overview of some of the programs designed to eliminate global nutrition disparities, including the Millennium Village Projects, the Grameen enterprises and "microcredit" initiatives in resource-poor countries, and the role that foundations play in providing the financial means to reduce poverty and, in turn, to alleviate hunger and malnutrition.

No matter how one defines "health," prevention and treatment of disease is an essential prerequisite for achieving health and well-being. Implicit in this is that the right medicine be available at the time and place of need. Reidenberg presents an overview of the WHO Essential Medicines Program. By definition, an "essential drug" is a drug needed to satisfy the health needs of the majority of the population. The essential drugs concept of purchasing a limited list of essential drugs for a health service and making them generally available has been accepted by 156 countries and most, if not all, donor organizations. The essential medicines idea was developed to help limited resource countries make choices to use their medical resources for the greatest good for the greatest number. Thirty years after the WHO initiated the Essential Medicines Model List, four out of five countries have adopted a national essential medicines list. More than 100 countries have a

national drug policy in place or under development. Furthermore, a network of 83 countries provides global monitoring for adverse drug reactions and as well as for potential safety problems. Regarding pricing, 30 years ago, virtually no publicly available price information was available, and few countries actively encouraged generic substitutions. In the 21st century, at least 33 countries provide such information.

Prevention and safety have long been an integral component of public health. Section 2 of volume 1 presents several essays on the topic. By focusing on ways to control risks, public health works toward making the environment a safer and healthier place in which to live. Silverstein presents an historical account of occupational health and safety in the United States. The Occupational Safety and Health Act of 1970 (OSHAct) declared that every worker in America is entitled to a safe and healthful workplace, and that employers are responsible for work being free from recognized hazards. Now, forty years later, many of the promises of the OSHAct have yet to be met. Silverstein reviews the history of occupational health and safety in the United States and exposes the barriers to OSHA's success (predominantly linked to the statutory design of regulation, inspection, and education) and the challenges that remain in preventing injury and illness at the workplace.

Hupert, Wattson, and Xiong present a sophisticated analysis of the complexity of planning for and responding to public health emergencies. Using the example of a large-scale aerosol anthrax exposure over an urban locale, they explore key determinants of health outcomes and health system surge capacity using several modeling techniques (state transition, queuing network). They suggest that such models can provide valuable insights for forecasting the logistical and staffing needs of large-scale prophylaxis campaigns for a range of intentional and natural disease outbreaks, such as the 2009 influenza A (H1N1) pandemic. While all model-based studies have their potential limitations, they may serve many functions in emergency preparedness and planning that cannot be provided through other means.

Food safety has periodically seeped into the consciousness of the lay public, almost always after a public tragedy involving tainted food. In 2006, there were 1,270 reported U.S. foodborne disease outbreaks, resulting in 27,634 illnesses and 11 deaths. Since then, many other well-publicized incidents have involved the safety of food products, including food recalls (berries from California, grapes from Chile, and so forth), contaminated beef or poultry, and recalls due to contamination (peanuts, almonds, and pistachio salmonella outbreaks occurred in 2009). Furthermore, public concerns over the use of food additives as well as use of pesticides have spurred interest in organically grown food products. Perhaps in response to the recent food outbreaks in the United States, the Food and Drug Administration (FDA) recently named a highly qualified food safety expert to be deputy commissioner for foods at the FDA. The newly created position is the first to oversee all the agency's numerous food and nutrition programs, and setting safety standards for produce is a top priority.

The article on foodborne illnesses by Tauxe, Doyle, Kuchenmuller, Schlundt, and Stein focuses on this important topic. Foodborne diseases are caused by a broad variety of pathogens and toxins. In their comprehensive and scholarly

article, the authors review the epidemiological, microbiological, and public health aspects of foodborne diseases resulting from the ingestion of contaminated foods and food products, and discuss the evolving public health approaches to the global challenges of foodborne infections. The global challenge of safeguarding the world's food supply is complicated by growing international trade, migration, and travel. Through the globalization of food marketing and distribution, contaminated food products can and do affect the health of people in numerous countries at the same time.

Pharmaceutical safety in the United States is under the purview of the FDA. Haas presents an overview of how the benefits and risks of pharmaceuticals are managed and discusses the implications for global drug safety. As more pharmaceutical products are manufactured in countries around the world (including Canada, China, and India), concerns about safety of the products are quite valid. The FDA does not have the money or the manpower to inspect each plant overseas; yet, the final product is distributed and marketed in the United States. Haas provides an excellent historical overview of key issues in drug safety, which led to regulation of the marketing of approved drug products, most notably the Food and Drug Amendment Act of 2007. The act mandated that product label changes for safety were to be imposed and executed promptly. To ensure an acceptable benefit-risk balance, the FDA was empowered to require additional studies or trials, and it could stipulate specific conditions limiting the market availability of a product to ensure its safe use. The FDA was instructed to promptly communicate evolving product safety concerns even if the available information was limited. In addition, the act mandated that virtually all clinical trials, regardless of sponsorship, be registered and that efficacy and safety results be publicly posted in a timely manner. The act created a major new safety information system (the Sentinel System) that would complement spontaneous adverse event reporting. Despite efforts to tighten the mechanisms to ensure drug safety, the system is not fail-safe. The goal for drug safety is to have a flexible and responsive system able to recognize potential risks early, collect information efficiently, and take action that is appropriate in the context of both benefits and risks.

Focusing on the needs of vulnerable populations is an important component of public health. Section 3 of volume 1 highlights health care issues of special population groups. Karpur, Bjelland, and Bruyère from the Employment and Disability Institute of Cornell University highlight the role of public health in improving the health, well-being, and overall quality of life for people with disabilities through the consideration of epidemiological trends in disability prevalence, issues related to health disparities, the legal and regulatory environment affecting access to preventive and curative health services, methods of measuring and tracking the population of people with disabilities, and specific priorities in public health. The Institute of Medicine (IOM) refers to disability as ''the nation's largest public health problem,'' one that affects not only the health of people with disabilities, but also their immediate families and the population at large. Key issues for people with disabilities requiring attention in the U.S. public health system need to be addressed at the global level. It is estimated that there are approximately 650 million people with disabilities in the world with about 80 percent living in developing countries. The authors discuss various models and strategies to improve

health and well-being for people with disabilities, focusing on health disparities for people with disabilities; addressing the unique considerations for youth, women, and the aging with disabilities; and working toward an equitable access to health care, health care insurance coverage, health promotion, and prevention of secondary conditions—universal concerns that public health systems in all countries should be taking into account in the development of their national strategy.

Immigrant health care traditionally has largely been ignored by health policy makers. Yet, in 2009, an estimated 16 million children lived in immigrant families in the United States, representing one of the fastest-growing segments of the population. Clearly, policies and programs are needed to support immigrant parents and children, but the reality is haphazard at best. Mohanty, Woolhandler, Himmelstein, Pati, Carrasquillo, and Bor present compelling data based on the 1998 Medical Expenditure Panel Survey (MEPS) and found that immigrants have less access to health care and less health care use than do U.S.-born individuals. They also found that per capita health care expenditures for immigrants were far lower than expenditures for the U.S.-born. The study convincingly showed that the widely held assumption that immigrants consume large amounts of scarce health care resources is not supported by the data. The authors conclude that the low expenditures of publicly insured immigrants also suggest that policy efforts to terminate immigrants' coverage would result in little savings.

The provision of health care (or lack thereof) to those incarcerated has a long, sordid history. Finkel presents statistical evidence to illustrate the spectrum of health problems in correctional facilities. Inmate health and medical conditions range the gamut from minor (colds or viruses) to the significant (HIV/AIDS, TB). In addition to the communicable diseases, the prevalence of mental health and psychiatric diseases and substance abuse is higher among the prison population than the general population. The public health consequences of not paying attention to the health of prisoners can be quite significant; infectious diseases transmitted or exacerbated in prisons have the potential to become full-blown public health problems when prisoners return to their communities. The scope of this chapter provides an overview of the state of health among prisoners, assesses the provision of health care to those incarcerated, examines the policies regulating care of prisoners, including the challenges governments face in their ability to provide health and medical care to inmates, and discusses the pros and cons of having the private sector (privatization) involved in prison health care delivery.

Lesbian, gay, bisexual, and transgender (LGBT) health care also has received marginal interest and attention among policy makers and in the medical school curriculum. Medical education in the United States, both during medical school and in residency, is often unlikely to include adequate cultural competency related to the care of sexual orientation and gender identity minorities. A survey conducted to assess curricula in U.S. medical schools found that less than 3.5 hours were dedicated to teaching about health issues related to homosexuality. Part of the problem is the paucity of data on population demographics and health status for this population. For public health departments and providers to plan appropriate services for this vulnerable population, it is essential to have reliable data. Until recently, many of the research studies conducted in

the LGBT community were community-based studies using nonprobability sampling techniques. Radix and Mayer discuss the barriers to access to care as well as the health issues prevalent among the LGBT community. Of course, each group has its own set of health care needs, as lesbian health care is necessarily different from gay health care. The authors make the point that LGBT individuals have specific health needs that require targeted and culturally appropriate interventions.

The first ever surgeon general's report on oral health in the United States was published in 2000. The report highlighted a "silent epidemic" of dental and oral diseases, especially among the poor, the elderly, and children. Globally, too, oral disease burden and disability, especially in poor and marginalized populations, is a huge unmet issue. Oral health is much more than the pains of a toothache. Oral diseases such as dental caries, periodontal disease, tooth loss, oral mucosal lesions, and oropharyngeal cancers are major public health problems worldwide. Poor oral health has a profound effect on general health and quality of life. The burden of oral diseases and conditions is greatest among the economically disadvantaged, which include a disproportionately large number of racial and ethnic minorities and underserved populations. The major risk factors for oral disease are known and they are common with other chronic diseases: diet, smoking, alcohol, and risky behaviors. Canto and Cruz provide an epidemiologic overview of the state of oral health care as well as discuss preventive measures initiated to reduce dental caries, including exposure to fluoride (community water fluoridation, for example, has done much to reduce dental caries), use of dental sealants, practice of good oral hygiene, and reduction in sugar intake. The unmet need for dental care is a serious problem that needs to be acknowledged and addressed.

Taking care of the health care needs of the growing geriatric population is complex, challenging, and, to some extent, costly. Adelman, Finkelstein, Mehta, and Greene present an overview of the challenges of providing high-quality care to a rapidly aging population. They examine the medical, psychological, and social components of older age and explore the needs of this heterogeneous cohort. Issues such as dementia and Alzheimer's disease, elder abuse, ageist bias, the risks of polypharmacy, and long-term care issues are discussed.

Section 4 of volume 1 focuses on population-based prevention strategies. Adolescent substance abuse (alcohol, drugs, tobacco) has been well studied over the past decades; yet, the problem remains. Prevention and control programs have received considerable attention over the past decades as well. Botvin, Griffin, and Murphy, leaders in adolescent substance abuse prevention and cessation studies, raise a number of important issues related to adolescent substance abuse, including prevalence trends and types of prevention-based program modalities used by schools, families, and communities. The authors state that the most effective approaches target salient risk and protective factors, are guided by psychosocial theories regarding the etiology of substance use and abuse, and are implemented over many years. Many school-based prevention programs, for example, focus on skill-building in the area of drug resistance as well as life-skills training. While progress has been made in the field of substance abuse prevention, continued efforts must be made in the area of skill-building to prevent adolescents and children from taking drugs in the first place.

The issue of violence has been the subject of numerous reports by private and public organizations over the past decade. The public health consequences of all forms of violence are considerable as violence is associated not just with fatalities, but also with substantial morbidity and costs. It is estimated that in 2006 the health costs of violence (both fatal and nonfatal) in the United States exceeded $70 billion. Anderson and Sidel take a global approach to the discussion of violence and its sequela, and lay out a public health approach to violence prevention. They posit that the goals of public health—to prevent disease, and injury and premature death and to promote healthy living conditions for all—are identical to the goals of violence prevention. The disciplines and methods of public health—analyzing the causes of diseases, injuries, and premature deaths and of poverty and despair and determining methods to counter them—can strengthen efforts to prevent violence. And the ability of public health workers to gain trust both nationally and internationally can bring new skills and vigor to violence prevention.

Women, especially women in resource-poor nations, are an especially vulnerable group in terms of economics and in health care. In the industrial and developing world, gender-based violence (GBV) is endemic. Not only is it a major public health and human rights problem, but also for the victims it can, and most often does, have devastating personal, health, societal, and economic consequences. Meshkat and Landes eloquently delineate the types of GBV ranging from sexual, psychological, and physical, and depict the global burden of the problem. In addressing the issue, it is important to understand that public health initiatives often are bound tightly to existing legal frameworks, and this holds especially true in the case of GBV. It is excellent that GBV is now recognized as a major global public health and human rights issue, but efforts to stem its practice still stymie those involved in the prevention and management of GBV. Much work remains both in the industrial and developing world to ensure the safety and well-being of women of all ages against all forms of GBV.

Few issues in public health have fostered as much controversy as contraception and abortion. Passions run high on both sides of the debate; religion, politics, and policy regularly clash. Henshaw, who has spent his career conducting research in this area, presents a comprehensive statistical report on the issue of contraception and abortion, and explores some of the barriers that inhibit or even prevent women from controlling their fertility. Focusing on unintended pregnancy, contraceptive use, and abortion in the United States, he clearly and concisely presents the statistical evidence showing trends and highlighting the barriers that exist to prevent women of all ages from controlling fertility. Regarding abortion, since the legalization of this procedure in 1973, it is estimated that 35 percent of women in the United States will have had at least one abortion by age 45. Regarding birth rates, recent figures show that in 2007 more babies were born in the United States than in any other year in U.S. history.[11] This increase reflects a larger population of women of childbearing age. Births to teenagers (ages 15 to 17), after declining for many years, increased, reasons for which are poorly understood. Mississippi has the nation's highest teen pregnancy rate, which was 60 percent higher than the national average.

Although tremendous advances have been made in the eradication of once-deadly diseases, the development of vaccines probably is the most significant

reason for the decline in morbidity and mortality from such diseases. Perhaps the world's greatest achievement in this area is the eradication of smallpox. Rosen takes a global look at disease prevention through vaccination, presenting an historical overview and then focusing on the challenges that remain. There is a staggering disparity in vaccination efforts worldwide; millions of children are needlessly dying from preventable infections. Closing the gap will require multinational efforts and significant amounts of manpower and financial resources. The WHO and the Global Alliance for Vaccines and Immunization are deeply involved in coordinating immunization plans, especially in the resource-poor nations.

VOLUME 2

Volume 2 focuses on the determinants of health and disease. Section 1 addresses the treatment and prevention of chronic diseases. Since the mid-20th century, there has been a huge explosion in the number of individuals diagnosed with diabetes mellitus. Endocrinologist Baker's chapter focuses on the global epidemic of diabetes and discusses the public health, medical, and economic implications of dealing with this disease as well as the consequences to patients and to society. Diabetes is a growing and serious disease that affects rich and poor alike. According to the WHO, diabetes is likely to be one of the most substantial threats to human health in the 21st century. In the United States. alone, the direct medical costs of treating diabetics will be $336 billion. This does not take into account the growing proportion of overweight children and teenagers who are at high risk for developing diabetes and does not factor in immigration or the growing population of ethnic minorities who also suffer from diabetes at much higher rates than the U.S. white population. Without significant changes in public or private strategies, the burden of treating diabetics will place a significant strain on an already overburdened health care system. Ironically, and perhaps tragically, diabetes is among the most preventable of major illnesses. Clearly, as Baker discusses, much more needs to be done to stem the epidemic domestically as well as internationally. For a chance of success, prevention efforts must include a partnership between the individual and the health care provider.

Cardiovascular disease, too, is among the leading causes of morbidity and mortality globally. In 2004, according to the WHO, 17.1 million people died from cardiovascular disease, which represents 29 percent of global deaths. By 2010, cardiovascular disease is predicted to be the leading cause of death in developing countries. Kassahun and Borden explore the surge in cardiovascular disease and its risk factors worldwide, the characteristics and implications of this growth, as well as public health initiatives that can stem or even reverse this trend. They discuss the social, environmental, and cultural determinants of cardiovascular health, such as obesity, tobacco use, and access to health care that need to be addressed globally to reduce the incidence of cardiovascular disease.

The surging prevalence of obesity in the United States and around the world is growing faster than that of any other public health condition. This trend is alarming from a medical and an economic perspective. The ever-increasing prevalence of obesity has been accompanied by a host of inherently associated comorbidities. As a result, obesity is fast becoming the major cause of premature death

in the industrial as well as the developing world. Cardiologists Bornstein and Cooper examine the implications of the huge explosion of overweight and obesity in the world in the 21st century. Over the past two decades, the number of overweight and obese adults, adolescents, and children has increased dramatically. Of great concern are the children and adolescents who already have early obesity-related degenerative diseases, such as hypertension, dyslipidemia, metabolic syndrome, and type-2 diabetes mellitus, as well as manifestations of early preclinical atherosclerotic cardiovascular disease that previously has not been observed in this age-group. The economic costs of obesity are examined as are preventive means of addressing the epidemic. For example, health care spending for obese American adults soared 82 percent between 2001 and 2006.[12] Health care costs related solely to obesity could easily total $344 billion in the United States by 2018, or more than one in five dollars spent on health care, if the trends continue. The central message is that if nothing is done to stem the rise in obesity, the economic, medical, and personal consequences will be even more difficult to deal with.

The global burden of asthma is explored by Shirtcliffe and Beasley from the Medical Research Institute of New Zealand. The rising global burden of chronic, noncommunicable diseases over recent decades has been labeled "the neglected epidemic." Over recent decades, asthma has become one of the most common chronic diseases in the world and is now the most common chronic disease of childhood in many countries. The authors present a comprehensive epidemiologic overview of the disease and address the probably causes of the increase of asthma worldwide, including climate change and urbanization. The economic burden of asthma is considerable both in terms of direct medical costs, such as the cost of pharmaceuticals and hospital admissions, and indirect medical costs, such as time lost from work and premature death. The GBD to governments, health care systems, families, and patients is substantial. Indeed, the authors argue that the burden of asthma in many countries is of sufficient magnitude to warrant its recognition as a priority disorder in government health strategies.

Mild to moderate hypertension is generally an asymptomatic disease. It aptly has been called the "silent killer" because it usually produces no symptoms and increases gradually and slowly over the years. People with high blood pressure usually have no idea that they have this problem, and they do not go to the doctor specifically because of elevated blood pressure. The detection and treatment of hypertension is thus a major public health challenge. Cheung and Ong focus on the growing burden of hypertension worldwide, and especially in the United States. Although hypertension is seldom curable, the more practical aim is to control the blood pressure. The authors present an epidemiologic overview of the disease and discuss the known risk factors for hypertension. Medical management and improving compliance with treatment are discussed.

Arthritis, especially osteoarthritis (OA), is a prevalent condition among most of the older population. As the baby boomers age, the number of new cases of OA are likely to increase. Perhaps not surprisingly, there has been a concomitant increase in the number of joint replacements being performed. Lyman and Nguyen focus on OA of the knee and the explosion in the number of total joint replacements being performed. While total knee replacement, in particular, is an elective procedure to treat severe arthritis of the knee, the increase in this surgery is driven

both by the aging of the population as well as the obesity epidemic. Weight loss interventions may be the single most efficacious method of prevention of knee OA, but barring that, surgical intervention is increasingly being used to treat OA and enhance quality of life. The authors discuss the economic burden of OA, which is substantial (direct and indirect costs associated with OA).

Section 2 focuses on advances in cancer screening and the challenges that remain. For years, the dominant view about screening was that early detection and aggressive treatment would lead to increased longevity. Screening for cancer targets healthy, asymptomatic individuals. The purpose is to detect the disease at an early stage to initiate treatment, which hopefully will extend life. A key principle is that the potential benefits of screening should outweigh the harms of testing. The physical, psychological, and economic sequelae of follow-up testing should the screening test be positive needs to be compared with the number of lives saved as a result of screening.

Cancer is the nation's number two killer behind heart disease and accounts for nearly a quarter of annual deaths. The good news, however, is that cancer death rates and the number of new cancer cases in the United States continue to decline.[13] The conclusion drawn was that early detection and new therapies are major contributors to this effect. Almost at the same time as this report was released, a new study on the effectiveness of mammogram screening also was released. An expert panel from the U.S. Preventive Services Task Force recommended that mammography screening to detect breast cancer should be scaled back. This bombshell recommendation, in direct conflict with the recommendations from the American Cancer Society and other medical groups, caused considerable confusion, distrust, and even anger. Studies evaluating the effectiveness of the prostate-specific antigen (PSA) test to screen for prostate cancer also have yielded questionable results, making a clear recommendation for or against this test almost impossible. The American Cancer Society, a staunch defender of most cancer screening, has said that the benefits of detecting many cancers in particular prostate cancer has been overstated. The PSA prostate cancer screening test has not been shown to prevent prostate cancer deaths. The dilemma for breast and prostate screening is that it is not usually clear which tumors need aggressive treatment and which can be left alone.

Some studies focusing on routine early cancer screening found that the screening did not save lives, thus calling into question why screening was being advocated in the first place. In some cases, widespread screening increased the detection and treatment of small, slow-growing tumors that may well never have caused harm. In some cases, the tumor might regress or even disappear. While almost all of the cancer screening tests in wide use are minimally invasive, fairly inexpensive, and generally accepted by the public, none are 100 percent accurate; positive test results require further workup, which often are invasive and costly and usually lead to overtreatment. Screening does come with medical risks. In many cases, disease is not evident, such as in the case of false-positive test results.

Trevena presents a scholarly overview of the benefits and risks of screening and early detection of disease. She then examines the evidence for screening for colorectal cancer. The issue of whether an individual benefits from early detection of cancer is not as straightforward as it may seem. For some diseases, a preclinical

phase may be so short that the disease is not likely to be detected by screening. Or, even if detected, options for cure may not exist. Also, not every preclinical case will progress to clinical disease. Trevena examines the screening options for colorectal cancer, including fecal occult blood testing, flexible sigmoidoscopy, colonoscopy, and a new screening option, CT colonography. She presents the pros and cons of each modality. Some countries recommend that a fecal occult blood test be used while others advocate for colonoscopy. The accuracy of the test, including false-positive results, needs to be weighed against the potential benefit in reducing colorectal mortality.

Elkin and Blinder of the Memorial Sloan Kettering Cancer Center in New York City focus on breast cancer screening. Mindful of the current mammography screening controversy, the authors present a comprehensive overview of the advances made in reducing and preventing breast cancer. Because so many of the risk factors for breast cancer are not modifiable, much attention has been devoted to other means of breast cancer prevention such as understanding the role of hormones in breast cancer etiology. Much of the chapter focuses on the current controversy in mammogram screening. The questions of when screening should be initiated, on whom, and how frequently remain controversial. The authors present a scholarly assessment of the evidence, including a discussion on the realities of false-positive results. The chapter concludes with a discussion of advances in breast cancer treatment, including surgical treatment, radiation therapy, systemic therapy, hormonal therapy, and chemotherapy.

Lung cancer, the leading cause of cancer mortality worldwide, typically exhibits symptoms only after the disease has spread to other organs, unfortunately making it difficult to cure patients with such advanced disease. The overall prognosis of this cancer is poor when compared with other cancers, such as breast or colon, and is dependent on where the cancer is located, the size and type of tumor, and the overall health status of the patient. The two types of lung cancers, small-cell lung cancer and non-small-cell lung cancer, grow and spread in different ways and also have different treatment options. To date, screening for lung cancer is not advocated for these reasons. Yet, we have known for decades that tobacco smoking is the leading cause of this cancer, and evidence is quite clear that if individuals stopped smoking (or never started), the incidence of lung cancer would be greatly reduced. Lung cancer can be prevented. Mazumdar's chapter on lung cancer prevention, screening, and treatment reviews the epidemiology of this cancer as well as focuses on the "effectiveness," "efficiency," and "efficaciousness" of treatment regimens. A national initiative for comparative effectiveness research (CER) for clinical decision making is described. A discussion of CER provides a review of ongoing research and initiatives in this area, and highlights the gaps in information and research. Overall, much works is needed to find a cure for lung cancer and in being able to bring the best possible care to patients of all race, gender, and socioeconomic status.

Controversy over prostate cancer screening and treatment options continues to play out in the lay and professional media. Nguyen and Kattan's chapter reviews the current status of prostate cancer screening and assess its benefits and potential deleterious effects, to determine ways to improve its predictive accuracy and efficacy. To better understand the controversy surrounding prostate cancer

screening and perhaps offer a solution, they provide a review of current screening modalities, assess their accuracy and utility in contemporary medical practice, and suggest future directions for improvement of prostate cancer screening.

Over the past decades, the incidence of skin cancer has increased substantially. The chapter by Berwick, Erdei, Gonzales, Torres, and Flores focuses on the epidemiology and genetics of skin cancer and illustrates the growing public health burden of this particular form of cancer. Advances in screening and treatment are discussed and preventive measures are explored. The incidence of melanoma, a potentially deadly form of skin cancer, has soared over the past few decades perhaps because of an increased interest in screening for the disease. Indeed, the increase might be due to a growing tendency to identify and treat benign lesions as malignant cancers. It is quite difficult, and sometimes impossible, to tell a malignant lesion from a melanocytic nevus, a type of benign mole. The authors discuss how to protect oneself from skin cancer and provide informative information on sunscreens, tanning beds, and genetic susceptibility. Although sun exposure is the major risk factor for skin cancer, it is also necessary for synthesis of vitamin D, necessary for bone and muscle health and a possible protective factor for many diseases, including colon cancer. Given the worry about sun exposure and skin cancer, the question remains: How does one achieve favorable vitamin D levels yet also practice skin cancer prevention?

Cervical cancer, so easily and inexpensively prevented, remains a major killer among women in resource-poor nations. Without screening intervention, morbidity and mortality will continue to increase. Sankaranarayanan, Thara, Ngoma, Naud, and Keita have published groundbreaking research on the topic, and in this chapter they present a comprehensive overview of the issue, including evidence convincingly showing that screening for human papillomavirus (HPV) can yield a significant reduction in the numbers of advanced cervical cancers and deaths from this disease. They review the current status and future prospects for controlling cervical cancer in developing countries in this chapter. Low-tech screening methods (often used because most rural areas cannot realistically conduct Pap smear screening) and a single round of screening for HPV can and does result in a dramatic reduction in the incidence of advanced cervical cancer. There is a huge potential to reduce the cervical cancer burden by means of HPV vaccination. The authors acknowledge that a recommendation for HPV vaccination for adolescence women for a disease that occurs during adulthood is a major paradigm shift in cervical cancer control. Although HPV vaccination holds great promise, and has been licensed for use in more than 100 countries, there are several challenges (notably cost) for its widespread implementation through national immunization programs in high-risk developing countries. Cervical cancer reflects striking global health inequity, resulting in deaths of women in their most productive years in developing countries, with a devastating effect on the society at large. It remains as the largest single cause of years of life lost to cancer in the developing world.

Section 3 of volume 2 focuses on the treatment and prevention of infectious diseases. So much has been written about HIV/AIDS over the past two decades and so much progress has been made in extending life expectancy among those with the disease. Demars takes a broad view of the epidemic, tracing its history and focusing on the global burden of the disease. While recent trend data indicate

that the incidence in Africa has appreciably slowed, dealing with the disease's sociopsychological sequela and ensuring that progress made is not eroded remain challenges both in the industrial world and the resource-poor, hard-hit part of the world.

The WHO estimates that more than 500 million individuals worldwide are infected with the hepatitis B or C virus. Hepatitis viruses are found in every part of the world and often cause infections ranging in severity from acute infections that are asymptomatic to fulminate, chronic infections, which in some instances can lead to cirrhosis and hepatocellular carcinoma or even death. Aden presents a focused discussion on the most prevalent hepatitis viruses (A, B, and C) and explains how these diseases remain an important public health concern in both the developing and the industrial world. Whereas hepatitis B is a more serious type of infection than hepatitis A, hepatitis C infection can result in serious liver damage; hepatitis C is one of the leading causes for liver transplantation. While hepatitis A and hepatitis B are vaccine preventable, no vaccine is available for hepatitis C. Risk factors, population at risk, and treatment modalities are presented.

The ebb and flow of sexually transmitted diseases (for example, chlamydia, gonorrhea, and syphilis) has long been a focus of public health practitioners. All three diseases are preventable, treatable with medication, and, in the early stages, curable. Torrone and Peterman of the CDC present an overview of the topic and focus on syphilis specifically. The authors discuss the challenges of sexually transmitted disease (STD) control focusing on trends, efforts at prevention and control, and the challenges that remain. STD control and eradication is possible, but certainly not easy.

Section 4 of volume 2 addresses the treatment and prevention of mental health illness and disease. The burden of mental health disorders in the United States is substantial with approximately half of the population meeting the criteria for one or more such disorders in their lifetime and almost one-quarter meeting the criteria in any given year.[14] Treatment costs for mental disorders are substantial, rising from $35 billion (in 2006 dollars) to nearly $58 billion, making it the costliest medical condition between 1996 and 2006.[15] The most prevalent class of disorders is anxiety disorders (for example, phobias, panic disorders, and the like) followed by impulse-control disorders, mood disorders (for example, major depressive disorders, bipolar disorders, and the like), and substance abuse disorders (for example, alcohol abuse or dependence, drug abuse or dependence, and the like). The most prevalent type of disorder is major depressive disorders. Most individuals with a lifetime mental disorder had their first onset in childhood or adolescence. Little is known about the epidemiology of child mental disorders and controversy exists about how best to treat children. Contributors to this section focus on specific mental disorders, such as depression, suicide, and substance abuse.

Depressive disorders are prevalent conditions among the general population, and the medical, public health, and economic consequences of depression are considerable. Tedeschini, Cassano, and Fava present an overview of depressive disorders and focus on the recognition, management, and treatment of these diseases. The authors stress that depression is underdiagnosed and undertreated as only half of all Americans with depression receive treatment of any kind. Despite the

availability of numerous effective treatments, many depressive disorders are often misdiagnosed. Several factors contributing to the poor recognition of depression have been identified, ranging from the stigma of depression itself to the relative lack of systematic ascertainment of depressive symptoms by physicians.

Barber and Miller focus on the topic of suicide both within the Untied States as well as globally. In their informative and scholarly piece, they review the salient aspects of the epidemiology of suicide and the challenges posed by a purely clinical approach to its prevention. They outline a public health approach to suicide prevention, with an emphasis on reducing a suicidal person's access to lethal means of suicide. Their thesis clearly illustrates that although suicide is a global problem, a public health approach to prevention is still in its infancy. Public health strategies, such as changing cultural attitudes, increasing social support, improving access to high-quality treatment, and perhaps most important, reducing access to lethal means are measures that can and should be implemented.

Griffin examines the data on substance use and abuse across the life span from early adolescence to late adulthood. There is great diversity in patterns of alcohol, tobacco, and other drug use over the life course, with some individuals abstaining from use throughout their lives and others facing ongoing battles with substance abuse and dependence. The focus is primarily on substance use rather than abuse, because substance use is more prevalent than abuse in the general population and therefore has a greater public health impact. A goal of the chapter is to examine the extent to which substance use can be thought of as a developmental phenomenon not only among young people, but also throughout the life course. The implications of a life span developmental perspective to guide substance use prevention efforts are discussed. Griffin highlights a future challenge: the anticipated increase in substance use problems among the elderly and among the baby boomers, the eldest of whom will be reaching age 65 in a few short years. By taking into account how age-related developmental factors can affect substance use, we may be better able to address these and other new prevention challenges in the future.

VOLUME 3

Volume 3 shifts focus to health policy issues. In section 1, Finkel provides a historical overview of comparative health care systems illustrating why and how other industrial nations moved toward universal health care and why the United States did not. The organization, administration, financing, and delivery of health care in several countries are presented in an in-depth analysis. A critique of how health care is delivered and financed in other countries provides a stark contrast to how health care is provided and paid for in the United States.

Quality and patient safety, in addition to cost management, is an important issue in health care policy. Lazar, Dawson, Hyman, Collins, Regan, Kaplan, Green, Cook, and Graham from the New York-Presbyterian Hospital present an overview of quality assurance, quality metrics, and quality evaluation techniques. Performance improvement management methodologies designed to reduce medical errors and safeguard a safe workplace. In 1999, the IOM published a seminal report entitled *To Err Is Human,* which catalyzed an enormous shift in the understanding of medical errors. The IOM report defined an error as an event in which

there is a failure of a process to achieve the intended outcome, or where an incorrect process of care was selected initially. An adverse event was defined as an injury to a patient caused by medical management rather than the patient's medical condition. The IOM report concluded that medical errors were responsible for as many as 98,000 deaths in the United States annually. Estimated annual costs of these errors were in the range of $17 billion to $29 billion. The report further opined that injuries caused by errors are inherently preventable. Lazar and colleagues state that achieving better outcomes for patients, lowering overall costs, and improving the patient experience will require the continued investment of time and money to spur innovation and create reliable effectiveness, safety, and efficiency in clinical settings. Measurement and continuous performance improvement are the mainstays of a robust organizational quality assurance program.

Until the 1980s, most people with private insurance in the United States were covered by traditional indemnity plans. As remains the case, the vast majority got their coverage through employment-based plans provided as a tax-exempt benefit. These indemnity plans delegated shopping decisions about what care to buy and where to buy it to individual consumers and their physicians and then relied on consumer cost sharing to contain costs. Specifically, plans used deductibles and coinsurance to create financial accountability for purchases; the notion was that responsibility for resulting out-of-pocket payments would create incentives for cost-conscious shopping. By design, health plans were relegated to a passive role of paying the bills, while providers were reimbursed fee-for-service on a cost basis. Such open-ended insurance schemes laid the foundation for rising costs, which the United States is now trying to reign in. Managed care was an attempt to contain costs, but has not succeeded in doing so. White discusses an alternative, consumer-oriented strategy (Consumer Directed Health Plans, CDHPs) to address the concerns and shortfalls of managed care. The basic nation of CDHPs is that by placing consumers at risk for paying for substantial amounts of care with their own money, this simultaneously will restore control over shopping decisions and increase consumers' motivation for cost-conscious shopping, while introducing savings options will mediate the accompanying increase in exposure to financial risk. White provides an in-depth discussion of CDHPs and their potential effectiveness in managing health care cost increases.

The use of health information technology (health IT) has become an exciting and important field in medicine. Ancker, Kern, Patel, Abramson, and Kaushal present a scholarly overview of the present and future uses of health IT. Health IT has been promoted widely as a potential solution to managing the massive amounts of data and information as well as serving as a cost management tool. The authors discuss the various types of health IT systems and explore the barriers to development and implementation of these technologies. Health IT offers particularly exciting possibilities for improving the quality and efficiency of health care delivery by making essential individual-level medical data more readily accessible at the point of care; improving communication among clinicians, patients, and public health agencies; and providing evidence-based clinical decision support to help clinicians practice according to optimal care guidelines.

Section 2 of volume 3 focuses on the difficult issue of health care disparities. Health status and health outcomes vary markedly among racial and ethnic groups.

According to an IOM report *Unequal Treatment: Confronting Racial and Ethnic Disparities* race and ethnicity remain a significant factor in determining whether an individual receives high-quality care and in determining health outcomes. Race has been shown to be a determinant of the characteristics and qualifications of physicians who patients see, the types of hospital to which a patient is admitted, and the types of procedures they will undergo. The explanations are complex.

Boutin-Foster focuses on diversity and the public health implications of a growing racially and ethnically diverse America. She examines the role academic medical centers can and should play in providing care to this multicultural population. An argument is made for the need to bring the issues of cultural diversity to the forefront of medical education. While progress has been made in increasing the proportion of racial and ethnic minorities in the health care field, the racial and ethnic composition of the health care workforce does not match that of the general population. Would systematic biases in treatment be reduced if the composition of the workforce resembles more closely that of the patient population? While no studies have been done to empirically answer this question, given the extent of disparities and unequal treatment (which have been researched), one could assume that it certainly would not hurt.

Section 3 of volume 3 discusses ethics and human rights issues. Atkinson explores why human rights is crucial to the work of public health, and argues that human rights is a necessary framework for public health. Her chapter explores why a human rights framework is crucial to the work of public health. The human rights framework—in concert with traditional medical ethics—articulates certain values and standards that specify how we should conduct ourselves. She presents an argument for an ethical and legal framework for moving forward the global public health agenda. She believes that the human rights framework offers us a reason to believe in the possibility of change.

Bioethicist de Melo-Martin addresses the ethically charged topic of genetic testing and public health. She presents some of the most significant ethical concerns that arise in relation to the use of genetic tests, discusses matters related to the analytic and clinical validity and utility of genetic tests, and explains how these aspects result in ethical quandaries. She then focuses on the concerns that the use of genetic tests, if such tests prove beneficial for the populations' health, might contribute to furthering existing health inequities. Finally, she discusses ethical issues related to obtaining, or omitting, informed consent and to protecting privacy and confidentiality. Ensuring that people are not unjustly discriminated against because of their genetic or health status requires careful attention to issues of privacy and confidentiality; yet, concerns about privacy need to be balanced against the legitimate public health needs. Focusing on these ethical concerns when making public-policy decisions about implementation of genetic testing and screening is necessary if we want to use these medical technologies in ways that will advance the public's health.

As chronic diseases, including cancer, surpass infectious diseases as the primary causes of death, and as individuals are living longer with their diseases, providing timely access to consistently high-quality end-of-life care has become an important international issue. How we manage the dying patient has both medical and ethical concerns. Tickoo and Glare present a comprehensive overview of the

palliative care movement both in the United States as well as in Australia, England, and India. At some point, all humans have to confront the inevitability of end of life. How one prepares for the eventuality of death is a personal and individual matter. What is necessary and important, however, is that end-of-life choices be made clear and available. Providing for end-of-life care is emotionally difficult, thus making it even more imperative that all patients have the option of timely access to palliative care services that are both appropriate and cost-effective.

Section 4 of volume 3 focuses on public health practice and education. Trushin and Bang present an interesting chapter on the role of epidemiology and biostatistics in health news reporting. In their thoughtful piece focusing on the role of uncertainty in science, they provide a comprehensive overview of the mechanisms of research and statistical analysis. The scientific method is based largely on common sense, and statistical thinking involves concepts that are accessible to all: an acceptance of chance and uncertainty, an appreciation of context, an ability to detect logical and factual flaws in information and ideas, and the realization that science is a fluid process whereby new empirical evidence is accumulated every day. The true spirit of science requires a healthy skepticism, which means suspended judgment and the use of reason to evaluate the validity of research results. Science thrives on these qualities, because they lead to a search for knowledge and ensure that the scientific method remains self-correcting.

Evidence-based medicine has been incorporated into the medical school curriculum, but it also has a role in public health practice. The design and use of public health actions that are effective in promoting health and preventing disease underlie the growing field of evidence-based public health (EBPH), which emerged in the 1990s to improve the *practice* of public health. Maylahn, Brownson, and Fielding describe the concepts and principles underlying EBPH, the analytic tools to enhance the adoption of evidence-based decision making, the dissemination and implementation in public health practice, and challenges and opportunities for more widespread use of EBPH, especially through state and local health departments. Unlike solving a math problem, significant decisions in public health must balance science and art, because rational, evidence-based decision making often involves choosing one alternative from among a set of rational choices. By applying the concepts of EBPH outlined in their chapter, the authors concluded that decision making and, ultimately, public health practice can be improved.

The American Association of Medical Colleges (AAMC) is the umbrella organization for U.S. medical schools. The AAMC's position on medical curriculum has far-reaching impact. Maeshiro of the AAMC presents an informative historical overview of the tensions and barriers to integrating the disciplines of clinical medicine and public health. The challenge of incorporating public health content into the standard medical curriculum is not new. Not surprisingly, the roots of this struggle are entwined with the historical events and trends that led to the separation, or "schism" as some have described, between the practice of medicine and the practice of public health in the United States. She relates how over time the disciplines have gradually moved toward an integrated whole both at the medical school curriculum level and at the postgraduate medical training level. The rise of

a specialty in preventive medicine, the development of a residency in preventive medicine, and the subsequent creation of board certification in this area are interrelated. The framework in which medical education exists (for example, accreditation criteria for both medical school and residency training, national examination content) acknowledge the need for physicians to have a population perspective.

Section 5 of volume 3 addresses some of the challenges public health faces as we move into the second decade of the 21st century. Few areas of biomedical science have aroused as much controversy as embryonic stem cell research. With advances in medical research and technology, stem cell research has proliferated around the world. Cauley addresses the stem cell debate, focusing on the medical, ethical, legal, and political aspects of the topic. He provides a scholarly overview of the short history of stem cell research and raises important questions that need to be addressed today and into the future.

Advances in computer science have opened a new area of research for global disease monitoring. McEntee, Castronovo, Jagai, and Naumova from Tufts University provide an overview of a number of advanced computational and analytical techniques that open new opportunities to examine the role of forecasting disease transmission and manifestation. They review applications of various remote sensing (RS) techniques and present the relatively nascent epidemiological applications of this technology. Public health applications of RS data are no longer new; spatial epidemiology is equally important as the strictly environmental applications for which RS was originally intended. This is not surprising because environmental studies and epidemiology are inextricably linked. Each provides information on human health conditions and the corresponding management of environmental resources. Climate and land-use change and variability can be measured remotely and corresponding effects of alterations in natural and built environments can be predicted. Their scholarly and thoughtful presentation of this new field illustrates the tremendous opportunities that can be tapped.

Smith's concluding remarks on thinking creatively about public health in the 21st century is an excellent historical wrap-up of key events in public health over the centuries. He provocatively asks what the public health field needs to do to meet the challenges. How should public health be shaped for the 21st century, both for its own sake as a critical field for the world's well-being and for the sake of the local and global public it serves? The answers to these questions, of course, are multilevel and multifaceted. He advocates that 21st-century public health should begin to look more rigorously at the multiple factors in a society that predict health outcomes. These factors include economics, housing, nutrition, sports and recreation, education, spirituality, family structure, gender relations, childcare, transportation, and whatever other factors make up an integrated human life. Those concerned with the improvement of health on a local, national, and global scale need to work collectively rather than in isolation. Health, after all, is a product of the multiple facets of society, and, as such, requires a multifaceted approach to health promotion, the prevention and treatment of disease, and, most important, the improvement of the quality of life for all people.

This is an exciting time for public health. As public health practitioners continue to work toward improvement in the health and well-being of populations around the world and focus on disease eradication and the prevention and control

of diseases, injury, and disability, this increasingly is being achieved in a global context for the potential benefit of all. It is the aim of this multivolume reference text to identify and analyze the diversity of the work being conducted in the contemporary public health landscape.

The tremendous effort that went into creating this multivolume text could not have been done without the generosity of the contributing authors. My appreciation for their time, their enthusiasm, and their scholarship, and especially their friendship, cannot be underestimated.

Madelon L. Finkel, PhD
New York City

REFERENCES

1. Winslow C-EA. The untilled fields of public health. *Science*. 1920; 51:23.
2. D. Cutler D, Miller G. The Role of public health improvements in health advances: the 20th century United States. http://www.nber.org/papers/w10511. National Bureau of Economic Research working paper w10511. Published May 2004.
3. Bunker JP, Frazier HS, Mosteller F. Improving health: measuring effects of medical care. *Milbank Quarterly*. 1994; 72:225–258.
4. Ten great public health achievements—United States, 1900–1999. *MMWR*. 1999; 48(12): 241–243. http://www.cdc.gov/mmwr/.
5. World Health Organization. *The World Health Report 2003-Shaping the Future*. Geneva, Switzerland: World Health Organization; 2003.
6. Health: Confronting the Challenges. http://www.usaid.gov/our_work/global_health/home/confrontingfactsheet.html.
7. G8. G8 communiqué. http://www.g8.utoronto.ca/summit/2000okinawa/finalcom.htm. Okinawa 2000.
8. Group of Seventy-Seven. Final communiqué adopted by the thirty-fourth meeting of chairmen/coordinators of the chapters of the Group of 77. http://www.g77.org/news/pr062703.htm. June 26–27, 2003; Geneva, Switzerland.
9. Alwan A, MacLean D, Mandil A. *Assessment of National Capacity for Non-Communicable Disease Prevention and Control*. Geneva, Switzerland: World Health Organization; 2001.
10. United Nations Development Programme. Millennium Goals. http://www.undp.org/mdg.
11. National Vital Statistics Report. Births, Marriages, Divorces, and Deaths. Vol. 57, No. 12. March 18, 2009. www.cdc.gov/nchs/data.
12. Trends in health care expenditures by body mass index (BMI) category for adults in the US civilian noninstitutionalized population. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st247/stat247.pdf.
13. Annual report to the nation on the status of cancer: 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. In BK Edwards, eds. *Cancer*. Lawrenceville, GA: Surveillance Research Program, National Cancer Institute.
14. Kessler RC, Wang PS. The descriptive epidemiology of commonly occurring mental disorders in the United States. *Annu Rev Public Health*. 2008; 29:115–129.
15. The five most costly conditions, 1996 and 2006: estimates for the US civilian noninstitutionalize population. http://www.meps.ahrq.gov/mepsweb/data_files/publications/st248/stat248.pdf.

This page intentionally left blank

SECTION 1

HEALTH CARE POLICY

This page intentionally left blank

CHAPTER 1

Comparative Health Care Systems: How Does the United States Stack Up?

Madelon L. Finkel, PhD

The U.S. Congressional Budget Office (CBO) projects that without significant changes in federal policy, total spending on health care would rise to 25 percent of the gross domestic product (GDP) by 2025, and rise to 50 percent by 2082. Furthermore, the two major public programs, Medicare and Medicaid, are estimated to account for 4 percent of the GDP in 2025 and almost 20 percent by 2082.[1] The CBO has gone on record as stating that the U.S. federal budget is on an unsustainable course primarily because of the rising cost of health care. Clearly, health care represents a large and growing component of the American GDP, and some sort of "meaningful reform" clearly is needed. Of course, the devil is in the details; what constitutes "meaningful reform" is being hotly debated. In addition, and adding to the current debate about health care reform, more than 45 million Americans do not have health care insurance, primarily because they cannot afford it. Half of all bankruptcies are a result of unpaid medical bills, and this situation exists both for those with and for those without health insurance coverage. Other nations do not spend nearly as much on health care, and every major country provides universal coverage for its citizens. What are they doing "right"? And, what is the U.S. doing "wrong"? This chapter seeks to provide an overview of comparative health care systems and then focus on the U.S. system. Historical events provide insight into how each system evolved and how the U.S. system, in particular, became so dysfunctional.

WHAT CAN WE LEARN FROM OTHER COUNTRIES?

Health care systems vary according to the extent of government involvement in providing care, ranging from nationalized health care systems (such as in the United Kingdom and Sweden) to decentralized private or nonprofit institutions (such as in Germany and France). Each country had to make a decision as to how the health care system would be organized, administered, and financed. A single-payer type of health insurance, for example, characterizes the Canadian system. The United

Kingdom has the most socialized system in the world, with the government controlling all aspects of health care delivery and financing. Multipayer systems are used in France and Germany in which health care is primarily financed by publicly controlled insurers.

In terms of financing a health care system, there are many different approaches. Publicly funded health care financing relies on a publicly managed fund. In some countries, the fund is controlled directly by the government or by an agency of the government for the benefit of the entire population. In contrast, private health insurance provides coverage only to those under contractual obligation between the insured or his sponsor and an insurance company, which basically seeks to make a profit by managing the flow of funds between funders and providers of health care. In terms of administration, health care may be administered and provided by the government (United Kingdom), or may be publicly funded, but health providers are private entities (the Canadian and French system). In other compulsory insurance models, health care is financed through funds (such as in Germany), which themselves are funded from a number of places, including employees' salary deductions, employers' contributions, and so forth.

Regardless of the type of health care system in place, health care spending around the world is rising at a faster rate than overall economic growth. Almost all countries have seen an increase in health care spending as a percentage of their GDP. Health spending is also rising faster than incomes in most developed nations, which raises questions about how each will pay for future health care needs. Before getting into actual dollar amounts, it is instructive to understand how individual health care systems evolved. Post–World War II is seen to be the pivotal point in history at which decisions had to be made.

The following provides a cursory overview of selected countries' health systems. Although it is not the intent of this chapter to provide an in-depth analysis of comparative health care systems, it is useful and important to highlight historical events that helped shape each individual nation's health care systems. Doing so both provides a point of comparison and illustrates what the United States might learn from other countries. In Europe, as in Canada, the guiding philosophy was, and still is, based on the principle of social solidarity, which implies that health care should be financed by individuals on the basis of their ability to pay, but it should be available to all who need it on roughly equal terms. Each of the countries discussed below mandates that individuals—regardless of economic status, employment status, age, sex, or any other demographic factor for that matter—must be covered by the state insurance scheme. Health care coverage is portable and comprehensive in scope of benefits. Nobody is denied coverage based on an inability to pay. All provide freedom of choice of providers.

GREAT BRITAIN

In Great Britain, during the early 1940s, Winston Churchill and his government made the decision to move millions of people from the cities to the countryside in the days before declaring war on Germany. Moving a large segment of the population to the countryside, however, necessitated that local social services be built up. Hospitals and physicians had to be available and a means to pay for their services

had to be put in place. The government established a national Emergency Medical Service to supplement the local services, and hundreds of hospitals were built, essential medical services (laboratory, radiology, ambulances, and so forth) were upgraded, and the Ministry of Health oversaw the operation of these services. As the war progressed, the government was compelled to provide free hospital treatment for civilians as well as for those in the military. During the German Blitz, most of the private hospitals and clinics were destroyed. Few individuals could afford to pay for private health care. As a result, the government ended up paying for medical care and doctors received a government salary for the portion of their time spent caring for patients.

While this emergency wartime health care system was intended to be temporary, after the war, both patients and providers alike did not want to see the new system discontinued. Sir William Beveridge was asked to prepare a postwar government plan for the provision of health care. The Beveridge Report on British Social Security called for comprehensive health care based on a tax-based national health service.[2] The report proposed widespread reform to the social welfare system, including the creation of the National Health Service (NHS). The National Health Service Act 1946 came into effect on July 5, 1948, and created the NHS in England and Wales; a separate NHS was created for Scotland at the same time. The NHS is accountable directly to the central government's Department of Health and Social Security, as well as to Parliament. The main principle was that a free, comprehensive health service should be available to all British citizens and legal residents, regardless of income. The system was very much a pragmatic outgrowth of postwar recovery efforts in war-damaged Britain.

A major characteristic of the NHS universal health care system is that it is universal, comprehensive, and paid for by tax-based financing.[3] Funding for the NHS is met from tax contributions paid by all persons over the age of 18 and employers in the United Kingdom. The system employs the physicians and nurses and owns most of the hospitals and clinics. The NHS also pays directly for the health care expenses. More than 1 million individuals are employed in some capacity by the NHS, and more than 2,500 hospitals and other health care organizations are included in the system. Essentially, the NHS is a single-payer system, in which the government is the nation's only health care purchaser. Individuals have the freedom to choose a provider, a general practitioner (GP), practicing in their geographic area.

Since its inception, the British NHS has followed a structured pattern, but that is not to say that reform efforts have not been initiated. As the costs of the system soared, as waiting time for hospital admissions increased, and as hospitals and health centers were showing signs of age and distress, the Thatcher government's response was privatization. In 1989, the NHS was reorganized to increase the private sector's role. The Conservative Party advocated the transfer of ownership to the private sector and tried to introduce more competitive market forces.[4] These initiatives were made on the basis of limited evidence as to whether the private sector was more efficient than the public sector.

Surveys showed little public support for the replacing the government-managed NHS with a mixed private-public health insurance program.[5] When the Blair Labour government gained control, the NHS system was in need of reform. The system was inefficient, waiting lists for nonurgent procedures were quite long, costs

were rising, and facilities needed upgrading. The Blair government allocated more money to the system, established the National Institute for Clinical Excellence as part of a modernization program, and created Strategic Health Authorities (SHAs), which are responsible for managing, monitoring, and improving local services in their regions. SHAs would be responsible for developing plans for improving health services in the local area, ensuring that the services are of high quality, and making sure that national priorities are integrated into local health service plans.

Furthermore, the Blair government established local Primary Care Trusts (PCTs), designed to organize and manage specific areas of medicine, such as mental health, primary care, ambulance services, acute care (hospitals), and social services. Collectively, PCTs are responsible for spending around 80 percent of the total NHS budget. PCTs have their own budgets and set their own priorities, within the overriding priorities and budgets set by the relevant SHA they belong to and the Department of Health (DH). Essentially, the PCT is responsible for planning primary care and community health services in its catchment area, and it contracts with hospitals and hospital consultants for specialty care as well as implementing quality improvement activities.

All physicians must belong to a trust, whether they practice in primary or specialty medicine. Every person in the United Kingdom has the right to choose to register with any GP of their choice practicing in their area. If the GP has contracted to provide NHS services, as virtually all do, then all consultations with the GP will be free of charge to the patient. While far from perfect, the British tax-based system in the 21st century does provide coverage to all. It is free at point of delivery, ensures choice of providers, and is comprehensive in scope. But, there is a chronic shortage of specialists in every field and there are the legendary long waiting lists, which stem from chronic underfunding and an undersupply of personnel and equipment. The NHS is well designed but chronically underprovisioned. Britain spends less of its GDP on health care spending than other nations yet faces massive deficits. The estimated cost of the NHS in England (the most populous part of the United Kingdom) in 2008 is £91.7 billion. The government has painful choices to make between cutting expenditures and raising taxes.[6] No other country has adopted the British system; each has implemented a unique universal system to reflect its specific needs and circumstances.

FRANCE

The 21st-century French health care system evolved from the 19th-century mutual aid societies. The 19th century had been marked by the rapid rise of the mutual benefit movement, which is still an important force in French political life. Manufacturers and unions formed collective insurance funds, which were financed through a self-imposed payroll tax; there were no set premiums. After World War II, the de Gaulle government, like the British government, had to focus on rebuilding the nation; money was limited and the pressing needs of society so great that the government was not willing or able to craft a new health care system.

The 21st century system of social security, including statutory health insurance, was established in 1945. With the end of the war, France established the *Sécurité Sociale*, an expanded program of health care and pension benefits. The postwar

government built on what already existed and expanded the payroll tax-funded system to cover all wage earners, their families, and retirees. (Self-employed and the uninsured were included decades later.) Although controlled by the French government, "social partners" (the trade unions and employer representatives) became active players in managing the French health care system.

The French National Health Insurance (NHI) scheme is an integral part of France's social security system, which ensures that all residents automatically be enrolled with an insurance fund based on their occupational status; health insurance is compulsory. The *Sécurité Sociale* is characterized as having universal coverage funded by payroll tax–financed insurance primarily through more than 140 independent, nonprofit, local insurance funds. Health insurance funds are not permitted to compete by lowering health insurance premiums or attempting to micromanage health care. Employer payroll taxes finance half of the NHI expenditures; "general social contributions" levied by the government on all earnings, including investment income, covers the rest. Employers, employee contributions, and personal income taxes fund the system. Of note is the fact that the working population has 20 percent of its gross salary deducted to fund the social security system.

The system is a public-private mix of hospital and ambulatory care. It is characterized as a solo-based, fee-for-service private practice for ambulatory care and public hospital system for acute institutional care. Patients are free to choose their hospital and their physician and are reimbursed under the NHI.[7] Nine out of 10 Frenchmen purchase supplementary health insurance to cover other benefits not reimbursed under the NHI. For ambulatory care, all health insurance plans operate on the traditional indemnity model—that is, reimbursement for services rendered. In the inpatient hospital services sector, there are budgetary allocations as well as per diem reimbursements. Physicians are private practitioners and are paid directly by the patient on the basis of a national fee schedule. Patients are then reimbursed by their local health insurance fund. The overwhelming number of practitioners comply with the *Tarif de convention* (tariff references), which sets reimbursement prices. Tariff references are the fixed rates to be used by doctors and are set by the national convention. Medical practitioners and clinics and hospitals that are not *conventions* have to display their prices. Public hospitals are paid on the basis of annual global budgets that are negotiated every year. Proprietary hospitals, which any insured citizen can go to, are reimbursed on a negotiated per diem basis.

As is the case with other peer nations, France is coping with rising health care costs. The deficit of the *Sécurité Sociale* is a concern to the government, which periodically considers reducing the amount of reimbursement to providers. As a result, more individuals are turning to *l'assurance complémentaire* (complementary insurance). This health insurance covers all or part of the costs not reimbursed by the national health system.

The French system is an interesting model for the United States. It is highly regarded and was ranked as the number one health care system in the world by the World Health Organization (WHO) in 2000. It provides universal coverage yet allows physicians to be private small business owners. It incorporates the private insurance sector yet is government administered. Individuals are guaranteed access to specialists and hospitals without going through a gatekeeper. There is no rationing

of care to speak of. Although far from perfect, the system is highly regarded by the French, as consumer and provider satisfaction is consistently high.

GERMANY

Germany has the world's oldest universal health care system, with origins dating back to Otto von Bismarck's Health Insurance Act of 1883. Germany's system is almost elegant in its simplicity, as described by health economist Uwe Reinhardt.[8] The system is decentralized (regionalized) with private practice physicians providing ambulatory care, and independent, mostly nonprofit hospitals providing the majority of inpatient care. Approximately 91 percent of the population is covered by a Statutory Health Insurance plan, which is financed by a payroll tax; the individual's premium is not a per capita levy but is income based. The plan provides a standardized level of coverage through any one of approximately 1,100 public or private sickness funds. Insurance payments are based on a percentage of income, divided between employee and employer.[9] Standard insurance is funded by a combination of employee contributions, employer contributions, and government subsidies on a scale determined by income level. Higher income workers sometimes choose to pay a tax and opt out of the standard plan, in favor of private insurance.[10]

The system is characterized by employer-employee financing rather than funding obtained solely from general taxes. Mandated coverage and employer and employee contributions have been the mainstay of the system from the beginning. The premiums paid by individuals and employers are collected by the government, which essentially serves as a risk-pooling entity for the entire system. Physicians and hospitals are paid by the sickness funds based on negotiated reimbursement rates. Although the national government sets national standards, the regions are powerful forces in shaping the federal legislation of standards for their region. The sickness funds, for example, traditionally have been allowed to set their own premium rates.

The German system includes portability of benefits should one move from one region of the country to another. Furthermore, health insurance remains unchanged for all workers during unemployment (unemployment insurance pays the premiums for unemployed individuals as their contributions to the national health insurance are paid for by a federally administered statutory unemployment insurance fund). An employee's nonworking spouse is covered automatically by the employee's premium, and retirees and the elderly are covered as well (pension funds share with the elderly in financing their premiums, which are set below actuarial costs for the elderly). Premiums for children are covered by the government out of general revenues.

The guiding principles of the German health care system (solidarity, decentralization, and nonstate operations) have not changed much since its inception. There is freedom of choice of providers and sickness funds; there is professional autonomy; there are comprehensive benefits; there are no waiting lines so to speak. Yet, as is true with every other nation, health care costs have been rising and there is a shortage of nurses. The most important topics for current and future reforms are financing and reimbursement, health technology assessment, and the fragmentation of health care between sectors and payers and collectivism versus competition.[11]

CANADA

The Canadian single-payer system is a universal health care system that is provincially based. That is, each of the 10 provinces and 3 sparsely populated northern territories is responsible for administering and financing health care for its legal citizens. Just after the end of the war in 1946, the first Canadian province, Saskatchewan, introduced near-universal health coverage. Saskatchewan, a rural farming province, had long suffered a shortage of doctors. It also had a long history of government involvement in providing health care—for example, in the early 20th century, the province established a municipal doctor program in which a town would subsidize a doctor to practice there, which laid the foundation for the single-payer system to be in place in the 21st century. In 1950, the province of Alberta introduced a similar program and later other provinces followed suit.

The passage of the Hospital Insurance and Diagnostic Services Act of 1957 (the HIDS Act) provided further impetus for provinces to establish their own health care systems. The act outlined five conditions for a health care system: public administration, comprehensiveness, universality, portability, and accessibility, and provided up to 50 percent of the costs of a health program for any provincial government that agreed to adopt these conditions. By 1961, all 10 provinces had some sort of health care system in place under the HIDS Act. Thus, a significant proportion of health care already was controlled and paid for by the provincial government. The HIDS Act served to lay the foundation for the Canada Health Act of 1984, which specified the conditions and criteria with which the provincial and territorial health insurance programs had to conform to receive federal money. These criteria require universal coverage for all "insured persons" and for all "medically necessary" hospital and physician services without copayments.

Ottawa provides funding for specific programs, but the provinces individually organize and administer their own health care system. To receive money from the federal government, each province has only to comply with a few conditions: operate a nonprofit insurance plan, provide comprehensive benefits, ensure that the plan is universal in scope to all insured persons of the province, ensure portability should insured people move between provinces, and be accessible.[12] Under the Canadian plans, physicians are small businessmen largely working with a high degree of autonomy under a fee-for-service system. Bills are submitted directly to the single payer, which greatly reduces the administrative overhead in administering the program. Individuals have freedom of choice of providers.

No system is perfect, and the Canadian system is no exception. There are waiting lists, varying considerably by procedure and province. Costs have increased and there is evidence of cost shifting. For example, inpatients receive full coverage for pharmaceuticals. Once they are discharged, these costs need no longer be paid for from public funds. Some provinces will pay for these costs, but others do not. As such, two-thirds of Canadians have some sort of supplementary private health insurance.

SWITZERLAND

Although the above countries' approaches to health care delivery and financing encompasses variations of a public-private mix, Switzerland, a confederation of states,

has a system that might be more palatable for the United States. Health care is regulated by the Federal Health Insurance Act of 1994. In 1996, Switzerland mandated that individuals purchase health insurance covering a range of benefits that are set out in detail in the Federal Act. Since it had no experience with a public system, the private nonprofit insurance industry filled the void. In this case, purchase is made from nonprofit private insurers who compete in the market place. The basic benefit package is set by law and is deemed to be quite generous; there is freedom of choice of providers. The government provides subsidies for those who cannot afford coverage.[13]

The compulsory insurance can be supplemented by a private insurance policy that would provide coverage for some of the treatment categories not covered by the basic insurance or to improve the standard of room and service in case of hospitalization. Although the level of premium can vary from one company to another, they must be identical within the same company for all insured persons of the same age-group and region, regardless of sex or state of health. This does not apply to complementary insurance in which premiums are risk-based.

Although a small and fairly homogeneous nation, Switzerland provides universal coverage, mandates that everyone purchase insurance from nongovernmental insurers, and delivers care without long waiting lists. Of course, there are issues such as containing rising costs but this system has been viewed as a model for the United States that might be acceptable to both Republicans and Democrats.

IMPLICATIONS FOR THE UNITED STATES

The countries discussed above each had to decide how best to provide health care coverage to their citizens. Each made the decision to provide, indeed mandate, affordable health care to its legal citizens. In terms of funding the health care system, each took a different path: Great Britain relies on a tax-based system; France collects general revenues and also levies a mandatory payroll tax to fund its insurance system. Germany relies primarily on work-based social insurance contributions as well as federal taxes. Canada funds its system through a national and provincial general revenue scheme, and Switzerland mandates that everyone purchase health care coverage from nongovernmental insurers. None is superior to the others; each "works," although cost control efforts and efficiencies in administration and organization clearly are necessary in every system.

In terms of satisfaction with one's health care system, citizens in these countries generally tend to be pleased. They have expressed concerns about rising costs but, by and large, people are satisfied with their system, and none reported that they would want the U.S. system to replace their own. A survey addressing this issue found that while the British were the most satisfied with their health care system, there were problems with wait times. Canadians, also very satisfied with their health care system, reported difficulties in seeing a specialist due to shortages of physicians. Wait times also topped the list of public concern.[14]

Regarding ability to get care in a timely manner or ability to see a doctor also in a timely manner, every country limits or rations health care to some extent. The United States is no exception, even for those with health care coverage. What other countries do not have is a burdensome, bloated administrative structure. Whereas the administrative costs for private insurance in the United States is approximately

25 cents on every dollar, the public programs, Medicare and Medicaid, have an administrative cost of between 4 and 8 cents per dollar. France's administrative costs are 4 cents per dollar and Canada's are 6 cents per dollar, as a point of comparison. The U.S. system is fragmented, costly, confusing, and unfair to many. In Europe, for example, nobody is denied services or care. In the United States, however, claims are denied routinely.

In terms of quality, an international comparison study found that each country performs well in some areas and poorly in others.[15] Based on 21 quality indicators, the study found that no country performed consistently better or worse than the others; each country in the study had at least one area of care where it could learn from the others. The study found that the United States performed relatively well, but because the American system is so expensive, it probably is not getting good value for its medical care dollar. More spending is not buying better outcomes.

The United States has the dubious distinction of being the only industrial country that does not have a universal health care system. Whether a government-mandated system of universal health care should be implemented in the United States remains a hotly debated political topic, with Americans divided in their views of the U.S. health care system and what should be done to improve it. What is clear, however, is that the United States spends far more per capita on health care, expends the highest proportion of its GDP on health care compared with peer nations, and has the most rapid growth rates in health spending than other nations. Little evidence indicates that the United States gets better value for its higher health care spending.[16]

OECD DATA

Data collected by the Organisation for Economic Co-operation and Development (OECD), a 30-country organization known as a premium statistical agency that publishes highly comparable statistics on a wide number of subjects including health care, are useful to make cross-national comparisons. Recent publications show that the United States continues to spend much more on health care both as a proportion of the GDP and per capita than any other OECD country, yet U.S. life expectancy is lower than would be predicted based on U.S. per capita income. Table 1.1 shows comparative data depicting total health expenditures per capita. Health spending per capita is at least 24 percent higher than in the next highest spending countries and more than 90 percent higher than in many other countries that are considered global competitors to the United States.[17] Whereas Switzerland, Sweden, and Denmark had levels of per capita health spending somewhat comparable to the United States in 1980, these countries had much lower average annual growth rates in health spending than the United States since then. This lower growth rate implies that, unlike the United States, other nations managed to control better levels of health spending.[17] Total health expenditures as a share of the GDP in the United States continues to outpace that for other comparable nations (see table 1.2).

Despite spending far more on health care, the U.S. health care system provides less access to health care resources. That is, the United States has fewer physicians, nurses, hospital beds, doctor visits, and hospital days per capita than the median OECD country.[18] What then is a plausible explanation for this situation? Analysts

have concluded that U.S. prices for health resources are much higher than in other OECD countries.[19] It is well known that most newly graduated physicians have crushing tuition debt. Many medical schools in other countries do not charge tuition or, if they do, the amount is quite modest. Perhaps, then, it is not surprising that health care workers' salaries are higher in the United States than in other countries.

Other factors also account for the difference. An inpatient hospital stay in the United States is more service-intensive and more expensive than in other nations, even though the length of stay tends to be shorter in the United States. Furthermore, the fragmented, dysfunctional administrative structure in the United States does not lend itself to cost efficiencies. Anderson et al. therefore concluded that the crucial role of prices as drivers of cross-national differences in health spending is the most likely explanation for the differences.[19] Other countries rely on supply constraints to control their health care spending—that is, limiting the number of hospital beds that can be built; putting controls on the diffusion of medical technology; and limiting the number of physicians, which has led to shortages in some countries.

U.S. STATISTICS

The need for fundamental transformation of the U.S. health care system has become increasingly apparent. Research reveals a fragmented system fraught with

Table 1.1
Total Health Expenditures per Capita, United States and Selected Countries, 2003

Country	Total Health Expenditure per Capita (in U.S. dollars)
Australia	$2,886
Austria	$2,958
Belgium	$3,044
Canada	$2,998
Denmark	$2,743
Finland	$2,104
France	$3,048
Germany	$2,983
Iceland	$3,159
Ireland	$2,455
Italy	$2,314
Japan	$2,249
Luxembourg	$4,611
Netherlands	$2,909
Norway	$3,769
Sweden	$2,745
Switzerland	$3,847
United Kingdom	$2,317
UNITED STATES	$5,711

Source: Adapted from the Organisation for Economic Co-operation and Development. OECD Health Data, 2006. http//www.oecd.org/health/healthdata.

Table 1.2
Total Health Expenditures as Percent Share of GDP, United States and Selected Countries, 2003

Country	Total Health Expenditure as Percent Share of GDP
Australia	9.2
Austria	9.6
Belgium	10.1
Canada	9.9
Denmark	8.9
Finland	7.4
France	10.4
Germany	10.8
Iceland	10.5
Ireland	7.2
Italy	8.4
Japan	8.0
Luxembourg	7.7
Netherlands	9.1
Norway	10.1
Sweden	9.3
Switzerland	11.5
United Kingdom	7.8
UNITED STATES	15.2

Source: Adapted from the Organisation for Economic Co-operation and Development. OECD Health Data, 2006. http//www.oecd.org/health/healthdata.
Note: GDP = gross domestic product.

waste and inefficiency. The administrative overhead in America is huge by international standards. For example, a study of administrative costs in the U.S. and Canadian health care systems showed that Americans spent more than $1,000 per capita on administration compared with only $307 in purchasing power parity dollars spent in Canada.[20] Purchasing power parity is an economic theory that estimates the amount of adjustment needed on the exchange rate between countries in order for the exchange to be equivalent to each currency's purchasing power. Although the disparity is huge, it is important to cite that these figures are based on 1999 dollars. Most certainly the gap has widened over the years. Contributing to this excess administrative overhead is the underwriting and marketing of the complex private insurance system. Ironically, the government programs (Medicare, Medicaid, and the Veterans Administration) have a much lower administrative overhead, comparable to that seen in Canada and Europe.

The excess spending inherent in the U.S. system refers to the difference between what a country spends per person on health care and what the country's GDP per person should predict that that country would spend.[21] The United States spends nearly 40 percent more on health care per capita than its GDP per capita would predict. High spending, however, has not translated into better health. The

WHO in 2000 commissioned an analysis of the world's health care systems. Using five performance indicators to measure health systems in 191 member states, it declared that France provides the best overall health care. The United States health care system was ranked first in both responsiveness and expenditure, but 37th in overall performance and 72nd by overall level of health. In contrast, the United Kingdom, which spent just 6 percent of GDP on health services, ranked 18th.[22]

Despite spending more on health care overall and per capita, Americans do not live as long as citizens of several other industrial countries, and disparities in the U.S. system are pervasive, with widespread differences in access to care based on insurance status, income, race, and ethnicity. In all fairness, the WHO report has been heavily criticized and, perhaps as a result, the WHO no longer produces ranking because of the complexity of the task. But, the fact remains that there are real and legitimate questions regarding the access, efficiency, and quality purchased by the high sums spent in the American system compared with all others.

Health care accounts for a large slice of the U.S. economic pie. Each year, health care spending has grown at an average annual rate of almost 10 percent, often outpacing spending on other goods and services. Since 1970, health care costs have grown on average 2.4 percent faster than the GDP.[23] The high and growing cost of health care is not a new issue or a new problem. Over the past decades, the proportion of dollars devoted to health care has increased steadily, causing significant concern among individuals, businesses, and governments (federal, state, and local). Businesses, for example, find it difficult to remain economically competitive when an ever-increasing amount of money has to be allocated to pay for health insurance for their employees. Individuals are chafing because of increases in deductibles and coinsurance, and more recently, because of cost-shifting. Out-of-pocket costs have increased steadily increased. Individuals, even those with insurance, are spending more of their income each year on health care costs despite having coverage.

Health care is expected to grow faster than most other sectors of the economy. Whereas education, transportation, and agriculture tend to grow at rates close to the economy, health care has not and does not. The United States was projected to spend more than $2.5 trillion on health care in 2009 (17.6 percent of the GDP) or $7,400 per person each year. By 2018, spending on health care is estimated to be more than $4.3 trillion. Since 1970, health care spending has risen 3.4 percentage points faster than the GDP, a trend that is not economically sensible.[23] Although health spending is fairly evenly split between the private sector and the public sector, both sectors have experienced persistent and growing increases in the cost of providing health care, creating the impetus for increases in insurance premiums, higher deductibles, cost-shifting, and cost-sharing. Health insurance premium increases have consistently outpaced inflation and the growth in workers' earnings. Premiums are rising faster than overall inflation and workers' wages, and for the 160 million people who receive their health insurance through their employer, the amount paid out of pocket is also increasing.

On the public side, costs for Medicare and Medicaid have increased significantly since their inception in 1965 despite efforts at cost containment. As the baby boomer generation becomes eligible for Medicare, the future anticipated cost burden could sink the program. Moreover, the financial meltdown of 2009 has left tens of thousands of individuals without a job and thus without health care coverage. It

is more than likely that many will qualify for Medicaid, thus putting additional pressure on this federal-state program. Not surprisingly, one of the major challenges in enacting health care reform is how to finance it without adding to the already huge federal budget deficit.

CONCLUSION

Equity is about fairness. In health care terms, it is about ensuring equal access to health services for those with equal need, irrespective of personal characteristics (race/ethnicity, sex, geographic area). Every developed nation in the world except the United States has a health care system that guarantees affordable health care for its citizens. For those Americans who have health insurance, the cost of purchasing insurance coverage continues to increase rapidly. Ever-increasing premiums represent a serious problem for many. A survey by the Kaiser Family Foundation found that the average premium for a family policy in 2009 was more than $13,000. In 1999, the cost was $5,800. Although premium increases have decreased from double-digit increases seen in the late 1990s and early 2000s, the 5 percent premium increase in 2009 still outpaced the 3.1 percent growth in wages. Also, premium increases are still going up faster than overall inflation and workers' wages.[23] The Kaiser report also presented some very scary numbers: if one takes the average of premium increases over the last 10 years, an annual health insurance premium in 2019, just nine years away, could top $24,000. Clearly the system cannot continue on its present path.

What we are seeing is a not-so-gradual unraveling of the employment-based system that has been in place for decades. Rising health care costs is the most significant driver of America's long-term debt and deficits. The need for reform is quite clear. How the United States got itself into this mess is the subject of the next chapter.

REFERENCES

1. Congressional Budget Office. http://www.cbo.gov/publications/collections/health.cfm.
2. Abel-Smith B. The Beveridge Report: its origins and outcomes. *Blackwell Synergy: Int Social Security Rev*. 1992;45(1/2):5–16. http://www.blackwell-synergy.com/doi/abs/10.1111/j.1468-246X.1992.tb00900.x.
3. Light DW. Universal health care: lessons from the British experience. *Am J Public Health*. 2003;93:25–30.
4. Privatization: everybody's doing it differently. *The Economist*. 1985;297:71–73.
5. Blendon RJ, Donelan K. British public opinion on national health service reform. *Health Affairs*. 1989;8(4):52–62.
6. Klein R. Britain's national health service revisited. *N Engl J Med*. 2004;350(9):937–942.
7. Rodwin VG. The health care system under French national health insurance: lessons for health reform in the United States. *Am J Public Health*. 2003;83:31–37.
8. Reinhardt UE. Global budgeting in German health care: insight for Americans. *Domestic Affairs*. 1993;94:159–194.
9. Altenstetter C. Insights from health care in Germany. *Am J Public Health*. 2003;93:38–44.
10. Universal Health Care. http://en.wikipedia.org/wiki/Universal_health_care#Germany.

11. Reinhardt UE. Health reform without a public plan: the German model. *Economix*. http://economix.blogs.nytimes.com. Published April 17, 2009.
12. Deber RB. Health care reform: lessons from Canada. *Am J Public Health*. 2003; 93:20–24.
13. Reinhardt UE. The Swiss health system: regulated competition without managed care. *JAMA*. 2004;292:1227.
14. Blendon RJ, et al. Common concerns amid diverse systems: health care experiences in five countries. *Health Affairs*. 2003;22:106–121.
15. Hussey PS, et al. How does the quality of care compare in five countries? *Health Affairs*. 2004;23:89–98.
16. Anderson GF, Frogner BK. Health spending in OECD countries: obtaining value per dollar. *Health Affairs*. 2008;27(6):1718–1727.
17. Organization for Economic Cooperation and Development. OECD health data, 2006: health care spending in the United States and OECD countries. http://www.oecd.org/health/healthdata.
18. Anderson GF, Frogner BK, Reinhardt UE. Health spending in OECD Countries in 2004: an update. *Health Affairs*. 2007;26:1481–1489.
19. Anderson GF, Reinhardt UE, Hussey PS. It's the prices, stupid: why the United States is so different from other countries. *Health Affairs*. 2003;22:89–94.
20. Woolhandler S, Campbell T, Himmelstein MD. Costs of health care administration in the United States and Canada. *N Engl J Med*. 2003;349:768–775.
21. Reinhardt UE. Why does U.S. health care cost so much? Part II: indefensible administrative costs. *Economix*. www.exocnomix.blogs.nytimes.com/2008/11/21. Published November 21, 2008.
22. Blendon RJ, Kim M, Benson JM. The public versus the world health organization on health system performance. *Health Affairs*. 2001;20(3):10–20.
23. Kaiser Family Foundation. Trends in health care costs and spending. http://www.kff.org.

CHAPTER 2

Clinical Quality and Patient Safety

Eliot J. Lazar, MD, MBA, Anthony Dawson, RN, Dan Hyman, MD, MMM, Karen Scott Collins, MD, MPH, Brian K. Regan, PhD, Steven Kaplan, MD, Robert A. Green, MD, MPH, Joseph T. Cooke, MD, and Phillip L. Graham III, MD, MPH

INTRODUCTION

The last 40 years have seen extraordinary advances in health care, not the least of which is the development of the science of quality and patient safety. New methods of implementing and measuring improvement in the outcomes patients' experience, and the relative safety in which medical care is provided, have spawned a virtual industry devoted to these goals. Positions such as chief quality officer, patient safety officer, performance improvement specialist, or "black belt" did not exist as few as 10 to 15 years ago. Improvement methodologies such as Lean, Six Sigma, or Plan, Do, Check, and Act have evolved from a demand by health care organizations to rapidly adopt and implement new policies and procedures designed to minimize errors and improve results. Measurement approaches grow increasingly more sophisticated, although true gold standards remain elusive. New organizations seem to spring up daily, each with a slightly different perspective on just what quality and safety means. This intense focus and emphasis not only underscores society's interest in better medical care, but also has increasing financial ramifications.

Given recent efforts in this area, one would assume that improved quality and patient safety would be relatively easy to achieve, or at least simple to define. As the subsequent text will demonstrate, however, the field is anything but simple. This chapter focuses on the progress made in quality assurance and the challenges that remain. For purposes of the piece, we will use the Institute of Medicine (IOM) definition of quality: "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge."[1]

QUALITY INDICATORS

Avedis Donabedian is heralded by many as the father of quality assurance. His seminal article "Evaluating the Quality of Medical Care," published in the *Milbank Memorial Fund Quarterly*, in 1966, provided the foundation for modern approaches

to the definition and measurement of contemporary clinical quality and safety.[2] Donabedian suggested that quality indicators could be characterized as measuring structure, process, or outcomes.[3] More recently, volume has been added as a fourth indicator. Ironically, in the same issue of the *Milbank Quarterly*, Odin Anderson opined that health care research had minimally affected public policy and that public health decisions were unsupported by scientific evidence.[4]

The last five years have seen an extraordinary number of quality indicators put into use, such as report cards, physician profiles, pay-for-performance, and a host of other applications. Although concerns about their validity persist, they have become a mainstay of public policy debate. Numerous entities have entered the arena of health care evaluation and clearly the discussions around health care reform include quality and access as cornerstones of future programs. Although acute care hospitals and long-term care facilities have been the focus to date, attention is being directed toward the performance of physicians and other health care providers.

DEFINING THE METRICS

If one accepts the modification of Donabedian's original construct, which categorizes quality metrics as volume, structure, outcomes, or process (VSOP), a brief discussion of the advantages and limitations of each is in order.

- **Volume:** In 1916, E. A. Codman postulated that experience was a key driver of hospital based clinical outcomes. "A hospital . . . organized to obtain the best results could not possibly allot such cases to its less experienced surgeons."[5] The basic concept is that higher volumes lead to improved quality. Although numerous papers have demonstrated that volume and quality are related, this has not been a universal finding.[6] Moreover, opinions differ as to whether institutional volume or operator volume is the more important driver, or whether excessively high volumes can erode performance. Despite these concerns, volume metrics remain in use because the data are available in administrative or financial data sets, and theoretically are easily understandable by the lay public.
- **Structure:** Structural metrics are defined by the ability to evaluate them based on yes or no answers—that is, they are binary, leading to easy aggregation and analysis. Structural metrics may be applied to institutions, programs, or practitioners. Examples may include designation as a specialty center, such as level-one trauma or cardiac surgery; qualifications including board certification; or the availability of equipment or technology, such as magnetic resonance imaging (MRI) or computerized practitioner order entry systems. Structural metrics are appealing because they are relatively easy to collect and do not require complicated data aggregation systems. Yet, the utility of structural metrics is limited, given the relative paucity of evidence indicating that they are true determinants of quality.
- **Outcomes:** In many ways, the "outcome of care" achieved by a health care institution or provider is the most critical information desired by stakeholders, especially the patient seeking to choose where to receive care and from whom. Defining outcomes and comparing them among providers is exceptionally challenging. There are confounding variables to consider when looking at even the simplest of outcome measures. For example, when evaluating mortality, does one examine the results at one week, one month, or one year? When does the time period for procedural mortality end? Depending on the methodology, various public health and systems issues may affect the results. The Center for Medicare and Medicaid Services (CMS) has adopted a 30-day mortality definition, which is designed to reduce regional and institutional variability.

Outcomes of care also are determined by a multitude of factors that include, but are not limited to, the capabilities or performance of the institution or organization as well as patient characteristics. Patient populations are different, making comparisons difficult. Risk or Severity Adjustment methodologies are employed to "normalize" the patient populations. At present no standard approach is available, and many of the adjustment algorithms are proprietary, which may pose problems for institutions trying to understand their own results and may create confusion for consumers trying to evaluate the results.

Outcomes data typically are obtained from administrative data sets, which are used primarily for billing purposes and regulatory submissions. Given the process by which data are collected, there is tremendous variability in the accuracy of administrative data. For example, the data are dependent on a process by which a medical records coder reads the medical record (often hand written) and, based on the presence of certain key words, converts the clinical narrative into numerical code. Frequently, symbols and abbreviations are not sufficiently documented. Variability in interpretation and documentation can result in significant differences in coding, which affect any subsequent risk adjustment. This is especially the case with secondary and comorbid conditions that are the basis for the majority of risk-adjustment algorithms.

Recently, the federal government mandated the inclusion of "present on admission" codes that indicate whether a particular comorbid condition was present at the time of admission versus having occurred during the inpatient stay. For example, was the patient admitted with stage II pressure ulcer, or was the pressure ulcer a result of the patient's hospitalization? The answer to this question will influence both risk-adjustment calculations as well as hospital reimbursement. Of course, some outcomes of care measures are well constructed and do provide value to health care institutions, consumers, and payers. Examples include outcomes for children with cystic fibrosis (CF) reported through the CF Foundation [7] and central line–associated blood stream infections.

- **Process Measures:** In contrast to outcomes measures, process measures assess whether particular elements of care, judged to be standard of practice, were provided to the patient. Process measures differ from outcomes measures in that they usually are abstracted and aggregated directly from the medical record rather than converted into numerical code. Data definitions are as important for process measures as they are for outcomes measures. Process measures are useful for organizations as part of their improvement efforts. Most focus on increasing the reliability and consistency of care being provided by teams working together in clinical settings.

The use of process measures for the purpose of public reporting is increasing, but it is of uncertain value to consumers, and clearly places a substantial burden on the institutions and organizations that are required to abstract them. Advancements in electronic medical records potentially will mitigate this burden; however, it is critically important that quality and patient safety leaders are involved in the design and implementation of these systems, as the usual focus is on the transactional nature of these electronic systems, rather than on data functionality. Furthermore, additional work is needed to ensure that extraction of data is both efficient and accurate.

IS MORE ALWAYS BETTER?

Over the last few years the number of quality and patient safety indicators has increased substantially. Careful attention to data definitions is critical to enhancing the value and utility of these measures to avoid excessively taxing precious organizational resources. For example, externally mandated submission of specific indicators may not be aligned with institutional focuses and priorities. Hospitals and their staff have a finite capacity for aggregating and analyzing data, particularly in times of constrained budgets. Institutional resources must be conserved and appropriately focused to achieve quality and patient safety goals. A thoughtful, evidence-based, streamlined system is needed to create and evaluate databases and registries.

In some situations, indicators are routinely collected, some over prolonged periods of time, with little variation in the results and often no attempt to act on them. Often, critical appraisal of the process or of its outcomes is lacking, thus creating a stagnant quality assurance program. We advocate the concept of determining whether an indicator is actionable—that is, whether it will drive performance improvement. Although there may be value in following some indicators even without a performance improvement intervention, it may be preferable to discontinue collection of some. Alternatively, sample size and intervals may be adjusted in an effort to conserve resources.

Careful attention to data definitions is critical to enhancing the value and utility of these measures. After all, each measure has limitations and disadvantages as well as advantages. Indeed, the recent proliferation of quality indicators may have unintended consequences—that is, creating challenges for the reporting institutions and not necessarily achieving the results a well-constructed measurement program should achieve. Numerous criteria for optimal measurement systems have been proposed, including the following: (1) the measures must be understandable to the public; (2) the measures must be related to high-priority clinical conditions; (3) the data are accessible, accurate, and permit comparisons between and among providers; and (4) an evidence-based approach to treating the condition should be embraced to ensure that the measure is then actionable.[8]

MEDICAL ERROR AND PATIENT SAFETY

Health care quality goes hand in hand with patient safety. The thrust in the 21st century is the assurance of a safe workplace for patients and for providers alike. In 1999, the IOM published a seminal report entitled "To Err Is Human: Building a Safer Health System,"[9] which catalyzed an enormous shift in our attention and understanding of medical errors. The IOM report defined an error as an event in which there is a failure of a process to achieve the intended outcome, or an incorrect process of care was selected initially. An adverse event was defined as an injury to a patient caused by medical management rather than the patient's medical condition. The IOM report concluded that medical errors were responsible for as many as 98,000 deaths in the United States annually. Estimated annual costs of these errors were in the range of $17 billion to $29 billion. The report further opined that injuries caused by errors are inherently preventable.

The IOM report was based on two seminal studies: the Harvard Medical Practice Study [10] and another by Thomas et al. focusing on adverse events in Colorado and Utah.[11] The Harvard Medical Practice Study examined more than 30,000

medical records of patients discharged from 51 different hospitals in New York State, and found that adverse events occurred in 3.7 percent of hospitalizations. The researchers postulated that more than half of these adverse events (58 percent) were preventable. In 2000, Thomas et al. published the results of a review of 15,000 hospital discharges from hospitals in Colorado and Utah. The researchers found an overall adverse event rate of 2.9 percent, of which 50 percent of the adverse events were preventable.

Similar results have been shown in other studies. A Canadian study published in 2004 found an adverse rate event of 7.5 percent across 20 hospitals in Canada; almost 40 percent were deemed to be preventable.[12] From a public health perspective, these findings are just the tip of the iceberg because almost all of the studies focus on inpatient stays. Since the majority of health care in the United States occurs in the outpatient setting, it is difficult to assess the magnitude of the problem of adverse events and preventable errors across the continuum of care. Clearly, there is a need for more studies in out-patient care settings.

PERFORMANCE IMPROVEMENT IN HEALTH CARE

The 21st-century health care system is highly complex, making it more than likely that an error or an adverse event could occur. Systems are defined as many interdependent parts working together to achieve an outcome. System failures may be attributable to both human and nonhuman factors. A number of models designed to better understand system failure have been utilized in health care settings, the most notable of which is the Reason's Swiss cheese model.[13] This approach acknowledges that systems may have multiple levels or layers of protection only one of which is required to prevent from an error reaching a patient. All too frequently each of these layers fail—that is, the holes in the slices of Swiss cheese line up, resulting in the error reaching the patient. System improvement is contingent on identification of the factors that may cause each of these layers to fail.

The recent shift in focus toward systems and how they fail has not addressed a critical issue in health care, namely, the notion of individual accountability. Paradoxically, health care as an industry has been criticized in the past for not being aggressive enough in weeding out low-quality or incompetent clinicians. A good systems approach looks beyond individual culpability. Concepts such as a blameless culture have been proposed primarily because almost all errors rarely are caused by one individual. More likely, the error occurred because of a system failure; yet, in practice, it is difficult to harmonize more aggressive action against those who deserve it while simultaneously promoting an open and transparent environment, in which the staff feel comfortable discussing and learning from their mistakes. The health care industry, however, can learn from others who have pioneered safety and quality initiatives. For example, the nuclear, aviation, and manufacturing sectors have focused on communication, team functioning, reliability, and error-reporting systems to reduce errors, resulting in excellent outcomes.

ERROR-REPORTING SYSTEMS

Mirroring corporate America, numerous approaches have been proposed to improve performance, minimize errors, and reduce unnecessary costs in the health

care setting.[14] Although many of these initiatives have demonstrated substantial value, the authors believe that rarely will a single methodology sustain an organization over time. Public health leaders and health care executives, however, must exercise caution to avoid dramatically changing the methodology too frequently, thus minimizing staff buy-in because of a perception that the program will be short-lived. Avoiding the appearance of the "initiative du jour" is crucial. We believe that employing a single approach as a foundation with the addition of elements from other methodologies over time is the most effective approach. A mechanic's tool kit is a cogent analogy. Starting with a high-quality tool kit, the master mechanic will then selectively add tools from other manufacturers as needed.

To build a safer environment, many hospitals have developed voluntary adverse event and near-miss reporting systems as an important adjunct to their patient safety programs. Theoretically, these systems provide valuable information; however, a significant limitation is the absence of the "true" incidence or occurrence rate because what is being counted is simply event reports, not the events themselves. This limitation is most significant when attempting to examine trends over time or to benchmark against other institutions. Is a significant change an actual change in occurrence or simply a spike, or dip, in reporting? Although these reporting systems are critical to effective hospital safety programs, they should be employed in the context of a comprehensive patient safety program.[15] In 2001, a second IOM report, "Crossing the Quality Chasm," described six components of a quality health care system: effectiveness, efficiency, safety, patient-centeredness, equity, and timeliness.[16] These components have substantially altered our concepts of quality and safety, resulting in initiatives that were not even contemplated 10 years ago.

PERFORMANCE IMPROVEMENT METHODOLOGIES: PUTTING THEORY TO PRACTICE

Six Sigma

Six Sigma, a quality management methodology, uses statistical tools to measure processes with the ultimate goal of reducing variability and defects.[17] Six Sigma can assist an organization in identifying the source of variability, defects, or errors, thereby facilitating the development of solutions. Sigma is the measure of variation that reflects how much a process deviates from expected performance standards. The Greek letter sigma (σ) is the symbol for standard deviation. The higher the sigma value, the less variation exists. Six Sigma is equal to 3.4 defects per million opportunities, while a value of 3 Sigma is equal to 66,807 defects per million opportunities. To illustrate the importance of low variation in a process, a typical hospital operating at 3.8 Sigma (99 percent perfect) would lose approximately 20,000 lab requisitions per year. A Six Sigma hospital operating at 99.9997 percent perfection would lose only seven requisitions per year.

Six Sigma uses the Define, Measure, Analyze, Improve, Control (DMAIC) approach to implement quality improvements. DMAIC is the process by which performance teams achieve their stated goals. *Define* requires that the improvement team understands and identifies factors that are critical to quality. Teams then can develop a "charter," which serves as the project plan and typically includes the

business case for improvement, statement of the problem, constraints or challenges to improvement, scope of project, players and responsibilities, and preliminary project plan. *Measure* is the phase during which process data are collected to quantify potential performance opportunities, thereby defining performance standards—that is, the values of acceptable performance. Measurement allows the team to validate data and define parameters for improvement before beginning a more complete in-depth analysis.

The *analysis* phase identifies sources of variation. An effective Six Sigma team will evaluate many potential causes of variation, preventing biases or past experience from influencing the team's thinking. Common areas for teams to investigate include *methods* (procedures or techniques), *machines* (technology and equipment), *materials* (data, instructions, forms), *measures* (data), *Mother Nature* (environmental elements), and *people* (how elements are processed together).

During the *improve* phase, teams identify an improvement strategy and pilot solutions, which if successful, will be implemented more broadly. The improvement strategy selected is highly dependent on the results of the preceding steps suggesting a sequential approach.

The final step for a Six Sigma project is *control,* which prevents the process from returning to its original state once the team completes its work. The control plan is essentially an audit and compliance functionality and should include both ongoing data monitoring as well as a defined reporting program. The success of a Six Sigma project, which can take up to six months to complete, relies heavily on the results of the control phase.

Six Sigma includes other tools that can be employed to promote change. For example, *work-out* is a highly structured, facilitated meeting designed to empower people to make decisions and drive change. A work-out approach frequently is employed to rapidly achieve consensus or to solve a complex problem. Another example, the *change acceleration process* (CAP), examines barriers to change and effectively works through those barriers to accomplish established goals.[18] CAP outlines the steps required to change a process: lead at the top, create a shared need, shape a vision, gain commitment, operationalize change, modify systems and structures, and monitor and control progress.

Lean

Lean is another management tool that has been utilized in manufacturing processes for many years, particularly in Japan. Lean thinking is based on the work of W. Edwards Deming, who is considered by many to be one of the founding fathers of the quality movement. He worked with Japanese industrialists during post–World War II Japan to change work processes. His approach was known as the 14 points, which stressed employee participation, reliance on data, and use of careful analysis to drive change.[19] In Lean, the definition of value is based solely on the customer. By minimizing steps that do not add value, waste can be removed and work completed more efficiently. Lean seeks opportunities to improve performance anywhere in the system, regardless of preestablished targets.

Lean thinking incorporates specific "tools" such as a *kaizen*, a multiday session during which the key constituents review current process and consider opportunities

for improvement. A kaizen begins with the creation of a value stream map, followed by development of processes that are less wasteful. It is important to stress that health care is an industry in which the consumer (patient) pays a high price for both the value of the produce, service, or treatment and the cost of waste (medical errors, delay in care, and so on). Lean helps hospitals create value and reduce waste.

Plan, Do, Study, Act

The Plan, Do, Study, Act (PDSA) methodology is based on the work of Walter Shewhart, an engineer. PDSA can be utilized in hospital performance improvement projects of varying size and scope, from small-unit-based to pan-institutional. An advantage of PDSA is that it is both flexible and highly structured, and is adaptable to many environments.[20]

Root Cause Analysis (RCA)

Although not a performance methodology, per se, an important tool in identifying sources of error is the process of Root Cause Analysis (RCA). Currently, the most common approach in health care is to determine the underlying cause of adverse events. It is imperative to evaluate and to understand the circumstances leading up to an event that causes, or almost causes, serious harm to a patient. The RCA is a linear evaluation of the event during which staff are interviewed and both human and systems issues are identified.

Regulatory bodies often have specific requirements for RCAs, although these may differ across states and organizations. Most formats require that the institution identify corrective actions in order to reduce the risk of a similar event in the future.[21]

The value of RCAs is frequently limited in that incidents rarely have a single root cause and can occur repeatedly because of other factors not unearthed during review of the initial event. If a second similar event occurs, institutions should move to a different process.

Failure Modes Effect Analysis

Failure Modes Effect Analysis (FMEA), increasingly used in health care, differs from the RCA process in that it attempts to identify all potential risks in a particular process. In industry, design FMEAs have been widely used to examine product components; the Joint Commission (formerly the Joint Commission on Accreditation of Healthcare Organizations) now requires FMEA as a systematic, proactive method to improve safety and reliability.[22] The FMEA process involves several key steps: (1) identifying steps in the process, (2) identifying failure modes (what could go wrong?), (3) identifying failure causes (how could it go wrong?), and (4) failure effects (what would be the consequences of the failure?). FMEA methodology assigns risk to a system or process based not only on the probability of failure, but also on the impact of failure on the eventual outcome.

Tools associated with FMEA include the "Five Whys," a method for identifying the "root" of a problem. By repeatedly asking why a variation or error occurred, you can retrace the steps in a process that led to the system failure. The required components of an FMEA exercise include identification of potential failure mode;

the effects for each failure mode; a root cause analysis for the most critical effects; development, testing, and implementation of system changes to reduce risk; and monitoring the impact of changes. FMEA exercises typically involve a multidisciplinary team with varying levels of experience. At times, an outside facilitator can be brought in to keep the process flowing smoothly and to avoid intragroup conflict. FMEA tools are available on the Web site of the Institute for Healthcare Improvement.[23]

Risk Resiliency

Risk resiliency is a new approach that combines strengths of the RCA process with those of FMEA. Initially presented in 2007 at the annual meeting of the Institute for Healthcare Improvement, this method currently is being employed and evaluated at a number of health care organizations. It focuses on proactive thinking about system defenses and the adaptations required to minimize risk.[24] The process begins with identification of existing "predesigned systems" that are in place to prevent errors. The next phase of the process involves an assessment of adaptability and escalation—that is, teams are asked to consider how risk or error is recognized, whether escalation occurs, and whether environmental factors could have been identified in advance.

Before implementing any methodology, the institution's leadership must evaluate carefully the principles of the proposed approach for compatibility with the organizational culture and philosophy. Leadership must be willing to reevaluate the organizational structure, reduce hierarchical layers, and reorganize staff based on operational products or services.

QUALITY AND PATIENT SAFETY: THE KEY PLAYERS

Few fields in public health are as dynamic as quality and patient safety (Q&PS). Evolving national and regional policy, the availability of new scientific knowledge, and innovative approaches to performance improvement have resulted in the introduction of new initiatives and indicators. Health care institutions are faced with the challenge of setting strategic agendas in Q&PS, making informed decisions about participation with external entities, and coping with the financial consequences of this new environment. The resulting panoply of activities often fails to align and, at times, these activities substantially conflict with each other. This has resulted in a somewhat fragmented approach to quality measurement and has provided the impetus for an alignment of quality indicators and priority setting at the national level. Following is a brief listing of some of the relevant entities in the field of Q&PS.

Centers for Medicare and Medicaid Services

CMS's focus on quality and patient safety evolved following the publication of the 1999 IOM report "To Err Is Human: Building a Safer Health System." One result of the report was congressional authorization of "pay-for-reporting" programs, which required hospitals to submit data on 10 quality indicators, with financial consequences for nonparticipation. CMS has since added new measures and currently makes this information available through the Hospital Compare Web site. Evidence suggests that public reporting of quality data has led to an improvement in

quality and patient safety, especially when linked to payment incentives, such as pay-for-performance programs.[25]

In 2006, Congress further authorized CMS to reduce payments for "hospital-acquired conditions" (for example, central-line infections). Under the prospective payment system of diagnostic-related groups (DRGs), hospitals generally are paid an average rate for each illness, regardless of actual expense. Under the DRG system, however, a case with a complication or comorbidity, such as an infection, is paid at a slightly higher rate to cover the additional costs of care. Under new rules, however, the incremental reimbursement would be prohibited if certain complications are not "present on admission" but develop in the hospital.

Peer Review Organizations

Peer Review Organizations (PROs) were established by Congress in 1984 to improve the accountability of Medicare and Medicaid programs. PROs are contracted by CMS to conduct peer reviews of clinical services provided by hospitals and individual practitioners. Specific focus areas are defined annually, although the review functions are conducted regionally. PROs cite providers for lapses of care and frequently request corrective action plans.

National Quality Forum

Established in 1999, National Quality Forum (NQF) is a "voluntary consensus standard setting body" as specified by the National Technology and Transfer Act of 1995. NQF endorses indicators and compiles "best practices."[26] The NQF is a nonprofit membership organization of several hundred members, including representation from providers, purchasers, health plans, and consumers. Consumers and purchasers hold a majority of seats on the board. NQF standards and indicators are developed by expert panels and are subject to comment by interested parties. CMS and other purchasers use NQF metrics, including the 400 indicators approved to date, as the "gold standard." NQF recognizes the need to harmonize or align measures, with the understanding that misaligned or conflicting measures will impose additional burdens on providers and lead to confusion for consumers. As such, CMS now adopts new performance measures only after they have been reviewed and endorsed through the NQF process.

Agency for Healthcare Research and Quality

The Agency for Healthcare Research and Quality (AHRQ) is an arm of the U.S. Department of Health and Human Services that funds health services research as well as studies of outcomes and effectiveness. Research goals include a means to reduce medical errors and improve patient safety as well as to develop effective methods for the delivery of quality care. AHRQ also maintains the National Guideline Clearinghouse, which is a free, Web-based compendium on specific illnesses and treatments. AHRQ provides oversight for the Patient Safety Task Force, and provides metrics and measurement tools to health care organizations.[27] AHRQ has published several sets of indicators, including the Inpatient Quality Indicators and the Patient Safety Indicators.

The Joint Commission

The Joint Commission is the primary accreditation body for hospitals and other institutional providers, deriving its authority from CMS in determining that providers are meeting the "conditions of participation" for Medicare reimbursement. Founded in 1951, the Joint Commission is an independent, nonprofit organization that accredits provider organizations in the United States by onsite surveys utilizing teams of physicians, nurses, and administrators and assesses each organization based on a comprehensive list of standards. A key element of the survey process is the concept of a "tracer" during which the course of a patient or provider is traced against the Joint Commission standards. Participating providers are required to periodically submit data to the Joint Commission regarding compliance with published standards (see their "Quality Check" Web site at www.qualitycheck.org).

The Leapfrog Group

The Leapfrog Group was created by a consortium of employers, the Business Roundtable, and is largely supported by large employers. The stated objective is to maximize employer purchasing power and to recognize and reward quality providers. By identifying "better" providers, employers hope to steer their employees to the high-quality providers.

CONCLUSION

Quality and patient safety has increasingly become a societal focus in all aspects of health care. Galvanized by the findings of the IOM reports on the topic, health care organizations have been driven by a desire (need) to ensure a high-quality, safe environment for providers and for patients. Although other industries have led the way in this area, the health care industry is rapidly adopting quality initiatives, such as Six Sigma, Leapfrog, and the like. In these competitive and economically challenging times, setting strategic priorities in quality and safety is an important driver of success.

Quality assurance is not a static entity. Through partnerships among providers, payers, government, industry, and consumers, the efforts made over the past decade or so can be improved upon. Achieving better outcomes for patients, lower overall costs, and improved patient experience will require the continued investment of time and money to spur innovation and create reliable effectiveness, safety, and efficiency in clinical settings. Measurement and continuous performance improvement are the mainstays of a robust organizational quality assurance program.

REFERENCES

1. Institute of Medicine, Committee on Quality Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
2. Best M, Neuhauser D. Avedis Donabedian: father of quality assurance and poet. *Qual Saf Health Care*. 2004;13:472–473.
3. Donabedian A. Evaluating the quality of medical care. *Milbank Mem Fund Q*. 1966;44:166–206.

4. Anderson O. Influence of social and economic research on public policy in the health field. *Milbank Mem Fund Q*. 1966;44:11–48.
5. Codman EA. *A Study in Hospital Efficiency*. Iowa City: Classics of Medicine Press, 1992.
6. Helm E, Lee C, Chassin M. Is volume related to outcome in healthcare? *Ann Intern Med*. 2002;137:511–520.
7. Cystic Fibrosis Foundation. http://www.cff.org.
8. The Healthcare Association of New York State Quality Institute. *Which Quality Improvement Initiatives Are Right for Your Organization*. New York: The Healthcare Association of New York State Quality Institute; 2008.
9. Institute of Medicine. In: Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 2000.
10. Brennan TA, Leape LL, Laird NM, et al. Incidence and adverse events and negligence in hospitalized patients: results from the Harvard Medical Practice Study I. *N Engl J Med*. 1991;324(6):370–377.
11. Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care*. 2000;38(3):261–271.
12. Baker GR, Norton PG, Flintoft V, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Can Med Assoc J*. 2004;170(11):1678–1686.
13. Reason J. Human error: models and management. *BMJ*. 2000;320:768–770.
14. Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *N Engl J Med*. 2006;355:2725–2732.
15. Cullen DJ, Bates DW, Small SD. Incident reporting system does not detect adverse drug events: a problem for quality improvement. *Jt Comm J Qual Improv*. 1995;21:541–548.
16. Institute of Medicine, Committee on Quality Health Care in America. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press, 2001.
17. Pande P, Holpp L. *What Is Six Sigma?* New York: McGraw-Hill; 2002.
18. Smith IJ, ed. *Using Performance Improvement Tools in Healthcare Settings*. Oakbrook Terrace, IL: Joint Commission Resources; 2007.
19. Going *Lean in Health Care*. IHI Innovation Series white paper. Cambridge, MA: Institute for Healthcare Improvement; 2005. (Available on www.IHI.org)
20. Walter Shewhart-The Grandfather of Total Quality Management. http://www.skymark.com/resources/leaders/shewart.asp#
21. Percapio KB, Watts BV, Weeks WB. The effectiveness of root cause analysis: what does the literature tell us? *Jt Comm J Qual Patient Saf*. 2008;34:391–398.
22. Failure Mode and Effects Analysis. Info Center. http://www.fmeainfocentre.com.
23. Institute for Healthcare Improvement. http://www.ihi.org.
24. Frey K, Hyman D, Resar R. Evaluation of the safety system surrounding adverse events. http://www.ihi.org. Published 2007.
25. Lindenauer PK, Remus D, Roman S, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med*. 2007;356:486–496.
26. Leape LL, Berwick DM, Bates DW. What practices will most improve safety? evidence-based medicine meets patient safety. *JAMA*. 2002;288(4):501–507.
27. Agency for Healthcare Research and Quality. http://www.ahrq.gov.

CHAPTER 3

Consumer-Directed Health Plans, Managed Care, and Future Directions in the Organization of Health Care Payment Systems

William D. White, PhD

INTRODUCTION

Health care is not only expensive, but also highly variable—that is, it often is difficult to predict the need for medical care. Health insurance provides a way to mediate risk and provide financial access to care. A central challenge for insurance schemes, however, is how to meet the goals of mediating risk and ensuring access to care while containing costs. One response is to turn to regulatory solutions and public systems of control. In private insurance markets in the United States, however, the approach has been to rely on market-based strategies whose premise is that costs can be constrained by creating incentives for cost-conscious shopping, but also being mindful that insulating consumers from the financial consequences of purchasing decisions may distort their behavior, leading to the inefficient use of care and increasing costs. To design more efficient and effective payment schemes, it is important to focus on who should pay, and how much each party should pay.

Until the 1980s, most people with private insurance in the United States were covered by traditional indemnity plans. As remains the case in the 21st century, the vast majority of individuals got their coverage through employment-based plans provided as a tax-exempt benefit. These indemnity plans delegated shopping decisions about what care to buy and where to buy it to individual consumers and their physicians and then relied on consumer cost-sharing to contain costs. Specifically, plans used deductibles and coinsurance to create financial accountability for purchases; the notion was that responsibility for resulting out-of-pocket payments would create incentives for cost-conscious shopping. By design, health plans were relegated to a passive role of paying the bills, while providers were reimbursed fee-for-service on a cost basis.[1]

Indemnity insurance substantially mediated financial risks for consumers compared with direct cash out-of-pocket payment, which was the primary method of paying for health care services in the United States until the 1930s. It also was associated with rapidly rising costs, however. Critics argued that a key factor in this

rapid cost growth was the design of indemnity plans and their reliance on consumers to shop for care. In particular, they offered two major critiques. The first was that existing levels of cost-sharing created inadequate motivation for consumers to shop for care—that is, that deductibles and coinsurance were too low to provide incentives for aggressive shopping. The second critique was that individual consumers were ill equipped to be effective shoppers because they faced prohibitively high costs in gathering and evaluating the price and quality information needed for informed shopping.[1] Some critics also raised concerns about consumers' judgment in responding to economic incentives, noting the possibility of myopic, short-sighted, "pennywise, pound foolish" behavior, such as putting off preventive services like screening tests that could yield large potential long-run benefits in order to realize short-term savings.[2]

One solution to the problem focuses on increasing the amount of consumer cost-sharing. Simply increasing consumers' exposure to out-of-pocket payments, however, conflicted with the goal of mediating risk. Instead, beginning in the early 1980s, the policy response was to alter dramatically market organization by shifting responsibility for shopping away from consumers and toward health plans. Under the banner of what has come to be known as "managed care," the rationale for this shift was twofold. First, the plans would act as more motivated shoppers than individual consumers because they would realize the full amount of any costs savings. Second, plans would be more effective shoppers than consumers because they would be able to realize economies of scale in gathering and evaluating information. A corollary claim was that if plans took over shopping for care, this would reduce the need for traditional consumer cost-sharing to contain costs, thus allowing additional mediation of risk.[3]

Managed care enrollments grew rapidly in the 1990s. In the 21st century, most Americans with private health insurance have coverage through a Managed Care Organization (MCO), such as a Health Maintenance Organization (HMO) or Preferred Provider Organization (PPO). The design of these MCOs varies considerably. But most combine relatively low levels of consumer cost-sharing with two major types of strategies seeking to constrain consumer choice. The first strategy is to impose constraints on *where* consumers can buy care and their choices of physicians and hospitals. MCOs typically selectively contract with individual providers on the basis of price and quality. They then use financial incentives to "steer" enrollees to providers in their contracting "network," where these incentives usually are framed in negative terms. Thus, HMOs usually "refuse" to reimburse for services from "out-of-network" providers unless specially authorized, leaving the consumer to foot the bill in full out of pocket. PPOs "allow" use of out-of-plan providers, but reimburse for these providers' services at substantially reduced levels—for example, if Hospital A is not in your PPO's network, you can use it, but you must bear a much larger share of the costs than at an in-network hospital.

The second major strategy is to impose constraints on choices by consumers and their physicians about *what* care to use through administrative controls and mechanisms, such as utilization review, case management, and clinical guidelines. Again, incentives associated with these efforts usually are framed in negative terms. Thus, failure to adhere to a plan's rules can lead to "denial" of reimbursement.[4]

Although managed care currently dominates private insurance markets, in recent years, it has faced major internal issues. Since the mid-1990s, consumer "backlash" against MCO constraints on choice has been increasing.[5] At the same time, after a period of decline following expansion in MCO enrollments in the early 1990s, growth in private health insurance premiums has reaccelerated.[6] Together, these trends have lead to a reexamination of strategies associated with managed care.

One response has been to seek to modify managed care strategies from within. Many MCOs have responded to consumer dissatisfaction by reducing constraints on choice.[7] For example, many MCOs have broadened networks to include additional hospitals and physicians, potentially reducing plan bargaining power. Many MCOs also have restructured utilization controls in two ways. First, they have sought to reduce controls on relatively low-cost services that tend to be used frequently by enrollees and often are an important source of tension without necessarily resulting in large savings. An example is dropping requirements that enrollees obtain permission from their primary care physician to get a referral to a specialist. Second, MCOs simultaneously have sought to reposition efforts to control costs for high-cost services involving major illness and to make these controls more consumer friendly. An example includes efforts to restructure case management programs for diseases like diabetes and congestive heart failure. Key themes have included increasing consumer "engagement" and developing more cooperative relationships with patients in seeking to manage their care.[8]

These measures may have eased consumer discontent. However, continuing questions remain about MCOs' future ability to contain costs. Reflecting these concerns, a second response to managed care's difficulties has been a growing interest in alternative, consumer-oriented strategies and, in particular, in what have come to be known as Consumer Directed Health Plans (CDHPs).

The thrust of CDHPs is to seek to put consumers back in charge of shopping for care. A major challenge for them is how to deal with the concerns about consumers' motivation and ability to shop that originally motivated managed care. Evoking an upbeat, pro-consumer rhetoric, CDHPs specifically seek to address these concerns through three major types of strategies. The first, signature strategy of CDHPs is to combine (1) a high-deductible catastrophic insurance policy, for example, with a stop-loss limit on total out-of-pocket spending of $10,000 or more, with (2) a tax-advantaged individual- or employer-funded savings account that can either be used to pay for current expenses or rolled over for future use. The basic notion is that by placing consumers at risk for paying for substantial amounts of care with their own money, this simultaneously will restore control over shopping decisions and increase consumers' motivation for cost-conscious shopping. Additionally, introducing savings options will serve to mediate the accompanying increase in exposure to financial risk. The second major strategy is to respond to concerns about consumers' ability to shop by providing them with Web-based price and quality informational tools to "empower" them in the marketplace. A third major strategy is to create incentives for consumers to use targeted services, such as preventive and chronic care, where the focus is on offering "carrots" to encourage the use of desired services, versus managed care's use of negative incentives as "sticks" to discourage the use of services plans deem undesirable.[9]

To date, enrollments in CDHPs have been modest; as of 2008, only 8 percent of workers with employment-based coverage were enrolled in a CDHP.[6] At least some advocates, however, have envisioned CDHPs as a vehicle for major health coverage restructuring.[10] For example, Scandien sees CDHPs and consumer-oriented strategies as a way "to begin to roll back the role of third-party payment and restore the control of resources to individual consumers."[11] Alternatively, Robinson et al. suggest that strategies associated with CDHPs may end up primarily complementing existing institutional arrangements and managed care.[12]

From a policy perspective, important questions include not only how CDHPs actually work, but also what role they could play in the organization of private (and public) health care payment systems in the future. This chapter considers the mechanics of CDHP plans and discusses possible design issues, examines how efforts to implement CDHPs have fared to date, and considers the possible implications of CDHPs for the organization of payment systems in the future.

Before proceeding, a brief note on language is in order. This chapter follows the convention in the consumer-oriented literature of describing patients as "consumers." From an economic perspective, this term has the advantage of encompassing both current and would-be buyers of services. It is used advisedly, however, because the term "consumer" is often used proscriptively in connection with arguments for increasing patients' involvement in actively "shopping" for care.[10]

PLAN DESIGN

Specifically turning to the design of CDHPs, a useful starting point is to begin by considering each of the main types of strategies associated with CDHPs.

Motivating Shopping

CDHPs seek to combine increased cost-sharing with tax-advantaged savings plans to simultaneously shift responsibility for shopping back to consumers, increase their motivation to shop, and mediate risk. To accomplish the first two of these goals (shifting responsibility for shopping and motivating shopping), CDHPs typically offer their enrollees high-deductible catastrophic insurance plans.

High-Deductible Catastrophic Plans

Catastrophic plans have similar designs to traditional insurance policies (a deductible and partial cost-sharing above this deductible up to a stop-loss limit) and when offered in the workplace, enjoy the same kind of tax advantages enjoyed by other employment-based coverage. They expose enrollees to higher levels of financial risk, however. For instance, consider a catastrophic insurance policy with a $5,000 deductible, a coinsurance rate of 20 percent, and a stop-loss limit on total expenditures of $10,000. In this case, a consumer will be fully responsible for paying for their first $5,000 of care out of pocket. Once their expenditures reach $5,000, they will be responsible for 20 percent of each additional dollar of expenditures until their total expenditures equal $10,000 ($5,000 for the deductible plus $5,000 in coinsurance payments). At this point, catastrophic provisions will apply and their insurance will pay in full for any additional expenses.

With a high deductible, consumers will be responsible for paying for substantial amounts of care with their own money and indeed, for the majority of consumers, the deductibles associated with CDHPs are likely to exceed their expected annual expenses. A direct implication is that at least until their deductible is met, consumers will be free to spend their money as they like without worrying about any administrative restrictions imposed by MCOs. Since they are fully responsible for their expenses up to their deductible, however, consumers also will have strong incentives for cost-conscious shopping. Specifically, potential savings may occur in two ways: (1) through selecting lower-cost providers; and (2) through selecting less costly treatment options. For example, the consumer chooses an X-ray over an MRI and elects to get it at a freestanding imaging center that charges less than the local hospitals.

Tax-Advantaged Savings Accounts

The problem with higher cost-sharing is that by definition it will expose consumers to higher levels of financial risk. Savings accounts can help mediate this risk by providing consumers with funds that can either be used to meet current expenses or be rolled over for future use. For consumers with employer coverage, these savings plans could compensate for the fact that less care may now be covered by tax-exempt insurance. For consumers without employer-based insurance, savings accounts provide a way to access tax advantages for health care purchases.

Looking specifically at savings options, CDHP plans usually are classified into two main types: (1) plans that provide an employer-based Health Reimbursement Account (HRA); and (2) plans that meet necessary federal requirements for an enrollee to establish an individual Health Savings Account (HSA). HRAs are funded, administered, and owned by an employer in conjunction with an employer-sponsored insurance plan. They first became available on a tax-advantaged basis following a 2002 Internal Revenue Service (IRS) ruling permitting tax-exempt employer contributions.[13] In a given year, employees can use these funds to pay for health care expenses or roll them over from year to year. They generally cannot take HRA funds with them if they leave their job, however.

Typically, an HRA-based plan is defined as a CDHP if there is a deductible of $1,000 or above for single coverage and $2,000 or above for family coverage, where total risk exposure depends on a plan's stop-loss limit. In 2008, the Kaiser/Health Research Educational Trust (HRET) Employer Survey found the average deductible for HRA plans was $1,552 for single coverage and $3,057 for family coverage and that the stop-loss limit for eligible out-of-pocket expenditures (deductible plus any copayments) averaged $2,543 for single coverage and $5,331 for family coverage. At the same time, the average HRA employer contribution was $1,249 for single coverage and $2,073 for family coverage, implying a substantial "gap" between the amount contributed and potential out-of-pocket expenditures in a given year.[6] For instance, if the HRA contribution for a family policy is $2,000 and the stop-loss limit is $5,300, the potential "gap" will be $3,300.

In contrast to HRAs, HSAs are individually owned and typically administered through an outside institution such as a bank or insurer. Under the provisions of Title XII of the Medicare Modernization Act of 2003, HSA plans may be established either in connection with an employer-sponsored health insurance plan or on

an individual basis. In either case, to set up an HSA, an individual must be enrolled in a health plan that meets two types of qualifications: a deductible equal to or above a minimum floor, and a stop-loss limit equal to or below some maximum amount. If the consumer elects to set up an HSA, the individual can make direct contributions to his or her accounts (either in qualified individual or employer plans). In addition, employers can contribute to employment-based plans. In either case, contributions are triple tax advantaged. Not only are no taxes paid on money put into HSA accounts, but also earnings on account balances and withdrawals from HSAs to pay for medical care are tax exempt. Furthermore, unlike accounts such as Investment Retirement Accounts, there are no mandatory withdrawal requirements after age 65.[13]

In 2009, for a health plan to be HSA qualified for single coverage, there needed to be a minimum annual deductible of $1,150 and a maximum stop-loss limit on total out-of-pocket payments (deductible plus copayments) of $5,800. Limits for family coverage were $2,300 and $11,600, respectively. At the same time, combined contributions from individuals and their employers could not exceed 100 percent of the deductible or a maximum of $3,000 for single coverage and $5,950 for family coverage. Finally, the purchase of first dollar supplemental coverage is prohibited.[13] Consequently, the potential again exists that in any given year a substantial gap may exist between the maximum allowed contribution and maximum possible total out-of-pocket payments, placing a consumer at risk for large "gap" payments.

Stepping back and briefly considering some potential design issues with these arrangements, a key question is whether the schemes described here can deliver on their promise to either motivate shopping or mediate risk. First, consider issues with financial risk. Placing a consumer at risk for large "gap" payments may permit reductions in premiums compared with traditional insurance. However, critics such as Jost [14] argue that CDHPs may have the effect of redistributing risk toward sicker consumers, especially those experiencing serious ongoing health problems who are likely to repeatedly reach levels of spending requiring large out-of-pocket payments.

The potential for large out-of-pocket payments has at least two important implications. First, while CDHPs may offer lower premiums overall, especially if savings are achieved from reduced utilization, moving to a CDHP may impose substantial hardship for sicker enrollees. Second, to the extent consumers have the option of choosing between insurance plans, imposing substantial risks on sicker individuals may lead to problems with risk selection in insurance markets. In particular, if CDHPs tend to systematically attract healthy consumers while creating incentives for sicker consumers to turn to plans with more generous coverage, this may set off a downward spiral in which high-risk consumers gravitate to more generous plans, pushing up their rates and potentially driving them out of the marketplace.[14]

Turning to issues with incentives, consumers clearly face strong incentives for cost-conscious shopping until their deductible is met, while to a lesser extent, coinsurance continues to create incentives until a stop-loss limit is reached. What happens, however, when they anticipate exceeding this limit? Given 100 percent reimbursement for catastrophic care, there clearly is no incentive to engage in cost-conscious shopping for expenditures above this limit. Furthermore, if a consumer anticipates

exceeding their stop-limit, there is no incentive for expenditures below this limit either—they will end up making out-of-pocket payments equal to their maximum limit anyway.[15] As noted, the majority of consumers in CDHPs are likely to have expenditures below their deductible. Concerns about incentive effects at high levels of expenditure are salient, however, because of the highly skewed nature of health care expenditures. Although only a small share of the population may have catastrophic expenditures, these consumers account for the majority of health care costs. For example, in 1996, the top 10 percent of patients in the United States were estimated to account for nearly 70 percent of total expenditures.[16]

At the other end of the spectrum, a related concern is how consumers with relatively modest anticipated levels of expenditures may view funds in savings accounts. If they take a long-term perspective and treat these funds as reserves against unknown future events (that is, savings for a rainy day), this might provide incentives to conserve funds. If, however, they take a short-term perspective and treat these funds as cash in hand to spend down in full each year, this will largely defeat the goal of motivating cost-conscious shopping while mediating risk. Instead, only "gap" spending is likely to be effective, raising the question of whether a smaller deductible without a savings account might not be better.[17] A particular issue for HRAs is that funds may be lost if a employee changes jobs, which might create strong incentives to spend down, especially for younger, healthier workers who may anticipate finding new employment within a few years.

Finally, even if consumers are well motivated, their ability to realize savings from competitive shopping will be limited by their ability to exercise choice in the marketplace. If they are in an MCO, their choices may be sharply constrained by the extent of their MCO's network. They can, of course, still go outside this network but reimbursement levels will be lower. Individual consumers also are unlikely to be able to negotiate as favorable rates with providers as those negotiated by their MCO.

Empowering Shopping

The primary problem consumers face in being an effective shopper is the high cost associated with gathering and evaluating price and quality information. Consumer-oriented efforts to increase the amount of information available in the marketplace through Web-based tools could be helpful. For instance, the Centers for Medicare and Medicaid Services' (CMS) "Hospital Compare" Web site posts extensive information on hospital quality indicators (see http://www.hospitalcompare.hhs.gov). Key logistical issues include what information to provide and how to support its use. To be of value to consumers, price and quality data need to be timely and relevant to a consumer's specific purchasing decisions. For example, suppose a consumer wants to compare prices for angioplasties at Hospital A versus Hospital B. One indicator of cost is average prices. From a consumer's perspective, however, the relevant "price" is not the average price, but the amount they personally can expect to pay out of pocket for this procedure given the specific reimbursement provisions of their plan. Furthermore, from a consumer's perspective, the relevant price includes not only expenses associated with the procedure, but also related services and the overall "bundled" cost likely to be involved with the relevant episode of care. Finally, quality also may be an important choice variable and simply listing a wide range of disparate, hard-to-interpret measures

may be not be very informative.[18] One approach is to collapse a range of measures into a single one, but this risks imposing preference orderings that may not be consistent with those of individual consumers. A more general issue, discussed below, is the degree to which simply providing information will necessarily resolve concerns about consumers' ability to be cost-effective shoppers.

Incentives to Use Preventive and Chronic Care Services

Many CDHPs exempt preventive services from deductibles and coinsurance and pay for these services in full as long as they are provided within the plan's network. HSA legislation limits exemptions to preventive services. IRS rules for HRAs are more flexible and permit plans to also exempt routine services for enrollees with chronic illnesses like diabetes and asthma.[13]

The rationale for these exemptions is straightforward: that they will encourage utilization of preventive and chronic care services that can reduce future costs and lead to better health outcomes. For example, cancer screening may allow earlier detection, while routine management of diabetes and asthma may avoid costly hospitalizations. However, as Berenson [19] observes, this raises an obvious question: if these services are so desirable, why are consumers not using them in the first place and what does this say about their ability to shop for care? If the problem is that consumers do not understand the potential benefits of preventive and chronic care services, why not solve this problem by simply providing consumers with more information and education?

Example of the Operation of a CDHP

The mechanics of CDHPs are a bit complex; therefore, an example is useful to show how one actually works. Consider the case of Joe, who is enrolled over a three-year period in a single-coverage employer HRA CDHP. Suppose this plan is based in a PPO with a deductible of $1,000, a maximum stop-loss limit of $3,500, and an annual employer contribution of $1,000. Suppose further that in Joe's PPO, in-plan preventive services are exempt, and he would be responsible for a 20 percent coinsurance rate for expenses above his deductible for in-plan services. If he goes out of plan, the coinsurance rate is 30 percent, and some of his costs may not be eligible for reimbursement at all.[1]

Suppose in Year 1 Joe has no major health problems and spends a total of $500 on health care, all from in-plan providers, including $300 for exempt preventive services. In this case, it will be a good year for Joe. He simply will be at risk for the $200 he spent on nonpreventive services from a convenient care center. Because he has a deductible of $1,000, he will be responsible for paying the bill in full. Deducting $200 from his HRA, no out-of-pocket payments will be necessary and an HRA balance of $1,000 – $200 = $800 will be rolled over.

---

1. PPOs typically tie reimbursement to Usual and Customary Rates (UCRs) in the relevant geographic area and only reimbursement for charges up to this amount (or some percentage of it), where charges above this amount are the responsibility of the enrollee (for example, if the URC is $100 and the charge is $150, Joe would be responsible for 30 percent of $100 plus $150 − $100 = $50, or $80 total).

Note several important features of Joe's choices. First, assuming that Joe would like to preserve as large a balance as possible in his HRA for future use, there will be strong incentives in Year 1 for him to engage in cost-conscious shopping. By selecting a convenient care center, Joe is able to realize the full $100 of savings associated with this choice. Second, while Joe uses $300 worth of preventive services, if they had not been exempt and he had to pay for them himself, he might have purchased less or none of these services.

Turning to Year 2, if his employer again contributes $1,000, Joe's total HRA balance will now be $1,800. However, suppose in Year 2 Joe has major health problems—he has a heart attack early in the year and he ends up spending $21,000. Suppose Joe purchases all of his services from in-plan providers, and they are fully eligible for reimbursement. Joe will owe the first $1,000 of the cost of his care (the deductible), plus 20 percent of the cost of his care above his deductible. Without a stop-loss limit, he would be responsible for $1,000 + (0.20 × $20,000) = $5,000. Because of his plans stop-loss limit of $2,500, his out-of-pocket expenses will be limited to $3,500. This, however, still will place him at risk for an out-of-pocket expense of $3,500 – $1,800 = $1,700, and his HRA account will be zeroed out.

Although there is no change in plan rules, once Joe approaches his stop-loss limit, he will face quite different incentives for cost-conscious shopping than in Year 1. If Joe anticipates surpassing this limit, there is no incentive to worry about costs above it. He also has no incentive to worry about charges below the limit, unless he expects savings that will bring him below it. Otherwise, even if he takes actions to reduce his costs (say from $21,000 to $20,000), he will end up with total spending equal to his limit, $3,500.

Finally, suppose in Year 3 that Joe's employer again contributes $1,000 to his HRA, resulting in a new balance of this amount, and that Joe's heart continues to give him problems and he spends $8,000 on care. In this case, Joe will now owe $1,000 + (0.20 × $7,000) = $2,400, which he will need to pay in full because this is less than his stop-loss limit. He could take $1,000 from his HRA and pay the balance of $1,400 out-of-pocket. In this case, Joe may benefit from cost-conscious shopping. For example, if he switches to a lower cost in-plan provider who charges him only $7,000 instead of $8,000, a reduction of $1,000, he will be able to realize out-of-pocket savings of 20 percent × $1,000 = $200. Alternatively, suppose that Joe considers using an out-of-plan provider who also charges $8,000. In this case, he will have out-of-pocket expenses equal to his deductible plus 30 percent of charges, or $1,000 + (30 percent × $7,000) = $3,100, a $700 increase in out-of-pocket spending compared with $2,400 if he used his previous in-plan providers with identical charges. Unless Joe values his potential gain in quality more then $700, he presumably will stay in network.

Taken together, Joe's experiences over this three-year period illustrate several important features of CDHPs. First, by design, Joe is shielded from the kind of catastrophic loss that would have resulted if he had to pay for all care in full; however, in any given year, he is potentially at risk for substantial out-of-pocket expenses because of the $2,500 gap between his employer's contribution of $1,000 and his stop-loss limit of $3,500. His HRA account may reduce some of this risk exposure, especially if he is able to build up balances over time. But he still may be at risk for substantial out-of-pocket payments, especially if he has several bad years and exhausts any rollover balances. Second, Joe's incentives for cost-conscious shopping will vary with his needs

for care. Third, the apparatus of managed care continues to hover in the background, where payment rules create strong financial incentives for Joe to limit his search to in-plan providers. If Joe's PPO included case management for catastrophic illness, this could add an additional level of complexity. Finally, consistent with concerns about risk selection, "Joe healthy" is more likely to be attracted to a CDHP then "Joe sick," who, if he has a choice of plans, may prefer a standard plan with lower deductibles and less overall risk exposure.

IMPLEMENTATION OF CDHPs

Enrollments

Over time, opportunities to enroll in CDHPs have grown, but participation in these types of plans remains modest. Focusing on employment-based plans, the share of firms offering their employees the option of an HRA with a family deductible of more than $2,000 ($1,000 single) for an HSA-qualified plan grew from 4 percent in 2005 to 13 percent in 2008. In 2008, these firms accounted for 25 percent of all covered workers. During the same period, enrollments in CDHP, HRA, and HSA plans rose from 2 percent of covered workers, or one worker in 50, to 8 percent, or almost 1 worker in 12.[6] These data suggest that while access to CDHPs is now available to about one-quarter of American workers, less then one-third of the eligible workers (8 percent) actually enrolled in them.

Breaking down enrollment patterns by firm size, in 2008, larger firms were somewhat more likely to offer CDHP plans than small ones (13 percent for firms with fewer than 200 workers versus 15 percent for those with 200 to 999 workers and 22 percent for firms with 1,000 or more workers). The proportional share of workers actually enrolled in CDHPs in small firms was higher than in larger ones (13 percent of employees for firms with less than 200 workers versus 5 percent for firms with 200 employees or more).[6]

At least one explanation of higher participation in CDHPs in smaller firms is that these firms are more likely to limit workers' options to a single insurance plan. For example, in 2008, 86 percent of firms with fewer than 200 employees offered a single-choice plan. Consequently, in most small firms, if a CDHP is offered, it is likely to be the only insurance option available. In contrast, in larger firms, CDHPs are likely to be offered as part of a menu of alternative plans and a consumer need not enroll in one to get coverage. Looking to the future, survey data from employers suggest that rising insurance premiums could accelerate current trends, where 60 percent of firms with CDHPs in 2008 named cost as the primary reason for offering them.[6]

Self-Selection

Studies of enrollment patterns for CDHPs find evidence that CDHP enrollees tend be more educated, have higher incomes, and have lower levels of prior health care utilization and chronic conditions. Among HSAs, one factor that may be potentially attracting older enrollees is opportunities for tax-advantaged retirement savings, especially in the case of individuals with higher incomes. When consumers have a choice, the literature suggests that people are more likely to enroll in such plans when the CDHP actively seeks to educate them about plan features. Interestingly, CDHP

enrollees were considerably more likely to say they would seek to change plans if they developed a chronic condition requiring more care. For instance, in the case of Joe, he might look for a new plan if he anticipates continuing needs for care with his heart condition.[20]

These findings have several implications. First, they lend some credence to concerns about the possible market dynamics that could be associated with introducing CDHPs. Second, they suggest an important role for consumer education in promoting CDHPs. Consistent with this, 37 percent of small firms (less than 200 employees) and 71 percent of large firms (more than 200 employees) list "educating and communicating change in benefit" as the biggest challenge associated with introducing CDHPs.[6]

Empowering Shopping and the Provision of Cost and Quality Information

Efforts to implement strategies to "empower" consumers in the marketplace have focused on the provision of Web-based informational tools. Reports from the field indicate that most CDHP plans have some sort of Web presence. Interestingly, reports further suggest that, in many cases, insurers who develop Web tools in conjunction with CDHP products also tend to end up sharing these tools with all their enrollees, including those in conventional managed care plans.[18] Although plan provision of Web-based tools appears fairly widespread, existing evidence suggests that the use of these tools remains limited. For example, a 2006 Kaiser survey asked CDHP enrollees questions about information provided by their plans in a broad range of possible areas, including information on both prices and provider quality. In most cases, more than half of these enrollees did not know whether their health plan even had a relevant Web site. Furthermore, 10 percent or less of enrollees said they actually went to a plan Web site to get information on the cost or quality of providers, although a somewhat higher percentage (19 percent) indicated they used a plan Web site to look up drug prices.[20]

One factor that may explain this low level of utilization is simply that these tools are relatively new and have yet to gain much traction with consumers. But two other related contributing factors may be data quality and limitations on the kinds of shopping opportunities offered by CDHPs. Common complaints about data provided by Web tools include their timeliness and relevance as well as problems with interpreting these data. For example, price information may be outdated. In addition, it may not be useful for determining the actual amount a consumer can expect to pay if they purchase a service from a particular provider, where plans may be reluctant to post information that could reveal the prices they have negotiated with specific providers, which often is considered proprietary information. In addition, CDHPs rely heavily on MCO networks that may offer relatively limited opportunities for comparative shopping. This does not necessarily mean that consumers would not value accurate information about their potential costs of services or the quality of these services if it were made available.

Benefits Design

Cost-Sharing

As described earlier, a central goal of CDHPs has been to increase consumers' exposure to financial risk to motivate shopping. Pragmatically, an important question is

whether, as currently constituted, CDHPs significantly affect risk exposure. Historically, the growth of managed care has been associated with declines in consumer cost-sharing. Since the late 1990s, however, consumer cost-sharing has been increasing in conventional managed care plans. Consumers bear the full impact of this cost-sharing because out-of-pocket expenses usually are not tax exempt. In contrast, the tax benefits associated with tax-exempt savings accounts may be substantial even for relatively low-income consumers because the exemption includes not only income taxes but also payroll taxes. In addition, stop-loss caps may be high in conventional plans.[2]

Taking into account tax considerations, Remler and Glied [21] simulate the implications of representative payment designs. Despite the rhetoric of CDHPs, they conclude many consumers in conventional plans, including those accounting for the bulk of health care expenditures, would likely see either no change or a decrease in their cost-sharing if they switched to a CDHP. This suggests that if CDHPs are really going to have much impact on incentives at the margin, substantial changes may be required in cost-sharing arrangements.[22]

Managed Care

Theoretically, a CDHP could simply offer an insurance product that includes a traditional indemnity high-deductible catastrophic plan with a savings option, informational tools, and incentives to use preventive and chronic care services. Once an enrollee reaches the stop-loss limit, however, this kind of plan would not provide any mechanism for controlling expenditures. Furthermore, it would place consumers in the position of having to buy services on their own in the open market where individual consumers without an MCO may face substantially higher prices.

Linking a CDHP to an MCO potentially could provide a solution to both problems, in which case an MCO could offer two major types of complementary services. First, linking a CDHP to an MCO can enable the CDHP to give its enrollees access to the MCO's provider network and associated price discounts. Second, it also can provide access to an MCO's tools for controlling health care utilization. The Kaiser/HERT Employer Survey does not provide information specifically on whether CDHPs are similar to managed care plans or what their relationship looks like.[6] However, the literature suggests that currently most CDHPs have close associations with MCOs, especially with PPO plans.[23] Potentially, a CDHP could simply rent access to a PPO network.

Significant business advantages for CDHPs appear to be linked to MCOs with strong, established networks. For example, two CDHPs, Definitely and Lumenos, were bought out by major insurers (United Health and WellPoint, respectively).[12] Anecdotal evidence suggests that reliance on managed care tools for controlling utilization is common. For example, a description on the Web site of a New Mexico Blue Cross Blue Shield CDHP plan site lists among its leading features: "Integrated health care management including disease management programs, prenatal program, case management services."[24]

---

2. Using data from the Kaiser/HRET 2005 survey, Remler and Glied [22] find that 21 percent of workers have no maximum on out-of-pocket spending and 55 percent either have no maximum or a maximum of $2,000 or more.

Further research on business relationships between CDHPs and MCOs clearly would be desirable. The kinds of relationships described herein suggest that important complementary relationships are evolving between CDHPs and MCOs. Looking at some of the implications of these relationships, reliance on PPO networks offers CDHPs a way to provide network access without completely eliminating a consumer's option of going out of network without plan permission, as would be the case with an HMO. If a PPO has a large network, there may be some allowance for comparative shopping. But, this is still a far cry from a model in which individual consumers are the engines of accountability in the marketplace. The flip side is that at least from the perspective of selective contracting, linking up with a CDHP seems likely to make relatively little difference in how an MCO conducts its business, leaving its basic organizational structure intact. Specifically, if a PPO is already engaged in selective contracting with hospitals and physicians, the main effect may be to increase the number of enrollees the PPO represents and give it additional clout in the marketplace.[3]

Regarding utilization controls, CDHPs and MCOs may complement each other in at least two ways. First, tools like case management may help CDHPs to control costs for enrollees whose expenditures have exceeded stop-loss limits on cost-sharing. CDHPs could conceivably develop their own tools for this purpose. If, however, MCOs already have such tools in place, economies of scale may be achieved from sharing them, especially within organizations offering multiple insurance products. Second, aspects of CDHPs may help MCOs address some of the issues created by consumer backlash. As discussed earlier, one particular source of tension with consumers has been MCO restrictions on the use of relatively low-cost, frequently used services like referrals to specialists. With many plans pulling back from using these types of controls, high-deductible plans may provide a way for MCOs to fill this gap and induce consumers to self-ration these kinds of services—for example, if you are paying the full cost of self-referring to a specialist, you may think twice before doing so.

Finally, CDHPs' use of carrots versus sticks to encourage the use of services like preventive and chronic care may offer an alternative to the kind of punitive models used by many MCOs in the past. It may be more effective to offer to "pay" consumers to enroll in a disease management program or participate in case management rather than to threaten them with penalties.

CDHP Impact on Costs and Quality

At the employer level, CDHPs generally are perceived as saving on benefits costs. Consistent with this, in 2008, 42 percent of employers in the Kaiser/HRET survey reported that CDHPs offered lower costs,[20] and, not surprisingly, insurers use this as an important selling point for their CDHP.[25] Benefits costs only include costs borne by firms and not those borne by workers. Studies on the issue show some savings, although results are not uniform and there are issues with where the cost data were obtained.[22, 26, 27]

---

3. One proviso, however, is that providers may be more reluctant to accept CDHP enrollees then conventional PPO enrollees, because CDHP enrollees may be more likely to default on copayments to providers because of greater cost-sharing.

In interesting new work, Lo Sasso, Helmchen, and Kaestner address selection issues by looking at workers at small firms that only offered a single plan in which the specific CDHP plan included separate deductibles for outpatient services and pharmacy and inpatient and outpatient surgery spending.[17] They found that for outpatient services and pharmacy, raising deductibles reduced spending, albeit modestly. At the same time, increases in deductibles for inpatient and outpatient surgery had minimal effects. The first finding is consistent with expectations that increased cost-sharing will reduce utilization, especially for consumers with moderate levels of expenditure. The second lends support to the concern that in the case of high-cost services, such as inpatient hospitalizations for which costs are likely to exceed stop-loss limits, even relatively high cost-sharing may have little impact. What is more surprising about their findings is that spending increases dollar for dollar for increases in employer contributions to HRA and HSA savings plans. Based on these results, when consumers receive additional funds, they appear to turn around and spend it on care immediately, thus raising questions about the value of accounts in mediating risk.

Regarding quality, there is a growing concern about the effects of high levels of cost-sharing on consumer purchasing decisions. A number of studies examined utilization services for employees enrolling in CDHPs compared with those who did not. A limitation with these types of studies, as noted earlier, is that there may be self-selection between consumers electing CDHPs versus other types of plans. In any case, these studies find evidence not only of lower utilization but also of a negative impact on quality. For example, Hibbard et al. examined impacts on the use of medical office visits, where evidence-based criteria were used to categorize visits into "high priority" and "low priority." They found that CDHP enrollees had fewer office visits, regardless of whether they were classified as high priority or low priority, suggesting that when cutting back services, consumers did not distinguish between the two types of visits.[28]

In another study, Greene et al. found that CDHP patients were more likely to discontinue medications for chronic conditions.[29] Survey data compared enrollees in CDHPs with those in more traditional employment-based coverage The researchers found that CDHP enrollees were more likely to say that they did not get needed services because of the costs.[20] Furthermore, recent work by Reed et al. suggests that consumers in CDHPs often took actions to reduce utilization without having a good understanding of plan rules.[30] These findings suggest that as currently structured, higher cost-sharing associated with CDHPs may have problematic effects on quality. On the positive side, Buntin et al. found evidence that enrollment in CDHPs led to more frequent use of preventive services, which is consistent with plan strategies to encourage the use of these services.[31]

DISCUSSION

There has been considerable discussion about empowering consumers in the marketplace. Whereas the basic strategy of managed care has been to shift responsibility for shopping and financial accountability away from consumers toward third-party payers, the strategy of CDHPs is to shift both back to consumers. The discussion in this chapter suggests that rather than a revolution, the

introduction of CDHPs has been marked by a process of mutual accommodation, such as that discussed by Robinson [12] and Robinson and Ginsburg.[32] In practice, most CDHPs rely on MCOs to shop for care and then offer their enrollees access to a plan network, thus allowing MCOs to preserve their basic business model and continue with selective contracting intact. CDHPs also draw on managed care strategies like case management to control costs for high-cost patients whose expenditures exceed the relatively high stop-loss limits associated with these plans. At the same time, findings suggest the potential for adoption of some important features of CDHPs by conventional MCOs. For example, CDHP cost-sharing strategies offer a potential alternative for filling in at least some of the gaps left by MCOs' retreat from utilization controls.

CDHP efforts to incentivize use of preventive and chronic care also offer MCOs a model of how to reframe the rhetoric of managed care in more consumer-friendly terms while continuing to seek to manage behavior. One could argue that if current trends of integration of CDHPs and MCOs continue, distinctions between the two types of plans will become increasingly blurred. One possible scenario is that the process of integration will continue to the point at which MCOs' use of major strategies associated with CDHPs and CDHPs' use of MCO networks and utilization tools become so ubiquitous that there is no meaningful difference between them. The fact that tightly managed HMOs eschewing high levels of cost-sharing have maintained a substantial market share suggests the possibility of continued differentiation in the marketplace. As health care is actively being discussed and formulated, it will be interesting to see how CDHP will be integrated into the reformed delivery system.

REFERENCES

1. Starr P. *The Social Transformation of American Medicine*. New York: Basic Books; 1982.
2. Camerer, CF, Loewenstein G, Rabin M (eds). *Advances in Bahavioral Economics*. New York: Princeton University Press; 2004.
3. Dranove D. The case for competitive reform in health care. In: Arnould R, Rich R, White WD, eds. *Competitive Approaches to Health Policy Reform*. Washington, DC: Urban Institute Press; 1993.
4. White WD. Market forces, competitive strategies and health care regulation. *University of Illinois Law Review*. 2004;1:137–166.
5. Blendon RJ, Brodie M, Benson JM, et al. Understanding the managed care backlash. *Health Affairs*. 1998;17(4):80–94.
6. Kaiser Family Foundation, Health Research and Educational Trust. Employer health benefits 2008 annual survey. http://ehbs.kff.org/2008.html. Published September 2008.
7. Lesser C, Ginsburg P. *Back to the Future? New Cost and Access Challenges Emerge*. Washington, DC: Center for Health Systems Change; 2001. Issue Brief No.35.
8. Hurley RE, et al. Early experiences with consumer engagement initiatives to improve chronic care. *Health Affairs*. 2009;28(1):277–283.
9. Congressional Budget Office. *Consumer-Directed Health Plans: Potential Effects on Health Care Spending and Outcomes*. Washington, DC: Congressional Budget Office; 2006.

10. Herzlinger R. *Market-Driven Health Care: Who Wins, Who Loses in the Transformation of America's Largest Service Industry*. Reading, MA: Addison-Wesley; 1997.
11. Scandlen G. Consumer-driven health care: just a tweak or a revolution? informed, motivated consumers can have an impact on the entire health care system. *Health Affairs*. 2005;24(6):1554–1558.
12. Robinson JC. Managed consumerism in health care. *Health Affairs*. 2005;24(6):1478–1489.
13. US Internal Revenue Service. Health savings accounts, publication 969. http://www.irs.gov/publications/p969/ar02.html#en_US_publink100038739. Published 2009. Accessed September 27, 2009.
14. Jost TS. *Health Care at Risk: A critique of the Consumer-Driven Movement*. Durham, NC: Duke University Press; 2007.
15. White WD. Consumer directed health plans. In: Muller RM. *Encyclopedia of Health Services Research*. Thousand Oaks, CA: Sage; 2009:230–234.
16. Berk M, Moneit A. The concentration of health care expenditures, revisited. *Health Affairs*. 2001;20(2):8–18.
17. Lo Sasso A, Helmchen LA, Kaestner R. *The Effects of Consumer-Directed Health Plans on Health Care Spending*. Washington, DC: National Bureau of Economic Research; 2009. NBER Working Paper No.15106.
18. Regopoulos L, Christianson JB, Claxton G, et al. Marketwatch consumer-directed health insurance products: local-market perspectives. *Health Affairs*. 2006; 25(3):766–773.
19. Berenson RA. Which way for competition? none of the above: advocates of consumer-directed care do not speak the language of the public, no matter how often they invoke "consumer sovereignty." *Health Affairs*. 2005;24(6):1536–1542.
20. Kaiser Family Foundation. *National Survey of Enrollees in Consumer Directed Health Plans*. Washington, DC: The Henry J. Kaiser Family Foundation; 2006. Report No.7594.
21. Remler DK, Glied SA. How much more cost sharing will health savings accounts bring? *Health Affairs*. 2006;25(4):1070–1078.
22. Feldman R, Parente ST, Christianson JB. Consumer-directed health plans: new evidence on spending and utilization. *Inquiry*. 2007;44(1):26–40.
23. Kongstvedt PR. *Essentials of Managed Health Care*. Sudbury, MA: Jones and Bartlett Learning; 2007.
24. New Mexico Blue Cross/Blue Shield. National Children's Dental Health Plan at Los Alamos National Laboratory. http://www.bcbsnm.com/lanl/cdhp.htm. Accessed December 16, 2009.
25. CIGNA. CIGNA Choice Fund Experience Study. http://newsroom.cigna.com/images/56/825638_ChoiceFund_Study.pdf. Published January 2009. Accessed December 18, 2009.
26. Parente ST, Feldman R, Chen S. Effects of a consumer driven health plan on pharmaceutical spending and utilization. *Health Serv Res*. 2008;43(5):1542–1556.
27. Wharam JF, Landon BE, Galbraith AA, et al. Emergency department use and subsequent hospitalizations among members of a high-deductible health plan. *JAMA*. 2007;297(10):1093–1102.
28. Hibbard JH, Greene J, Tusler M. Does enrollment in a CDHP stimulate cost-effective utilization? *Med Care Res Rev*. 2008;65(6):764.
29. Greene J, Hibbard J, Murray JF, et al. The impact of consumer-directed health plans on prescription drug use. *Health Affairs*. 2008;27(4):1111–1119.

30. Reed M, et al. High-deductible health insurance plans: efforts to sharpen a blunt instrument. *Health Affairs*. 2009;28(4):1145.
31. Buntin MB, Damberg C, Haviland A, et al. Consumer-directed health care: early evidence about effects on cost and quality. *Health Affairs*. 2006;25(6):w516–w530.
32. Robinson JC, Ginsburg PB. Consumer-driven health care: promise and performance. *Health Affairs*. 2009;28(2):w272–w281.

This page intentionally left blank

CHAPTER 4

Health Information Technology and Public Health

Jessica S. Ancker, MPH, PhD, Lisa M. Kern, MD, MPH, Vaishali Patel, PhD, MPH, Erika Abramson, MD, and Rainu Kaushal, MD, MPH

INTRODUCTION

All public health efforts depend on massive efforts to collect, organize, interpret, and disseminate data. A complex public health task, such as coordinating a national response to epidemic influenza, for example, requires tracking cases of the disease nationwide, analyzing the resulting data rapidly, presenting the information to policy makers, and organizing and tracking interventions, such as manufacturing and distributing vaccine. Rapid and high-quality information is similarly critical to other functions performed by local, state, and federal public health agencies, including assessing population health (through surveys, vital statistics, and specific reportable disease tracking), investigating epidemics, promoting healthy behavior through social marketing and public policy, monitoring and ensuring the quality of food and drinking water, and conducting research. In fact, all three of public health's essential functions—assessing population health, developing policy, and ensuring population health by applying these policies [1–2]—can be seen as information management tasks. To accomplish these functions, public health agencies have created information systems, such as electronic disease registries, to collect and aggregate the data.

Information is similarly essential to the practice of medicine. Clinical medicine has developed a different set of electronic information systems designed to track and collect individual patient data; these systems include electronic health records and billing systems. These clinical and public health information systems have remained separate not only because of a technical inability to exchange data between them, but also because the two communities have had different information needs and stakeholders.[3] Public health relies on group-level data on millions of people, is publicly funded, and involves nonclinician stakeholders such as public officials, schools, community health workers, and epidemiologists.[4] Public health also emphasizes prevention and often includes a focus on social and environmental determinants of disease, such as social support, sexual behavior, nutrition, and physical environment.[3]

Clinicians, on the other hand, require detailed chronological information about much smaller numbers of patients. Clinicians also need this data in real time, whereas public health data collection often takes place on a relatively slow timeline. Because most of the medical care in America is provided by the private sector, clinicians need this information not only to provide optimal care, but also to bill insurers or public payers such as Medicaid.

By the end of the 20th century and the beginning of the 21st century, at a time of rapid advances in computing power, several compelling new incentives emerged to improve data about personal and public health, to better link clinical and public health data, and to reform the practice of medicine. One of these incentives was an urgent new public focus on bioterrorism and biosurveillance, stemming from such events as the terrorist attacks of September 11, 2001, followed in the same year by the anthrax attacks (in which spores were mailed to several public figures and news organizations), and in 2003, the emergence of severe acute respiratory syndrome (SARS) as a novel infectious disease that rapidly crossed state and national boundaries.[5–6] These events raised concerns about the inadequacy of the public health information infrastructure, and created a new perceived need for much more timely and thorough biosurveillance.

Another incentive pertained to out-of-control medical cost related to U.S. medical care becoming more technology intensive, the ageing of the population, and the shifting of the national disease burden from acute infectious disease to costly chronic or long-term problems such as obesity, diabetes, and cancer.[7–9] Furthermore, national experts began to call attention to serious problems within the medical system itself and their cost and quality implications, such as medical errors, fragmented medical care, and unwarranted regional variations in medical practice.[7, 10–12] Engaging the public in their own health and health care began to be seen as a more important way to help improve both medical care quality and public health.[13]

The term health information technology or health IT describes a variety of methods for electronically managing and transmitting clinical and health information,[14] from a desktop record system in a doctor's office, to a Web site in which a patient can look up her Papanicolaou (Pap) smear results, to a nationwide data collection system for disease surveillance. Health IT has been widely promoted as potential solution to most of the issues mentioned in the preceding paragraphs: improving collection and management of health information; linking health data sets; improving communication between public health agencies, clinicians, and consumers; and facilitating improvements in medical practice particularly through the provision of clinical decision support. These health IT systems have the potential for organizing data in a structured fashion so that it can be shared seamlessly; improving communication within and between medicine, public health, and consumers; and shaping human behavior (either at the point of decisions or through feedback afterward) to prevent errors, improve quality, or facilitate decisions.

If these potentials can be realized, health IT could substantially improve public health,[13] reduce the cost of care, and enhance quality of medical care.[10, 15] Creating these systems is technically challenging and costly, however, and implementing them can alter behavior, work practices, and professional relationships. A variety of factors pose barriers to development and widespread adoption. Furthermore, evaluation research on health IT systems sometimes shows they fail to produce their posited advantages, and may even produce unanticipated adverse consequences.[16–18]

This chapter will—

- introduce several current types of stand-alone health IT systems in medicine and public health, with their demonstrated advantages as well as their unintended consequences;
- describe the goal of *interoperability*, or seamless health information exchange between these systems, and its potential for improving public health and medical care; and
- explain continuing barriers to health IT development, implementation, and interoperability, as well as policy initiatives and research aimed at overcoming these barriers.

DEFINITIONS AND TYPES OF STAND-ALONE HEALTH IT SYSTEMS

Systems Designed for Public Health

Among the health IT systems used in public health are disease surveillance systems and immunization registries. The law mandates that certain communicable diseases (such as sexually transmitted diseases) be reported to state public health agencies, which in turn provide data to the Centers for Disease Control and Prevention (CDC). The current reporting system relies on a hybrid of paper-based and electronic technologies. After diagnosing a new case, a clinician or a staff member generally must complete a form manually, and mail or fax it to the appropriate agency, where the data are reentered into databases either manually or through technologies such as optical character recognition. Not surprisingly, the system is plagued by a low compliance rate, delays in reporting, and substantial underreporting.[13] Also, as each agency builds its own database, the data may remain in isolated caches at the local, state, or federal levels.[19–20] Like disease data, vital statistics (birth and death records) are collected nationwide using a system centralized in the National Center for Health Statistics, which receives data from the individual states under a cooperative system.[20]

Electronic immunization registries are designed to manage the increasing complexity of immunization. In recent years, the number of recommended early childhood immunizations has risen in this country, and immunization schedules have become increasingly complex.[21] Public health agencies are not the only stakeholders who want to monitor immunization completeness; clinicians also need to know what other clinicians have administered, and parents need this information to ensure that children can attend school. Public electronic immunization registries are designed to address these needs. For example, New York City's Web-based immunization registry allows authorized users to enter and view data about both immunizations and results of blood lead tests for an individual child.[22] A nationwide success story regarding public immunization registries occurred after Hurricane Katrina in September 2005; within days after the storm, the Louisiana Immunization Network for Kids Statewide connected to the Houston-Harris County Immunization Registry in Texas to provide immediate access to the immunization records of children who were forced to evacuate New Orleans.[23]

Systems for Clinical Medicine

Health IT systems used in clinical medicine include electronic medical or health records (EMRs and EHRs), computerized provider order entry (CPOE), e-prescribing, and clinical decision support. An EMR is an electronic record of health-related

information on an individual patient that can be created, gathered, managed, and consulted by authorized clinicians and staff. An EMR has been defined as an electronic record available to individuals within one health care organization. By contrast, electronic health records (EHRs) can be used by clinicians and staff across more than one health care organization (in other words, they are interoperable; see "Systems for Patients and Consumers" for more details).[24] Beyond this definition, however, there is considerable variation in the features offered by different EHR products. A basic EHR permits tracking of a patient's clinical and demographic data. A fully functional EHR, by contrast, also may allow clinicians to add free-text progress notes, view and manage results of laboratory tests and imaging, and use embedded functions such as computerized provider order entry and clinical decision support.

Systems for computerized provider order entry allow clinicians to place orders for lab tests, procedures, and prescriptions electronically, thus facilitating delivery to the appropriate service providers and avoiding illegible handwritten orders.[25] E-prescribing systems electronically link to pharmacies and pharmacy benefit companies, so that the prescription can be automatically delivered to the pharmacy, checked against patient-specific data at the pharmacy (such as drug allergies and other concurrent prescriptions), or checked against approved formularies.

Clinical decision support refers to a variety of techniques for helping clinicians manage the complexity of modern medicine by providing evidence-based guidance at the point of decision making in the context of patient-specific information.[26] A simple form of decision support might be a checklist that reminds a clinician of the optimal procedures to follow during the clinical encounter with a patient with a specific diagnosis. Electronic systems can provide considerably more complicated guidance, such as recommendations based on evidence-based clinical guidelines, and alerts when a patient has a documented allergy to a proposed drug or when a new prescription would lead to a drug interaction with the patient's current prescriptions.

Evidence that these types of health IT can reduce medication errors, improve adherence to clinical guidelines, and improve quality of care is fairly strong.[15] However, much of the data originates from four large academic centers using home-grown systems, so far less is known about commercial systems or about the experience in community-based practices or smaller medical centers.[15] Many other studies have been observational analyses or efficacy studies conducted under highly controlled circumstances rather than effectiveness studies in practice.[27, 28] Evidence about the overall impact of EHRs on quality of care, efficiency, and other relevant endpoints across the spectrum of health care is not yet very strong.[15, 27, 29]

Systems for Patients and Consumers

Health IT systems for patients include the personal health record (PHR) and emerging ad hoc systems of social networking. A personal health record is an electronic application that allows individuals to collect and manage their own health information. Although there have been some products that reside on a flash drive,[30] most are now Web-based. Again, as with EHRs, there is considerable variability between PHR applications.[31] One form is the so-called tethered PHR, or portal, which provides patients with a view into some of their EHR data residing with one

particular health care provider or organization. Because these tethered PHRs are fully integrated into the organization's EHR, they also may provide secure messaging so patients can communicate with the provider, refill medications, and schedule appointments. They, however, cannot integrate data from other providers outside the organization. Another form of the PHR is the freestanding PHR, a product offered by a third party (two products available are hosted by software giants Google and Microsoft™) and are not integrally connected with an individual provider's system. Although these freestanding PHRs have the capability to receive downloads of data from multiple providers, providers may or may not choose to participate.[15]

It is widely expected that PHRs will be a valuable tool to engage patients in their own health care, raise their awareness of health issues more broadly, foster a sense of ownership and empowerment, and improve patient-provider communication.[13, 31–33] Small-scale studies suggest that when patients are given access to their medical record, they provide value by correcting data, providing missing data, and documenting their adherence to health maintenance procedures.[34] Standalone electronic decision support has been shown to be helpful to patients.[35] On the other hand, minority racial status, lack of Internet access, and lower education level are known to be associated with reduced use of PHRs,[36–37] and new users may find them confusing and difficult to understand.[38] Poor health literacy, numeracy, and computer skills impair patients' ability to make use of personal health information in electronic form.[39–42] This raises the possibility that any benefits of PHRs would be disproportionately concentrated among more advantaged, Internet-savvy patient populations. Research, however, has not yet been done to document the effects of PHR access on health or quality outcomes, or to learn how patient behavior may change with access to this information.

Social networking applications are beginning to be used by consumers not only to discuss health issues, but also to share their own health information. One example is PatientsLikeMe.com, which invites patients to input personal data (such as functional status while taking a particular medication) and then allows the patients to see the pooled data.[43] The effects of this type of communication are largely unknown. It could help people become better educated about their condition, empower them to better decisions, and provide much-needed social support. On the other hand, it might promote the spread of erroneous information. Also, people without a scientific background may not recognize the potential biases and flaws of the PatientsLikeMe.com database, which is made up entirely of self-selected patients and has not been vetted for accuracy.[43]

THE GOAL OF HEALTH INFORMATION EXCHANGE AND INTEROPERABILITY

The electronic systems described here are designed for collecting and aggregating health information, but typically they do not make it simple to share the resulting information. Nevertheless, exchanging health data—and doing so in real time—is critical to both public health and clinical medicine. Biosurveillance could be made vastly more powerful if emergency, ambulatory care, and pharmacy systems could be connected to the appropriate public health agency to enable event

monitoring in near real time.[44–46] Public health would also benefit if public health agencies could better communicate alerts and results of investigations back to the clinician, ideally at the point at which they are caring for a patient for whom the alert is appropriate.[3] Monitoring over-the-counter medication sales also could provide advance warning of subclinical outbreaks.[47]

In clinical medicine, providers need rapid access to data from laboratories, pharmacies, and radiology after they order tests, prescriptions, or images. They also need to communicate with other clinicians and health care organizations because Americans change health care providers frequently [48] and see multiple providers (particularly specialists) concurrently. Transitions in care, such as discharge from the hospital to the community and emergency department visits, are especially vulnerable moments during which time key information often is lost or inaccessible.[49–50] From the patient's point of view, interoperability in the PHR is especially important because almost all patients consult multiple providers (primary care providers, specialists, dentists, oculists and ophthalmologists, allied health providers such as physical therapists, and so on) and are interested in their insurance data as well. One organization has estimated that only truly interoperable PHRs are likely to bring net value to the health care system.[31]

For all these reasons, building an infrastructure that will allow rapid and seamless electronic *health information exchange* for public health and clinical medicine has become a priority in the United States.[44, 51] The proposed National Health Information Network (NHIN) would (1) rely on fully functional EHRs and CPOE in physician office practices, nursing facilities, and hospitals; (2) connect these stakeholders to enable them to share data about individual patients; (3) allow all these stakeholders to view the results of lab and radiology tests electronically; and (4) connect providers with payers to allow automated claims submission, eligibility checks, and referral processing.[51] The CDC's Public Health Information Network (PHIN) is a national initiative to enable public health to exchange information electronically (www.cdc.gov/phin). One component is the National Electronic Disease Surveillance System (NEDSS), which is being designed to detect outbreaks rapidly by enabling electronic delivery of information from clinical systems to public health departments. One goal is to automate the process to minimize clinicians' burden in providing this information (www.cdc.gov/NEDSS). An example of a functioning public health surveillance system is the Real-time Outbreak and Disease Surveillance (RODS) system. In Pennsylvania and Utah, hospitals send RODS data in real time as they are collected at physician-patient encounters in emergency departments and pharmacy transactions.[52–53] Automated reporting of certain disease results from electronic laboratory data has been shown to be more complete and timely than the current clinician-initiated, paper-based reporting system.[54]

An important step toward making health information exchange genuinely useful is achieving *interoperability*, or the ability for different electronic systems to work together to integrate each other's data seamlessly, with as little manual involvement as possible.[31, 55] This requires internationally agreed-on *standards,* including standard terminology to ensure that words mean the same thing to all stakeholders and standard data structures to ensure that information is computer readable.[19, 44, 56] An example of a terminology standard is LOINC (Logical Observation Identifiers Names and Codes), which is a controlled set of codes to describe laboratory

tests and other clinical observations. An example of a data format standard is the Health Level Seven (HL7) messaging standard, which specifies how information is packaged and sent electronically so that individual components of the message can be recognized by the computer software and interpreted accurately. Such standards mean the difference, for example, between e-mailing a PDF (portable document format) file of the patient's hemoglobin A1c level, and sending an electronic message such that the data would appear in the appropriate field of the EHR. Sending information regarding a PDF file is an example of so-called machine-organizable data exchange; the information is sent electronically, but human operators are needed to manually reenter the data into the electronic system or read it and make a decision on the basis of it. By contrast, machine-interpretable data exchange allows the data itself to be integrated into the record and used for electronic decision support or other purposes without the need for human involvement.

A key concept in interoperability is that data should be entered once and then reused multiple times. In a noninteroperable system of health IT, a clinician might open a child's EHR to enter her rubella immunization date, and then switch to a Web browser to access the public immunization registry, locate the child's record there, and reenter the same information in the format required by the registry. By contrast, in an interoperable system, the data once entered into an EHR could be seamlessly uploaded to the public registry, perhaps automatically to avoid burdening the clinician.

BARRIERS TO DEVELOPMENT, ADOPTION, AND USE OF HEALTH IT

Despite the potential benefits of health IT, adoption has been slow in the United States.[14, 57–58] A survey within the American Hospital Association found that only 1.5 percent of acute care hospitals had an EHR in all clinical units, and an additional 7.6 percent had a system in at least one unit.[58] Computerized provider order entry had been implemented in 17 percent of hospitals; larger hospitals, urban hospitals, and teaching hospitals were more likely to have EHRs. As of 2005, about 23 percent of U.S. physicians in ambulatory practice used some form of basic EHR, but only 9 percent had an EHR with capabilities such as CPOE and decision support.[14] Electronic prescribing is relatively rare, especially among community-based practitioners.[59] The barriers to be overcome before health IT is likely to become widespread include financial, technical, social, and human factors.

Financial

One financial barrier is the large start-up and maintenance costs for health care organizations seeking to implement health IT. Moreover, although in theory the decision support and other functions in the EHR could save money by reducing waste and inefficiencies, such as duplicated lab tests, the financial benefits may accrue primarily to the insurer, not to the practitioner.[29] This problem, in which the costs are borne by one stakeholder while the benefits accrue to another, is known as *misalignment of incentives*. A similar misalignment may hamper adoption of PHRs in which financial benefits of improved care are likely to go primarily to the insurer, whereas the costs of the PHR might be borne by either the individual or the health care provider.[31]

Technical

Continued improvements in technical and organizational architecture are needed before many of the health information exchange and interoperability benefits can be realized. For example, different EHR vendors must agree to follow the same standards and, in fact, interpret the standards in the same way before they will be able to exchange data.

Social

An important societal barrier is the American concern that sharing health information may compromise personal privacy and put individuals at risk of consequences such as stigma, loss of a job, or loss of health insurance. One way in which this concern has manifested is the popular opposition to a universal unique identification code for medical purposes. Congress currently has restricted the Department of Health and Human Services from implementing a unique identifier.[4] Without a unique identifier, it is challenging to confirm that two records in two different EHRs refer to the same individual patient and thus it is difficult to share data horizontally across data sources and longitudinally over time.

Human Factors

Human factors such as poor usability and information design, lack of integration into ordinary workflow, and the potential for disrupting professional relationships have proven to be serious barriers to adoption.[18, 60] In one famous case, physicians at Cedars-Sinai Medical Center in Los Angeles forced a halt to the implementation of a costly new CPOE system because it took a prohibitively lengthy sequence of steps to place an order, interrupted their workflow, and threatened patient safety.[61] Similarly, interface usability issues posed a serious problem for new users of a PHR.[38]

POLICY DEVELOPMENT TO PROMOTE HEALTH IT

One way in which these barriers are being addressed is through national and state policy development. A national standards organization, HL7, is continuing to develop standards for health information and health information exchange. For example, a recent HL7 product is a functional model that EHR systems should strive to meet.[56, 62] This all-volunteer, nonprofit organization includes representatives from academia and industry. The resulting standards are not mandated, but they are widely respected and adopted. One method of trying to ensure some level of uniformity in standards used in different products is to promote certification. The Certification Commission for Health Information Technology (CCHIT), a Chicago-based nonprofit organization, was assigned the task of developing certification criteria by the U.S. Department of Health and Human Services and began certifying EHR systems in 2006.[56] Certification focuses heavily on standards for data exchange, as well as security and system reliability.

National policy is beginning to focus on the problem of misalignment of incentives. The 2009 American Recovery and Reinvestment Act (ARRA) included a provision that will give providers financial incentives if they can demonstrate “meaningful

use" of certified EHRs.[63] Although specifics of how "meaningful use" will be defined are still in development, this has the potential to make major changes to the health IT landscape.

State-level policy also is addressing financial and technical barriers. For example, New York State is investing heavily in health IT through a program called Healthcare Efficiency and Affordability Law for New Yorkers (HEAL NY), which has provided seed money to local organizations that are willing to raise matching funds, and has promoted independent evaluation of the impact of the resulting technologies.[64] One major focus of this program has been to promote health information exchange, so many of the beneficiaries have been regional health information organizations (RHIOs). These organizations bring together key stakeholders within a community to encourage the sharing of health information and to build technical and governance structures for health information exchange.[26] RHIOs are likely to be an integral part of the NHIN.

AREAS OF ACTIVE RESEARCH

To ensure that the best possible options are chosen for policy, careful research is needed to identify problems, test solutions, and evaluate outcomes. Thus, *biomedical informatics* research is active and growing in importance. One important area of ongoing research is improving the usability of health IT and better integrating it with workflow.[18, 60, 65] Methods may include qualitative or quantitative usability analyses: examples include verbal protocol analysis, which analyzes a single user's actions and verbalized thoughts while working with the technology; and task analyses, which breaks down tasks to be completed by various stakeholders with the help of the health IT and assesses the results for cognitive difficulty or other problems. These methods are helpful for matching technology to individual cognitive needs as well as group dynamics.[66–67] Focus groups, semistructured interviews, ethnographic analyses, and other qualitative methods are valuable for learning about workflow and how it is affected by the introduction of IT.[68] These methods are critical for reducing resistance to the introduction of new technology by professionals [60], as well as for ensuring its suitability for patients.[69–70]

Research into health literacy, communication, and behavior change is important to understand the impact of health IT on patients. Innovative ways of representing and explaining medical information [71–72] and of guiding longer-term behavior change [73] may be needed before patients will be able to make productive use of PHRs. More information is needed to determine whether and how patients will change behavior in response to access to their medical information and online communication.[43]

Computer science methods for processing information also can have strong effects on the usability of health IT. For example, clinicians and other technology users generally find free text easiest to both input and understand. Most electronic systems force structured data input through such methods as dropdown menus, checkboxes, and controlled vocabularies, because structured data are needed for interoperable data exchange, decision support, and other health IT functions. Natural language processing research uses computer science methods to find better ways of parsing free medical text to extract structured data,[74–76] and conversely of

turning structured data into readable text summaries.[77] Ultimately, these methods could make systems easier to use for humans, without compromising the computer's ability to process information. Other active areas of research using computer science methods include patient record matching and de-identifying EHR data for public health reporting and research.[78–79]

At the level of program evaluation, it is critical to learn more about the costs, effectiveness, and comparative effectiveness of health IT systems.[64, 80–83] Initial evaluation of new programs (such as health information exchange) must focus on the platform to verify that it works as planned; for example, that data are accurate and exchanged in a timely fashion. Other processes and outcomes to be measured include system usage, the financial sustainability of the business case or model, clinical and administrative impact (including, at least for some projects, impact on quality of care), unintended consequences, comprehensive return on investment, and overall success of the program.[80] To demonstrate the value of EHRs, research must focus on both economic and clinical determinants of value. Also, to date, the potential benefits of health IT have been demonstrated primarily in a few large academic centers using home-grown systems iteratively refined over many years.[15] To maximize impact on public health, the focus of evaluation research must be broadened to include the experience of the wider variety of health care organizations that now are implementing health IT systems (in most cases, commercial ones). Providers, payers, public officials, and patients cannot be expected to adopt health IT wholeheartedly without better answers to the questions of whether health IT is saving money, improving public health, and adding value to the health care system.

CONCLUSION

Health IT, particularly interoperable health information exchange, has the potential to advance public health markedly by improving the assessment of population health, collecting data to inform health policy, and implementing policy to ensure population health. Linking clinical and public health systems to create a seamless, nationwide electronic data collection system could produce a much more powerful disease surveillance system with the ability to track population health status as well as rapidly emerging biological threats. Health IT offers particularly exciting possibilities for improving the quality and efficiency of health care delivery by making essential individual-level medical data more readily accessible at the point of care; improving communication among clinicians, patients, and public health agencies; and providing evidence-based clinical decision support to help clinicians practice according to optimal care guidelines.

Major barriers, however, remain to optimizing health IT systems and implementing them. User interface issues make much health IT difficult to use, and a lack of understanding of clinical or public health workflow on the part of software developers may result in systems that interfere with ordinary work processes instead of facilitating them. Technical challenges such as the need for additional standards also remain to be overcome before many of the benefits of health IT are likely to be realized. In addition, health IT systems sometimes have unanticipated adverse consequences when implemented in the real world. As a result, there are many important

opportunities for quantitative and qualitative research to improve these systems so that they live up to their potential.

In addition, more extensive evaluation research is needed to establish a firmer evidence base about the impact of health IT on population health. The value of health IT must be assessed as a function of both clinical factors, such as measured improvements in indicators of health or quality of medical care, and economic factors, such as efficiency, reduction in waste, or return on investment. When both clinical and economic effects are maximized, it will be clear that health IT is genuinely adding value to the health care system.

REFERENCES

1. Institute of Medicine. *The Future of the Public's Health in the 21st Century*. Washington, DC: National Academy of Sciences; 2002.
2. Institute of Medicine. *The Future of Public Health*. Washington, DC: National Academy of Sciences; 1988.
3. Kukafka R, Ancker JS, Chan C, et al. Redesigning electronic health record systems to support public health. *J Biomed Inform*. 2007;40(4):398–409.
4. Yasnoff W, Overhage J, Humphreys B, LaVenture M. A national agenda for public health informatics: summarized recommendations from the 2001 AMIA Spring Congress. *J Am Med Inform Assoc*. 2001;8(6):535–545.
5. Khan AS, Ashford DA. Ready or not: preparedness for bioterrorism. *N Engl J Med*. 2001;345(4):287–289.
6. Institute of Medicine. *Secret Agents: The Menace of Emerging Infections*. Washington, DC: National Academy of Sciences; 2002.
7. Institute of Medicine. *To Err Is Human: Building A Safer Health System*. Washington, DC: National Academy of Sciences; 1999.
8. Alemayehu B, Warner KE. The lifetime distribution of health care costs. *Health Serv Res*. 2004;39:627–642.
9. Yach D, Hawkes C, Gould CL, Hofman KJ. The global burden of chronic diseases: overcoming impediments to prevention and control. *JAMA*. 2004;291(21):2616–2622.
10. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press; 2001.
11. Aspden P, Wolcott J, Bootman J, Cronenwett L, eds. for Institute of Medicine. *Preventing Medication Errors: Quality Chasm Series*. Washington, DC: National Academies Press; 2007.
12. Wennberg J, Cooper M, eds. *The Dartmouth Atlas of Health Care*. Chicago, IL: American Hospital Association Press; 1999.
13. The Markle Foundation. *Connecting for Health: A Public-Private Collaborative*. New York: The Markle Foundation; 2003.
14. Blumenthal D, Glaser JP. Information technology comes to medicine. *N Engl J Med*. 2007;356(24):2527–2534.
15. Chaudhry B, Jerome W, Shinyi W, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med*. 2006;144(10):742–752.
16. Koppel R, Metlay JP, Cohen A, et al. Role of computerized physician order entry systems in facilitating medication errors. *JAMA*. 2005;293(10):1197–1203.
17. Han YY, Carcillo JA, Venkataraman ST, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics*. 2005;116(6):1506–1512.

18. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. *J Am Med Inform Assoc*. 2004;11(2):104–112.
19. Chute CG, Koo D. Public health, data standards, and vocabulary: crucial infrastructure for reliable public health surveillance. *J Public Health Manag Pract*. 2002;8:11–17.
20. O'Carroll PW, Yasnoff WA, Ward ME, Ripp LH, Martin EL, eds. *Public Health Informatics and Information Systems*. New York: Springer-Verlag; 2003.
21. Yasnoff WA, Rippen HE. The electronic health records system in population health. In: Lehmann HP, Abbott PA, Roderer NK, et al., eds. *Aspects of Electronic Health Record Systems*. New York: London Springer; 2006.
22. Papadouka V, Schaeffer P, Metroka A, et al. Integrating the New York Citywide Immunization Registry and the Childhood Blood Lead Registry. *J Public Health Manag Pract*. 2004;10:S72–S80.
23. Boom JA, Dragsbaek AC, Nelson CS. The success of an immunization information system in the wake of Hurricane Katrina. *Pediatrics*. 2007;119(6):1213–1217.
24. National Alliance for Health Information Technology. *Report to the Office of the National Coordinator for Health Information Technology on Defining Key Health Information Technology Terms*. Bethesda, MD: National Alliance for Health Information Technology; 2008.
25. Kuperman G, Gibson R. Computer physician order entry: benefits, costs, and issues. *Ann Intern Med*. 2003;139(1):31–39.
26. Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA*. 1998;280(15):1339–1346.
27. DesRoches CM, Campbell EG, Rao SR, et al. Electronic health records in ambulatory care: a national survey of physicians. *N Engl J Med*. 2008;359(1):50–60.
28. Kawamoto K, Houlihan CA, Balas E, Lobach D. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Br Med J*. 2005;330(7494):265.
29. Wang SJ, Middleton B, Prosser LA, et al. A cost-benefit analysis of electronic medical records in primary care. *Am J Med*. 2003;114(5):397–403.
30. Wright A, Sittig DF. Encryption characteristics of two USB-based personal health record devices. *JAMIA*. 2007;14(4):397–399.
31. Kaelber DC, Shah S, Vincent A, et al. *The Value of Personal Health Records*. Charlestown, MA: Center for Information Technology Leadership, Healthcare Information and Management System Society; 2008.
32. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. *J Am Med Inform Assoc*. 2006;13(2):121–126.
33. Public Health Data Standards Consortium. *Electronic Health Record: Public Health Perspectives*. Baltimore, MD: Public Health Data Standards Consortium; 2004.
34. Staroselsky M, Volk LA, Tsurikova R, et al. Improving electronic health record (EHR) accuracy and increasing compliance with health maintenance clinical guidelines through patient access and input. *Int J Med Inform*. 2006;75(10–11):693–700.
35. Green MJ, Peterson SK, Baker MW, et al. Effect of a computer-based decision aid on knowledge, perceptions, and intentions about genetic testing for breast cancer susceptibility: a randomized controlled trial. *JAMA*. 2004;292:442–452.
36. Miller H, Vandenbosch B, Ivanov D, Black P. Determinants of personal health record use. *J Healthc Inf Manag*. 2007;21(3):44–48.

37. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. *JAMIA*. 2009;16(5):683–689.
38. Britto MT, Jimison HB, Munafo JK, Wissman J, Rogers ML, Hersh WR. Usability testing finds problems for novice users of pediatric portals. *JAMIA*. 2009;16(5):660–669.
39. McCray AT. Promoting health literacy. *JAMIA*. 2005;12:152–163.
40. Ancker JS, Kaufman DR. Rethinking health numeracy: a multidisciplinary literature review. *J Am Med Inform Assoc*. 2007;14(6):713–721.
41. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: a systematic review. *J Am Med Inform Assoc*. 2006;13(6):608–618.
42. Poynton T. Computer literacy across the lifespan: a review with implications for educators. *Comput Human Behav*. 2005;21:861–872.
43. Ancker JS, Carpenter K, Greene P, et al. Peer-to-peer communication, cancer prevention, and the internet. *J Health Commun*. 2009;14:38–46.
44. Biosurveillance Technical Committee. *Standards Gap and Overlap Analysis Biosurveillance Use Case: Visit, Utilization, and Lab Results Data*. Washington, DC: Healthcare Information Technology Standards Panel; 2006. Report No.Contract HHSP23320054103EC.
45. Biosurveillance Technical Committee. *Selected Standards*. Washington, DC: Healthcare Information Technology Standards Panel; 2006.
46. Lazarus R, Klompas M, Campion FX, et al. Electronic support for public health: validated case finding and reporting for notifiable diseases using electronic medical data. *J Am Med Inform Assoc*. 2009;16(1):18–24.
47. Marx M, Rodriguez C, Greenko J, et al. Diarrheal illness detected through syndromic surveillance after a massive power outage: New York City, August 2003. *Am J Public Health*. 2006;96(3):547–553.
48. Reed MC. Why people change their health care providers. *Data Bulletin: Center for Studying Health System Change*. 2000;16:1–2. http://www.hschange.com/CONTENT/81/81.pdf.
49. Smith PC, Araya-Guerra R, Bublitz C, et al. Missing clinical information during primary care visits. *JAMA*. 2005;293(5):565–571.
50. Forster AJ, Murff HJ, Peterson JF, Gandhi TK, Bates DW. The incidence and severity of adverse events affecting patients after discharge from the hospital. *Ann Intern Med*. 2003;138(3):161–167.
51. Kaushal R, Bates DW, Poon EG, Jha AK, Blumenthal D, and the Harvard Inter-Faculty Program for Health Systems Improvement NHIN Working Group. Functional gaps in attaining a National Health Information Network. *Health Affairs*. 2005;24(5):1281–1289.
52. Tsui F-C, Espino JU, Dato VM, Gesteland PH, Hutman J, Wagner MM. Technical description of RODS: a real-time public health surveillance system. *J Am Med Inform Assoc*.2003;10(5):399–408.
53. Gesteland PH, Gardner RM, Tsui F-C, et al. Automated syndromic surveillance for the 2002 Winter Olympics. *J Am Med Inform Assoc*. 2003;10(6):547–554.
54. Overhage JM, Grannis S, McDonald CJ. A comparison of the completeness and timeliness of automated electronic laboratory reporting and spontaneous reporting of notifiable conditions. *Am J Public Health*. 2008;98(2):344–350.
55. Walker J, Pan E, Johnston D, Adler-Milstein J, Bates DW, Middleton B. The value of health care information exchange and interoperability. *Health Aff (Millwood)*. 2005;Suppl Web Exclusives:W510–W518.
56. Mon D. The difference between the EHR standard and certification. *J AHIMA*. 2006; 77(5):66–70.

57. Ford EW, Menachemi N, Phillips MT. Predicting the adoption of electronic health records by physicians: when will health care be paperless? *J Am Med Inform Assoc*. 2006;13:106–112.
58. Jha AK, DesRoches CM, Campbell EG, et al. Use of electronic health records in US hospitals. *N Engl J Med*. 2009;360(16):1628–1638.
59. Pagan J, Pratt W, Sun J. Which physicians have access to electronic prescribing and which ones end up using it? *Health Policy*. 2009;89(3):288–294.
60. Armijo D, McDonnell C, Werner K. *Electronic Health Record Usability: Interface Design Considerations*. Rockville, MD; 2009. Report No.09(10)-0091-2-EF.
61. Morrissey J. Harmonic divergence: Cedars-Sinai joins others in holding off on CPOE. *Mod Healthc*. 2004;24(8).
62. Health Level Seven Electronic Health Record Technical Committee. Home page. http://www.hl7.org/ehr/. Published 2006. Accessed December 5, 2006.
63. Blumenthal D. Stimulating the adoption of health information technology. *N Engl J Med*. 2009;360(15):1477–1479.
64. Kern LM, Kaushal R. Health information technology and health information exchange in New York State: new initiatives in implementation and evaluation. *J Biomed Inform*. 2007;40(6 Suppl):S17–S20.
65. Poon EG, Blumenthal D, Jaggi T, Honour MM, Bates DW, Kaushal R. Overcoming barriers to adopting and implementing computerized physician order entry systems in US hospitals. *Health Aff (Millwood)*. 2004;23(4):184–190.
66. Patel VL, Kaufman DR. Cognitive science and biomedical informatics. In: Cimino JJ, Shortliffe EH, eds. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. 3rd ed. New York: Springer; 2006.
67. Seol YH, Kaufman DR, Mendonca EA, Cimino JJ, Johnson SB. Scenario-based assessment of physicians' information needs. *Medinfo*. 2004;11(1):306–310.
68. Ventres W, Kooienga S, Vuckovic N, Marlin R, Nygren P, Stewart V. Physicians, patients, and the electronic health record: an ethnographic analysis. *Ann Family Med*. 2006;4(2):124–131.
69. Kaufman DR, Patel VL, Hilliman C, et al. Usability in the real world: assessing medical information technologies in patients' homes. *J Biomed Inform*. 2003;36:45–60.
70. Kukafka R, Khan SA, Hutchinson C, et al. Digital partnerships for health: steps to develop a community-specific health portal aimed at promoting health and well-being; *Proceedings AMIA Annual Fall Symposium*. 2007:428–432.
71. Kaufman DR, Pevzner J, Hilliman C, et al. Redesigning a telehealth diabetes management program for a digital divide seniors population. *Home Health Care Manag Pract*. 2006;18(3):223–234.
72. Ancker JS, Chan C, Kukafka R. Interactive graphics to demonstrate health risks: formative developed and qualitative evaluation. *J Health Commun*. 2009;14(5):461–475.
73. Kukafka R, Lussier YA, Eng P, Patel VL, Cimino JJ. Web-based tailoring and its effect on self-efficacy: results from the MI-HEART randomized controlled trial. *Proceedings of the AMIA Symposium*. 2002:410–414.
74. Hripcsak G, Elhadad N, Chen Y, Zhou L, Morrison F. Using empiric semantic correlation to interpret temporal assertions in clinical texts. *J Am Med Inform Assoc*. 2009;16:220–227.
75. Chen E, Stetson P, YA. L, Markatou M, Hripcsak G, Friedman C. Detection of practice pattern trends through Natural Language Processing of clinical narratives and biomedical literature. *Proceedings AMIA Annual Fall Symposium*. 2007:120–124.

76. Chen E, Hripcsak G, Friedman C. Disseminating natural language processed clinical narratives. *Proceedings AMIA Annual Fall Symposium*. 2006:126–130.
77. Elhadad N, McKeown K, Kaufman DR, Jordan D. Facilitating physicians' access to information via tailored text summarization. *Proceedings AMIA Annual Fall Symposium*. 2005:226–230.
78. Kukafka R, Ancker JS, Chan C, et al. Redesigning electronic health record systems to support public health. *J Biomed Inform*. 2007;40(4):398–409.
79. Beckwith B, Mahaadevan R, Balis U, Kuo F. Development and evaluation of an open source software tool for deidentification of pathology reports. *BMC Med Inform Decis Mak*. 2006;6(12).
80. Hripcsak G, Kaushal R, Johnson KB, et al. The United Hospital Fund meeting on evaluating health information exchange. *J Biomed Inform*. 2007;40:S3–S10.
81. Kern LM, Barron Y, Blair AJ III, et al. Electronic result viewing and quality of care in small group practices. *J Gen Intern Med*. 2008;23(4):405–410.
82. Kern LM, Barron Y, Abramson EL, Patel V, Kaushal R. HEAL NY: Promoting interoperable health information technology in New York State. *Health Aff (Millwood)*. 2009;28(2):493–504.
83. Kern LM, Dhopeshwarkar RV, Barron Y, Wilcox A, Pincus H, Kaushal R. Measuring the effects of health information technology on quality of care: a novel set of proposed metrics for electronic quality reporting safe. *Jt Comm J Qual Patient Saf*. 2009;35(7):359–369.

This page intentionally left blank

SECTION 2

HEALTH CARE DISPARITIES

This page intentionally left blank

CHAPTER 5

Racial and Ethnic Diversity in Academic Medicine

Carla Boutin-Foster, MD, MS

The 21st century has been marked by a significant growth in the proportion of racial and ethnic minorities in the United States relative to the total population. Currently, approximately 30 percent of the U.S. population is classified as being a racial and ethnic minority, with blacks and Hispanics accounting for approximately 12 percent each, and Asians, Pacific Islanders, and Native Americans representing approximately 5 percent. Based on vital statistics (births, deaths) and immigration statistics, it is projected that the demographic composition of the U.S. population will undergo a seismic change over the next decades. The most dramatic change will be observed among the composition of the number and proportion of Hispanic/Latinos in the population. That is, by the year 2050, the proportion of individuals of Hispanic/Latino origin is expected to double from 12.5 percent to 24.4 percent of the total population; the black population is projected to rise from 12 percent to 14.6 percent; and the proportion of people of Asian origin is expected to double from 4 percent to approximately 8 percent.[1] The changing demographics will necessarily reflect a different, multicultural population characterized by a wide diversity in languages, cultural practices, and beliefs. And with this change, there most certainly will be new health challenges.

Studies have shown that although immigrant populations tend to be healthier than their native-born counterparts, over time, as the immigrant population adopts the health habits of the existing population, their health worsens.[2] Therefore, as more immigrant populations come to the United States and the longer they live in the United States, the greater the likelihood that will adopt health behaviors characteristic of their adopted country—that is, unhealthy eating and physical inactivity, which lead to diabetes, heart disease, asthma, and other chronic conditions.[3] Furthermore, second-generation immigrants also tend to quickly become "Americanized," which often leads to behavioral and lifestyle health conditions contributing to chronic disease development. As such, public health organizations and academic medical centers, in particular, will need to shift their focus to meet the challenges of this growing diverse minority population. One way to better meet the needs of this

population is to develop a culturally diverse environment, including a new focus on research of the health care needs of minority populations and the establishment of culturally sensitive projects that are designed specifically to address the health problems of this population. Importantly, academic medical centers as well as community clinics must begin to educate and train physicians and health care providers who will care for this culturally diverse population.

The focus of this chapter is on the public health implications of a growing racial and ethnically diverse America and the role academic medical centers can and should play in providing care to this multicultural population. An argument will be made for the need to bring the issues of cultural diversity to the forefront of medical education. Moreover, a case will be made for advocating for an increase in the proportion of racial and ethnic minorities in the physician workforce. As used herein, racial group will be defined as a group of persons defined by reference to race, color (including citizenship), or ethnic or national origin. For purposes in this chapter, ethnicity refers to the relating to, or characteristic of a sizable group of, people sharing a common and distinctive racial, national, religious, linguistic, or cultural heritage.

In the United States, academic health centers have a long tradition of caring for the public, particularly the less fortunate who tend to seek care at medical center clinics. To continue to provide both clinically excellent and culturally sensitive care, academic health centers need to address the issue of race and ethnicity as well as diversity of their workforce. I argue that the workforce should mirror the composition of the patient population. To achieve this objective, there will have to be an increase in the proportion of medical students, faculty, and staff from diverse backgrounds that can feed directly into the workforce to reflect the demographic shift. History shows, however, that achieving this goal is not as simple as it seems.

HISTORICAL PERSPECTIVES ON ACHIEVING DIVERSITY IN THE PHYSICIAN WORKFORCE

By the middle of the 19th century, approximately 14 medical schools were developed specifically to train black physicians. Shortly after the Civil War, however, approximately one-half were defunct, leaving seven medical schools that admitted only black students. These schools were almost exclusively located in the South and included Howard University Medical School located in Washington, D.C.; Leonard Medical School (Shaw University) in Raleigh, North Carolina; New Orleans University Medical College in Louisiana; and, in Tennessee, Meharry Medical College in Nashville, Chattanooga National Medical College, Knoxville College Medical Department, and University of West Tennessee College of Physicians and Surgeons in Memphis.[4]

A further decrease in the number of black medical schools occurred after the release of the report by Abraham Flexner, a research scholar at the Carnegie Foundation for the Advancement of Teaching. In his 1910 report, "Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching," Flexner conducted an assessment of medical education in the United States and Canada. This report has been hailed as being the impetus for overhauling medical education in the United States and is credited with creating

more rigorous scientific standards for medical schools, which were designed to improve the quality of American medical education. Based on his evaluation of the medical schools, Flexner recommended that several schools be closed. After the release of this report, only two of the seven black medical schools remained: Howard University in Washington, D.C., and Meharry University in Tennessee. Flexner argued that the other five medical schools for black students should be closed because they were "ineffectual" and in no position to make any contribution of value.[4] The Flexner report also led to the closure of the only three medical schools for women. The rationale was that women showed a decreasing inclination to enter the medical profession and that there apparently was not a strong demand for women physicians at the time.[5] In one sense, the sweeping closure of these medical schools had a substantial impact on the medical education for black students and for women. At the time, there were few medical schools that even admitted blacks and women, thus closing those schools that did accept blacks and women was particularly restrictive in training black and female physicians.

After recommending these closures, Flexner made efforts to support these schools through financial support from the Carnegie Foundation. Nonetheless, the impact of these closures cannot be minimized.[6] In the wake of these closures, with few exceptions, the majority of the overwhelmingly white medical schools were not eager to admit blacks or women. In fact, by the early 1960s, less than 3 percent of the students entering medical school were black. This situation led to the impetus for development of affirmative action policies, which were designed to redress policies of institutionalized bias well entrenched in institutions of higher education, not just in medical schools.

In 1961, President Kennedy introduced the term "affirmative action" in Executive Order 10925, an order that prohibited discrimination on the basis of race, creed, color, or ethnicity. Affirmative action was a legal attempt to correct legally sanctioned racial and gender discrimination.[7] This order mandated that projects financed through federal funding "take affirmative action" to ensure that hiring practices were free of racial bias. The assassination of President Kennedy could have torpedoed the civil rights and affirmative action movement, but newly sworn in President Johnson strongly supported Kennedy's civil rights bill introduced in Congress in 1963. This bill included provisions to ban discrimination in public accommodations, and enabled the U.S. attorney general to join in lawsuits against state governments that operated segregated school systems, among other provisions. Johnson pushed for the passage of the bill and, in 1964, the Civil Rights Act of 1964 was enacted. This act was a landmark piece of legislation in U.S. history essentially outlawing racial segregation in schools, public places, and employment.

In 1965, President Lyndon Johnson mandated via Executive Order 11246, the expansion of affirmative action to federal contractors. Under this order, the Office of Federal Contract Compliance Programs stated that any agency with 50 or more employees and that received $50,000 or more in federal funds must have a written affirmative action plan for achieving a proportion of women and racial minorities that is proportionate to their availability in the general labor pool of women and members of racial minorities.[8, 9]

These affirmative action policies did not go unchallenged and resulted in debates that spanned more than three decades. One of the earliest cases that involved a

medical school was the 1978 case of *Bakke v. the University of California-Davis School of Medicine*. In this landmark case, Bakke, a Caucasian medical school applicant, charged that University of California's policy of reserving a portion of their seats in the incoming freshman medical class for minority students was in violation of Title VI of the Civil Rights Act and the Equal Protection Clause of the Fourteenth Amendment. In the landmark ruling that followed this case, Supreme Court Justice Lewis Powell opined that race could be included as one of many factors to achieve diversity, which, he concluded, was in a compelling governmental interest.[10] In contrast to the Bakke case, in *Hopwood v. the University of Texas*, the Fifth U.S. Circuit Court of Appeals ruled in 1996 against the University of Texas School of Law's use of race as a factor in deciding which applicants to admit or as a basis for admission or financial aid because it was said to be a race-based policy.[11] *Hopwood v. Texas* was the first successful legal challenge to a university's affirmative action policy in student admissions since the Bakke case. After seven years as a precedent in the Fifth Circuit, the Hopwood decision was abrogated by the U.S. Supreme Court in 2003. In *Grutter v. Bollinger*, the Supreme Court ruled in favor of the University of Michigan's Law School stating that the narrowly tailored use of race in admissions process to further a compelling interest in obtaining a diverse student body was not a violation of the Equal Protection Clause of the Fourteenth Amendment or Title VI of the Civil Rights Act. In *Gratz v. Bollinger*, however, the Court ruled against the undergraduate school stating that the use of a point system that treated groups of applicants differently based on their race was in violation the Civil Rights Act and the Fourteenth Amendment.[8]

These legal attacks against affirmative action have had a rippling effect, leading to states such as California, Connecticut, Michigan, and Nebraska to reconsider and even ban affirmative action policies. In 1996, Proposition 209, a ballot initiative for the elimination of race-conscious admissions at the state's public institutions was approved in California. In the year following that initiative, the number of minority California residents accepted to one or more California medical school fell from 233 in 1993 to 157 in 1997 and subsequently to 156 in 2001. These trends have been attributed at least in part to legal efforts dismantling affirmative action.[12]

The legal challenges to affirmative action probably have contributed to the current low enrollment of racial and ethnic minorities in medical schools and subsequently among medical school faculty. Notwithstanding, within the medical community, the issue of affirmative action has been discussed and debated. In particular, the Association of American Medical Colleges (AAMC), the umbrella organization for medical schools, has developed clearly defined policies related to affirmative action and cultural diversity. In particular, the AAMC advocates for preparing physicians and scientists to meet the nation's evolving health needs, and, more pertinent to this discussion, working toward ensuring that the nation's medical students, biomedical graduate students, residents, fellows, faculty, and the health care workforce are diverse and culturally competent. The AAMC, created to advance medical education in the United States, represents all 131 accredited medical schools in the United States and 17 in Candada; approximately 400 major teaching hospitals and health systems, including 68 Department of Veterans Affairs medical centers; and nearly 90 academic and scientific societies. Through these institutions and organizations, the AAMC represents 125,000 faculty members, 75,000 medical students, and 106,000 resident physicians.

The AAMC originally used the term underrepresented minority (URM) to recognize the under-representation of certain ethnic groups. The term URM has been most often used to refer to blacks, Hispanics/Mexican Americans, and Native Americans, including American Indians, Alaska Natives, and Native Hawaiians. The new AAMC definition, however, allows for the designation of URM to be modified based on changing demographics to refer to those racial and ethnic populations that are underrepresented in the medical profession *relative to* their numbers in the general population.[12] Based on AAMC statistics, from 1970 to 1990, URM enrollment in U.S. medical schools increased from approximately 5 percent to 11 percent and has remained steady at approximately 12 percent.[13] During this same period, however, minority representation in the U.S. population increased to 26 percent of the total population. Clearly, enrollment of URM students in medical schools has not kept pace with that of the population, which is a cause for concern.[12]

The need for improving diversity in medical schools is best illustrated by looking at statistics compiled by the AAMC.[14] Table 5.1 and table 5.2 show the distribution of medical school applicants by race and ethnicity and the proportion of U.S. medical school matriculants by race and ethnicity. In 2007, the overwhelming majority of medical school applicants were white (57 percent) and one-fifth were Asian, whereas only 7.4 percent were black, 7.1 percent were Hispanic/Latino; and Native Americans, Alaskan Natives, Native Hawaiians, and Pacific Islanders accounted for less than 1 percent of the total applicants.

Similarly, among actual medical school matriculants, blacks accounted for approximately 6.7 percent, Hispanic/Latinos 7.4 percent, Native Americans and Alaskan Natives 0.4 percent, and Native Hawaiians and Pacific Islanders 0.3 percent, whereas whites and Asians represented 60.8 percent and 19.9 percent of the total, respectively. The underrepresentation of racial and ethnic minorities in medical school enrollees and matriculants contributes to the paucity of racial and ethnic minority faculty. Racial and ethnic minorities make up only 5 percent of the total U.S. medical school faculty, and, of this, approximately 20 percent of the minority faculty is concentrated in only six medical schools. Moreover, racial and ethnic minorities are less likely to be promoted to senior academic rank and are one-third less likely to attain senior rank when compared with white faculty.[15] Thus, in addition to increasing enrollment of racial and ethnic minority medical students, increased diversity and equity is needed in medical school academic faculty and leadership.

THE CASE FOR DIVERSITY

Increasing the racial and ethnic participation in the health care workforce, especially in academic medical centers where so many poor and minority groups seek care, is important from a medical and public health perspective. The 2004 Institute of Medicine (IOM) report, "In the Nation's Compelling Interest: Ensuring Diversity in the Health Care Workforce," provides several compelling reasons for achieving diversity in the health care workforce. For example, the report states that it is essential to have diverse faculty that reflect the multicultural composition of the population. It also is important to have a diverse workforce to teach students and trainees, as well as advocates for more research to address the pressing health care issues regarding providing high quality care to all, regardless of race, color, and creed.[16] Indeed,

Table 5.1
Percentage and Number of U.S. Medical School Applicants by Race and Ethnicity, 2007

Race/Ethnicity	Percent (Number)
White	57.0 (24,136)
Asian	19.8 (8,390)
African American	7.4 (3,133)
Hispanic or Latino	7.1 (2,999)
Non-U.S. Resident or Permanent Resident (Foreign)	4.3 (1,810)
Other	4.4 (1,696)

Source: Adapted from Association of American Medical Colleges data.

one major compelling reason for diversity is the elimination of racial and ethnic disparities.

The report also states that racial and ethnic minority health care providers are more likely to serve in medically underserved communities, thereby, increasing access to care for this needy population. Racial and ethnic minority patients report greater levels of satisfaction with care provided by minority health professionals perhaps because of a perceived comfort level of being treated by a race-concordant provider. Certainly, cultural and linguistic barriers can be breached more easily if the provider can speak the same language as the patient. Evidence seems to support this assumption. A 2003 study of 252 adults in a primary care setting found that

Table 5.2
Percentage of U.S. Medical School Matriculants by Race and Ethnicity, 2006–2007

Race/Ethnicity	Percent Matriculants
Whites	
In 2006:	60.8
In 2007:	59.9
Asian	
In 2006:	18.7
In 2007:	19.9
Hispanic/Latino	
In 2006:	7.4
In 2007:	7.2
African American	
In 2006:	6.7
In 2007:	6.4
Non-U.S. or Permanent Resident (Foreign)	
In 2006:	1.6
In 2007:	1.8
Other	
In 2006:	4.8
In 2007:	4.8

Source: Adapted from Association of American Medical Colleges data.

race-concordant visits were longer, these visits were associated with greater patient satisfaction, and the physicians were more participatory.[17] Concordance may reflect a greater interest in personal beliefs, values, and the patient's community.[18]

Data from the Commonwealth Fund Minority Health Survey also found that among each race and ethnic group, respondents who were race concordant reported greater satisfaction with their physician compared with respondents who were not race concordant.[19] In another study of HIV/AIDS patients, findings showed that having race-concordant providers eliminated the disparity in initiation of protease inhibitors.[20] This is not say that patients and physicians must be racially concordant to achieve such results, rather the issue speaks to physician and patient cultural sensitivity as well as physician willingness to discuss personal and social health concerns with patients with whom they can better relate because of a similar ethnic or racial background. Clearly, efforts should be made to understand factors that contribute to a sense of greater patient satisfaction, compliance, and understanding in order to promote better understanding among all physicians, regardless of racial or ethnic background.

Dr. Louis W. Sullivan, former U.S. Secretary of Health and Human Services, established the Sullivan Alliance to increase diversity in the health professions to reduce racial and ethnic health disparities. The three main objectives of the Sullivan Alliance are to (1) raise awareness of the importance and value of achieving racial and ethnic diversity in the health professions, (2) disseminate information about "best practices" and resources that enhance diversity, and (3) stimulate academic programs in the health professions of medicine, dentistry, nursing, psychology, and public health to create new or more effectively implement existing diversity initiatives (see http://www.jointcenter.org/hpi/pages/sullivan-alliance). In 2004, the Sullivan Alliance issued its final report, "Missing Persons: Minorities in the Professions," which described increasing the diversity of medical students, residents, and faculty as an "indispensable tool in efforts to improve access to health care for underserved populations."[2, 21] This report provided the public with the foundation for achieving cultural diversity and competency among health care professions. The following sections focus on the rationale for increasing diversity in the physician workforce.

HEALTH DISPARITIES DO MAKE A DIFFERENCE

Fundamental to addressing many of the current ills in the health system is addressing the role that race and ethnicity play in health care access and outcomes. Despite significant advances in civil rights, the fact remains that when it comes to both the type and quality of health care that an individual receives, race and ethnicity must be taken into account. According to the IOM report, "Unequal Treatment: Confronting Racial and Ethnic Disparities," race and ethnicity remain a significant factor in determining whether an individual receives high-quality care and in determining health outcomes.[22] Race has been shown to be a determinant of the characteristics and qualifications of the physician that patients see, the types of hospital to which a patient is admitted, and the types of procedures they will undergo.[23–26] Upon inspection, it becomes evident that race and ethnicity are important factors that are at the core of the health care debate. Of the 47 million uninsured American, more than 50 percent are racial and ethnic minorities.

Epidemiologic studies clearly show persistent differences in health outcomes and health status among racial and ethnic groups. For example, high blood pressure is more prevalent and is associated with poorer outcomes in blacks than whites. As a consequence, blacks are nearly twice as likely to have a first stroke and more likely to die from one compared with whites.[27–29] Among adults older than age 20, racial and ethnic minorities have a higher prevalence of diabetes than whites. Racial and ethnic minorities also have poorer glycemic control. Although this may be attributed to physiologic differences, differences in treatment of patients with diabetes also have been implicated.[30, 31]

Similar disparities are observed in cancer. Although cancer death rates have declined, the rate of decline is lower in blacks. Black men have the highest incidence for prostate cancer and are more than twice as likely to die from this disease.[32] White women have the highest incidence rate for breast cancer, but black women are most likely to die from breast cancer, usually at a more advanced stage, which most likely is due to a lack of early detection or suboptimal treatment.[33] Among all racial and ethnic groups, black males and black females have a higher incidence and death rate from both colorectal and lung cancer.[34, 35] Both blacks and Latinos are more likely to experience a mental disorder than their white counterparts and are less likely to seek treatment. When they do seek treatment, they are more likely to use the emergency room for mental health care and are more likely than whites to receive inpatient care rather than outpatient care.[34] With regard to mental health, women are two times more likely to suffer from depressive symptoms than men, especially Hispanic women.[36]

Not only are outcomes different among racial and ethnic groups, but treatment, too, differs. Studies have shown disparities in pain management, with minorities less likely to have access to appropriate pain management services and treatments, less likely to have their pain documented by health care providers, and less likely to receive pain medications. URMs are more likely to use the emergency department for pain care, but they are less likely to receive adequate care. They also experience poorer health and quality of life related to pain. In one study, only 25 percent of pharmacies in nonwhite neighborhoods had opioid supplies compared with 72 percent of pharmacies in white neighborhoods.[37]

THE IMPORTANCE OF HEALTH LITERACY

Health care diversity does not exclusively refer to only black and white. Diversity also applies to language. Health literacy is a major contributor to health disparity and can contribute to and exacerbate health disparities. Health literacy refers to the ability to read and comprehend health-related materials that ensure successful functioning in the role of the patient.[38] This includes the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions.[39] Although the reasons for health literacy are many (for example, language barriers and poor reading skills being major contributors), the consequences are quite serious. Poor health literacy has been linked to poorer adherence to medical instructions, poorer functional status, and poorer medication adherence.[40] Patients who have low health literacy also are those patients who are more likely to have poorer health outcomes.

Because language barriers often accentuate health literacy gaps, it is so important to have bilingual personnel provide instructions and to have culturally relevant education material written at a fifth-grade reading level in different languages. Having a physician or other health care provider who can speak the patient's language can make a tremendous difference. For example, a study that looked at Russian patients with type 2 diabetes found that the introduction of a Russian-speaking provider was associated with improvements in low-density lipoprotein (LDL), mean hemoglobin A1c, and both systolic and diastolic blood pressure, and improvements in medication adherence. Interpreters are important; however, they may limit the types of information patients provide.[41, 42] Thus, having physicians who speak the same language as their patients would improve communication and, hopefully, health outcomes.

Many studies have shown that language concordance is associated with greater patient comprehension and adherence to instructions. One study, for example, found that among Asian patients, those who had language-concordant physicians rather than interpreters were less likely to have unanswered questions. Conversely, patients who used interpreters were more likely than language-concordant patients to report having questions about their care or mental health that they wanted to ask but did not. Thus, although interpreter services are helpful, patients may be less likely to ask certain questions.[41]

PROVISION OF HEALTH CARE

Diversity in the health care workforce also is important for providing medical services to racial and ethnic minority populations. Although black and Latino physicians make up approximately 7 percent of the national physician pool, they provide care for up to 25 percent of black and Latino patients. Bach et al., for example, found that black patients were more likely to be seen by black physicians, physicians who were less likely to be board certified and physicians who reported greater difficulty in securing high-quality care.[24] Other studies have found that black and Latino physicians are more likely to practice in communities that are densely populated by black and Latino residents and are more likely to care for patients who are insured by Medicaid or who are uninsured. Komaromy et al. found that black physicians practiced in areas where there were five times as many black residents than in areas where other physicians practiced. Similarly, in communities where Latino physicians practiced, the proportion of Hispanic residents was twice as high as in areas where other physicians practiced.[43] This also may be attributable to patient preference; black and Latino patients may be more likely to seek out racial and ethnically similar physicians.[44]

As the proportion of blacks and Latinos in the population increases, there undoubtedly will be more black and Latino patients. From a manpower perspective, the number of minority physicians probably will be insufficient to meet this demand. This raises the question asked by Saha: "Can the current physician workforce pool keep up with this demand?"[44] More important, how will the policies and practices change to meet this need? Increasing diversity in the medical workforce potentially can enhance and expand current health disparities research.

The root causes of health disparities are complex and require a multidisciplinary approach to developing interventions. First, medical practitioners must have a

better understanding of the social determinants of health disparities—that is, lower socioeconomic status, less than high school education, insufficient access to healthy foods, and poor health literacy.[45] Social determinants of disease often lead to a myriad of health disparities that disproportionally affect racial and ethnic minorities. Second, more research in this area is needed to better understand the complex components to effect change. For starters, diverse researchers are needed who can interact with and engage community members and understand the social and political climate in which they live.

There are many challenges to conducting health disparities research, including difficulty in recruiting diverse research subjects, lack of conducive settings for interventions, insufficient resources that reduce the likelihood of sustainability, and lack of trust of research and academic institutions from the prospective participants.[46, 47] Addressing health disparities also relies on community involvement and developing grassroots efforts. Community-Based Participatory Research is a methodological approach that involves equal partnership between communities and academic researchers. This research method has been used to address many medical conditions such as diabetes, asthma, and cancer.[48, 49] As part of this, bilingual researchers are needed who can increase the representation of participants who are not fluent in English. This increased representation will enhance the generalizability of research findings and extend applicability to a broader target audience.[50, 51]

MAKING MEDICAL EDUCATION MORE DIVERSE

Having a diverse faculty also might help to recruit a diverse student body and house staff. Diversity in higher education and allied health professions training settings is associated with better education outcomes among *all* students.[52, 53] One study reported a significant association between the proportion of URMs in a class and student ratings of their ability to care for a diverse population. The higher the proportion of URMs, the greater the likelihood of students rating themselves as highly prepared to care for minority populations. In schools with the highest proportion of URMs, white students also were more likely to have strong attitudes toward endorsing equitable access to care. URM students were substantially more likely than white or nonwhite and non-URM students to plan to serve the underserved.[53, 54]

A diverse faculty whose members are able to participate in the design and development of curricula programs is needed, for example, in developing a curriculum on cultural competency. The AAMC and Liaison Committee on Medical Education (LCME) have set recommendations to ensure that medical schools integrate cultural competence training as part of their curriculum. Cultural competence training involves training students to interact and communicate effectively in cross-cultural settings. A more diverse faculty can demonstrate cultural competence in practice and in theory. Having a diverse faculty creates an environment of mutual learning in which students and health professionals can learn from each other. A study of minority faculty in other higher education settings demonstrated that faculty of color employ a broader range of pedagogical techniques compared with their white counterparts, including class discussions and writing assignments that include diverse perspectives.[54] Therefore, more racially and ethnically diverse faculty may enhance

the overall quality of medical education. Such diversity most certainly would reflect the demographic realities of the population.

CHALLENGES FOR THE FUTURE

The 21st century has witnessed new threats to health care that affect the most vulnerable and disenfranchised populations. For example, Hurricane Katrina forced thousands of people to relocate, the majority of whom were poor African Americans and the elderly poor. The threats of new pandemics such as H1N1 virus or biological threats such as anthrax have provided the impetus for public health officials to develop and have in place strategies for the rapid vaccination of not only those who have access to medical care, but especially those most vulnerable populations, including the chronically ill, the poor, and children and pregnant women. There is a substantial challenge for health care facilities to quickly mobilize its workforce to identify and treat those in greatest need. For this to happen expediently, medical professionals need to have a familiarity with the population in need and an understanding of their social context. Having a diverse workforce that can reach out to this population and that has the necessary social, linguistic, and cultural skills to gain trust and entry into these communities is imperative to tackling such threats.

Ideally, a diverse workforce should be present at all levels of academic medicine, especially in positions of academic leadership where policy reform is developed. Institutions must evaluate their climate for diversity through periodic confidential surveys of all students and health professional staff and ensure that diversity in numbers is accompanied by diversity in the approach to education, patient care, and research. The task for promoting diversity cannot be relegated to just a few but must be an institutional goal that is promoted from the academic medical center's leadership, that is, from the top down. This leadership must be ready to change policies that are overtly discriminatory in nature and to review the current recruitment, promotion, and tenure practices of the institution as it relates to racial ethnic minorities and women. Individuals who are familiar enough with the organizational structure and history of diversity at the institution should lead designated central offices. Institutions should review their current affirmative action policies to ensure that they reflect national standards.

To ensure diversity in academic leadership, search committees should include representation of women and racial and ethnic minorities. Leadership should compile data showing how the proportion of underrepresented minorities and women at the institution compares with national standards. Additionally, programs should be in place to ensure that adequate pipelines of talented students from underrepresented backgrounds are available. This requires developing community partnerships with local colleges, minority-serving institutions, and historically black colleges and universities. Programs also should be in place to provide social support and proper career development once diverse students and faculty join the institution.

To ensure that a diverse body of students, residents, and faculty contributes to the academic mission, there should be forums for all members of the institution to contribute to different modalities of education, clinical care, and research. A diverse research faculty can bring skill sets that can inform novel research directions. Therefore, the institution must begin to embrace different yet methodologically sound

approaches to research, such as community-based participatory research, working with community partners, and conducting research in nonacademic and nontraditional venues.

In summary a diverse workforce is essential to addressing the needs of a diverse U.S. population, enriching cultural-competency education for students, improving quality of care and access to underserved populations, and developing culturally relevant approaches to research and to addressing the most pressing public health needs. There are compelling reasons to promote health care diversity. Just about every health care organization, from the AAMC to the LCME to the government, has embraced the need for cultural diversity in the education system. Academic health centers contribute to health by upholding their tripartite mission of educating future physicians, developing innovative biomedical research, and providing quality patient care. Fulfilling this mission has become increasingly challenging in light of the ever-changing demographic landscape of the United States. Therefore, strategic planning must occur in order to make diversity in the health care workforce a priority and part of the fabric of the institution.

REFERENCES

1. US Census Bureau. US interim projections by age, sex, race, and Hispanic origin: 2000–2050. http://www.census.gov/population/www/projections/usinterimproj/. Accessed November 11, 2009.
2. Goel MS, McCarthy EP, Phillips RS, Wee CC. Obesity among US immigrant subgroups by duration of residence. *JAMA*. 2004;292:2860–2867.
3. Park Y, Neckerman KM, Quinn J, Weiss C, Rundle A. Place of birth, duration of residence, neighborhood immigrant composition and body mass index in New York City. *Int J Behav Nutr Phys Act*. 2008;5:19.
4. Harley EH. The forgotten history of defunct black medical schools in the 19th and 20th centuries and the impact of the Flexner Report. *J Natl Med Assoc*. 2006;98:1425–1429.
5. Flexner A. *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching*. New York: Carnegie Foundation for the Advancement of Teaching; 1910. Bulletin No. 4.
6. Hoover EL. Did Flexner's Report condemn black medical schools? not so, in my opinion. *J Natl Med Assoc*. 2006;98:1432–1434.
7. Strelnick AH, Lee-Rey E, Nivet M, Soto-Greene ML. Diversity in academic medicine no. 2: history of battles lost and won. *Mt Sinai J Med*. 2008;75:499–503.
8. Smith SG, Nsiah-Kumi PA, Jones PR, Pamies RJ. Pipeline programs in the health professions, part 2: the impact of recent legal challenges to affirmative action. *J Natl Med Assoc*. 2009;101:852–863.
9. Crosby FJ, Iyer A, Sincharoen S. Understanding affirmative action. *Annu Rev Psychol*. 2006;57:585–611.
10. Regents of University of California v Bakke. 438 US 265; 1978.
11. Hopwood v Texas. 78 F3d 932; 5th Cir, cert denied. 116 S Ct 2581; 1996.
12. Cohen JJ. The predictable price of Proposition 209. *Acad Med*. 1999;74:528.
13. The American Medical Student Association, Foundation Study Group on Minority Medical Education. *Findings from Literature Search and Anecdotal Data*. Reston, VA: American Medical Student Association; 1996.

14. Association of American Medical Colleges. Diversity in medical education: facts and figures 2008. http://www.aamc.org/data/facts/.
15. Palepu A, Carr PL, Friedman RH, Amos H, Ash AS, Moskowitz MA. Minority faculty and academic rank in medicine. *JAMA*. 1998;280:767–771.
16. Institute of Medicine. *In the Nation's Compelling Interest: Ensuring Diversity in the Health Care Workforce*. Washington, DC: The National Academy Press; 2004.
17. Cooper LA, Roter DL, Johnson RL, Ford DE, Steinwachs DM, Powe NR. Patient-centered communication, ratings of care, and concordance of patient and physician race. *Ann Intern Med*. 2003;139:907–915.
18. Street RL Jr., O'Malley KJ, Cooper LA, Haidet P. Understanding concordance in patient-physician relationships: personal and ethnic dimensions of shared identity. *Ann Fam Med*. 2008;6:198–205.
19. Laveist TA, Nuru-Jeter A. Is doctor-patient race concordance associated with greater satisfaction with care? *J Health Soc Behav*. 2002;43:296–306.
20. King WD, Wong MD, Shapiro MF, Landon BE, Cunningham WE. Does racial concordance between HIV-positive patients and their physicians affect the time to receipt of protease inhibitors? *J Gen Intern Med*. 2004;19:1146–1153.
21. Missing Persons Minorities in the Professions: A Report of the Sullivan Commission on Diversity in the Healthcare Workforce. 2004. http://www.sullivancommission.org.
22. Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: National Academies Press; 2003.
23. Popescu I, Vaughan-Sarrazin MS, Rosenthal GE. Differences in mortality and use of revascularization in black and white patients with acute MI admitted to hospitals with and without revascularization services. *JAMA*. 2007;297:2489–2495.
24. Bach PB, Pham HH, Schrag D, Tate RC, Hargraves JL. Primary care physicians who treat blacks and whites. *N Engl J Med*. 2004;351:575–584.
25. Brown CP, Ross L, Lopez I, Thornton A, Kiros GE. Disparities in the receipt of cardiac revascularization procedures between blacks and whites: an analysis of secular trends. *Ethn Dis*. 2008;18(Suppl 2):S112–S117.
26. Nguyen LL, Hevelone N, Rogers SO, et al. Disparity in outcomes of surgical revascularization for limb salvage: race and gender are synergistic determinants of vein graft failure and limb loss. *Circulation*. 2009;119:123–130.
27. Pickering TG. Hypertension in blacks. *Curr Opin Nephrol Hypertens*. 1994;3:207–212.
28. Cooper RS, Liao Y, Rotimi C. Is hypertension more severe among US blacks, or is severe hypertension more common? *Ann Epidemiol*. 1996;6:173–180.
29. Lindhorst J, Alexander N, Blignaut J, Rayner B. Differences in hypertension between blacks and whites: an overview. *Cardiovasc J Afr*. 2007;18:241–247.
30. Thackeray R, Merrill RM, Neiger BL. Disparities in diabetes management practice between racial and ethnic groups in the United States. *Diabetes Educ*. 2004;30:665–675.
31. Allsworth JE, Toppa R, Palin NC, Lapane KL. Racial and ethnic disparities in the pharmacologic management of diabetes mellitus among long-term care facility residents. *Ethn Dis*. 2005;15:205–212.
32. Odedina FT, Akinremi TO, Chinegwundoh F, et al. Prostate cancer disparities in black men of African descent: a comparative literature review of prostate cancer burden among black men in the United States, Caribbean, United Kingdom, and West Africa. *Infect Agent Cancer*. 2009;4(Suppl 1):S2.
33. Hershman D, McBride R, Jacobson JS, et al. Racial disparities in treatment and survival among women with early-stage breast cancer. *J Clin Oncol*. 2005;23:6639–6646.

34. Polite BN, Dignam JJ, Olopade OI. Colorectal cancer model of health disparities: understanding mortality differences in minority populations. *J Clin Oncol*. 2006;24:2179–2187.
35. Flenaugh EL, Henriques-Forsythe MN. Lung cancer disparities in African Americans: health versus health care. *Clin Chest Med*. 2006;27:431–439,vi.
36. Boutin-Foster C. An item-level analysis of the Center for Epidemiologic Studies Depression Scale (CES-D) by race and ethnicity in patients with coronary artery disease. *Int J Geriatr Psychiatry*. 2008;23:1034–1039.
37. Morrison RS, Wallenstein S, Natale DK, Senzel RS, Huang LL. "We don't carry that": failure of pharmacies in predominantly nonwhite neighborhoods to stock opioid analgesics. *N Engl J Med*. 2000;342:1023–1026.
38. Leyva M, Sharif I, Ozuah PO. Health literacy among Spanish-speaking Latino parents with limited English proficiency. *Ambul Pediatr*. 2005;5:56–59.
39. Ratzan SP. Health literacy. In: Seldon CR, Ratzan SC, Parker RM, eds. *National Library of Medicine Current Bibliographies in Medicine*. Washington, DC: National Institutes of Health/U.S. Department of Health and Human Services; 2000.
40. Gazmararian JA, Kripalani S, Miller MJ, Echt KV, Ren J, Rask K. Factors associated with medication refill adherence in cardiovascular-related diseases: a focus on health literacy. *J Gen Intern Med*. 2006;21:1215–1221.
41. Green AR, Ngo-Metzger Q, Legedza AT, Massagli MP, Phillips RS, Iezzoni LI. Interpreter services, language concordance, and health care quality: experiences of Asian Americans with limited English proficiency. *J Gen Intern Med*. 2005;20:1050–1056.
42. Mehler PS, Lundgren RA, Pines I, Doll K. A community study of language concordance in Russian patients with diabetes. *Ethn Dis*. 2004;14:584–588.
43. Komaromy M, Grumbach K, Drake M, et al. The role of black and Hispanic physicians in providing health care for underserved populations. *N Engl J Med*. 1996;334:1305–1310.
44. Saha S, Taggart SH, Komaromy M, Bindman AB. Do patients choose physicians of their own race? *Health Aff (Millwood)*. 2000;19:76–83.
45. Marmot MG, Bell R. Action on health disparities in the United States: commission on social determinants of health. *JAMA*. 2009;301:1169–1171.
46. Musa D, Schulz R, Harris R, Silverman M, Thomas SB. Trust in the health care system and the use of preventive health services by older black and white adults. *Am J Public Health*. 2009;99:1293–1299.
47. Rajakumar K, Thomas SB, Musa D, Almario D, Garza MA. Racial differences in parents' distrust of medicine and research. *Arch Pediatr Adolesc Med*. 2009;163:108–114.
48. Horowitz CR, Goldfinger JZ, Muller SE, et al. A model for using community-based participatory research to address the diabetes epidemic in East Harlem. *Mt Sinai J Med*. 2008;75:13–21.
49. Lengerich EJ, Kluhsman BC, Bencivenga M, Allen R, Miele MB, Farace E. Development of community plans to enhance survivorship from colorectal cancer: community-based participatory research in rural communities. *J Cancer Surviv*. 2007;1:205–211.
50. Cohen JJ, Gabriel BA, Terrell C. The case for diversity in the health care workforce. *Health Aff (Millwood)*. 2002;21:90–102.
51. Stoff DM, Forsyth A, Marquez ED, McClure S. Introduction: the case for diversity in research on mental health and HIV/AIDS. *Am J Public Health*. 2009;99(Suppl 1): S8–S15.

52. Hung R, McClendon J, Henderson A, Evans Y, Colquitt R, Saha S. Student perspectives on diversity and the cultural climate at a US medical school. *Acad Med*. 2007;82:184–192.
53. Saha S, Guiton G, Wimmers PF, Wilkerson L. Student body racial and ethnic composition and diversity-related outcomes in US medical schools. *JAMA*. 2008;300: 1135–1145.
54. Umbach P. The contribution of faculty of color to undergraduate education. *Res Higher Educ*. 2006;47(3):317–345.

This page intentionally left blank

SECTION 3

ETHICS AND HUMAN RIGHTS

This page intentionally left blank

CHAPTER 6

Human Rights: A Necessary Framework for Public Health

Holly G. Atkinson, MD

In these early years of the 21st century, human rights and public health intersect in an exciting and vibrant way. By working to achieve public health goals, we actually advance a number of human rights. More importantly, by embracing human rights as an ethical and legal basis and guiding framework, we adopt the most promising and compelling strategy for moving forward a global public health agenda. Human rights are a necessary framework for informing public health, for without it, public health loses much of its transformative potential and moral focus. When left to its own devices, public health does not always improve the health of communities and their members. In fact, public health initiatives sometimes—either by acts of omission or commission—fail to protect the most vulnerable among us or even themselves commit violations of human rights. The Tuskegee Study is an appalling example of a monumental public health failure.

For 40 years, between 1932 and 1972, the U.S. Public Health Service (PHS) conducted a study on nearly 400 men in the late stages of syphilis. The specific goal of the study was to follow the progression of *untreated* syphilis in African Americans men to learn more about the natural course of the infection. The men, most of whom were illiterate sharecroppers from Alabama, were never told what disease they had, or how serious it was. Their informed consent was never sought; in fact, they were deliberately misinformed about the nature of the study. Furthermore, when penicillin became available in 1943, treatment was purposely withheld from the group. When some of the men figured out their disease and sought medication from the PHS, even then they were denied treatment. The study ended only when the Associated Press ran an exposé on the sordid affair in July 1972. The Tuskegee Study, designed to advance knowledge and contribute to public health, now ranks among the most egregious violations of human rights in the history of American medicine.[1]

This chapter explores why a human rights framework is crucial to the work of public health. First, the human rights framework—in concert with traditional medical ethics—articulates certain values and standards that specify how we should conduct ourselves. The Tuskegee episode illustrates the disastrous consequences of

public health initiatives failing to respect, promote, and protect human rights. A human rights framework, however, offers more to public health than professional standards alone. The human rights framework also provides a vast armamentarium—a vision of a more just world, an agenda for change, an array of initiatives shared with civil society, benchmarks for vulnerable groups, and legal strategies to hold state actors accountable—for achieving public health goals. This chapter argues that human rights is a *necessary* framework for public health.

HUMAN RIGHTS AND THE RIGHT TO HEALTH[2]

Human rights is a concept that, while rooted in antiquity, came to the vanguard only during the 18th century, when the American and French revolutions created a new ethical standard. In justifying America's independence from England, Thomas Jefferson stated in the Declaration of Independence, "We hold these truths to be self-evident, that all men are created equal." As historian Lynn Hunt writes in *Inventing Human Rights*, "With this one sentence Jefferson turned a typical eighteenth-century document about political grievances into a lasting proclamation of human rights."[3] Thirteen years later, the Declaration of the Rights of Man and Citizen—born of the French Revolution—declared the "natural, inalienable and sacred rights of man" to be the foundation of government. Although lacking in explicit definition, the idea of the "rights of man" was here to stay, and the nature of politics was forever changed around the world.

Not until the 20th century, however, did the concept of *universal* human rights take hold, culminating as the organizing principle of the Universal Declaration of Human Rights (UDHR). The UDHR is an unprecedented document, born of the tragedies and horrors of World War II and adopted by the United Nations General Assembly on December 10, 1948.[4] The breadth, depth, and universality of human rights articulated by the UDHR stands as one of the landmark achievements of the 20th century.

How do we best define human rights today? In concept, a human right is *any basic right or freedom to which all human beings are entitled and in whose exercise a government may not interfere*. Operationally, human rights refer to "an internationally agreed upon set of principles and norms that are embodied in international legal instruments."[5] These principles and norms were generated through a painstaking, consensus-building process among member states of the United Nations, and thus human rights derive from human institutions created to serve the people.

In the 21st century, human rights advance through a system of laws designed to promote and protect fundamental individual freedoms and human dignity. Rights are categorized into five basic groups: civil, political, economic, social, and cultural rights. Human rights are principally concerned with the relationship between the individual and the state. In this regard, a human rights approach to public health offers much promise, as governments have major roles to play in promoting public health and undertaking or supporting specific public health measures.

What does the "right to health" mean? Every person has the human right *to the highest attainable standard of health*, without discrimination of any kind. The World Health Organization's (WHO) 1946 Constitution was the first to declare, that the enjoyment of the highest attainable standard of health is one of the fundamental rights

of every human being. In the Constitution's Preamble, health is defined as a state of complete physical, mental and social well-being, and not merely the absence of disease or infirmity." The definition goes on to highlight the importance of health promotion, which is "the process of enabling people to increase control over and to improve their health." To do so, "an individual or group must be able to identify and realize aspirations, to satisfy needs, and to change or cope with the environment."

This right to the highest attainable standard of health, or "right to health" for short, embodies a number of specific rights, which in the 21st century—extending beyond the WHO Constitution—are spelled out in various human rights treaties and other documents and should be considered in their totality. They include the following human rights:

- the right to the highest attainable standard of physical and mental health, including reproductive and sexual health;
- the right to equal access to adequate health care and health-related services, regardless of sex, race, or other status;
- the right to equitable distribution of food;
- the right to access to safe drinking water and sanitation;
- the right to an adequate standard of living and adequate housing;
- the right to a safe and healthy environment; and
- the right to a safe and healthy workplace, and to adequate protection for pregnant women in work proven to be harmful to them.[6]

How should we make sense of these rights? How do we make good on the promise of the right to the highest attainable standard of health? In defining the overarching concept of the right to health, Mary Robinson, the former United Nations High Commissioner for Human Rights, stated the following:

> The right to health does not mean the right to be healthy, nor does it mean that poor governments must put in place expensive health services for which they have no resources. But it does require governments and public authorities to put in place policies and action plans which will lead to *available and accessible health care for all in the shortest possible time.* To ensure that this happens is the challenge facing both the human rights community and public health professionals. (emphasis added)[6]

Robinson rightly affirms the centrality of the principle of *progressive* realization of human rights. This means that governments and other important actors must move as expeditiously and effectively as possible to realize a right, especially the right to health. To be sure, governments around the world do not possess the equal capacities to realize the right to health. Some governments possess an abundance of resources and thus have great ability to act on behalf of their populace. Other governments have meager resources at best, leaving them unable to promote the well-being of their citizens. Yet all too often, governments are *unwilling* to act, rather than *unable* to act. Under these circumstances, citizens and concerned groups can and should use rights language and legal instruments to hold their governments accountable.

In summary, human rights are a cohesive set of values and standards that provide us with a powerful framework to move forward the public health agenda. Human rights articulate our ideals, inspire us to act on their behest, set standards for

judging health and social policy, allow for the identification of violations, provide legal concepts to hold actors accountable, and protect the most vulnerable and disenfranchised among us.

PUBLIC HEALTH AND HEALTH STATUS

Writing in *Science* magazine in 1920, C. E. A. Winslow defined public health as

> the science and the art of preventing disease, prolonging life, and promoting physical health and efficiency through organized community efforts for the sanitation of the environment, the control of community infections, the education of the individual in principles of personal hygiene, the organization of medical and nursing services for the early diagnosis and preventive treatment of disease, and the development of the social machinery which will ensure to every individual in the community a standard of living adequate for the maintenance of health.[7]

In the 21st century, the American Public Health Association more succinctly states, "Public health is the practice of preventing disease and promoting good health within groups of people, from small communities to entire countries."[8] Both of these definitions underscore the two distinct aspects of public health work: the *prevention* of disease and the *promotion* of health, rather than the treatment of disease *per se*. Regardless, the focus of public health is on population health, rather than on individual health, which is the primary focus of clinical medicine.

The dichotomy between public health and individual health is an ancient one. The Greeks worshipped two divinities with distinct domains: Hygeia was the goddess of good health and represented what is now the realm of public health: cleanliness, sanitation, good health habits, and ultimately the prevention of illness. Asklepios, her father, was the god of medicine and healing, and one of the most important and widely worshipped divinities of antiquity. The original Hippocratic Oath pays homage to both Asklepios and Hygeia, "I swear by Apollo the physician, and Asklepios, and Hygeia, and Panakeia, and all the gods and goddesses, that, according to my ability and judgment, I will keep this Oath and this stipulation."[9] Although belief in these gods and goddesses has long since passed, the chasm they represent remains alive and well in the 21st century. Ironically, public health initiatives remain one of the most powerful ways to advance individual health and well-being, as well as to contribute to the development of healthy, vibrant communities.

What determines individual health status? Health is influenced primarily by five major factors—personal behavior, genetic influences, social circumstances, access to medical care, and environmental exposures. In the United States, behavior is the major determinant of health, accounting for about 40 percent of all premature deaths in the United States. Medical care, or more accurately the lack of it, plays a surprisingly small role, contributing to only about 10 percent of premature deaths in the United States. The remaining 50 percent of premature deaths is due to genetic influences (roughly 30 percent), social circumstances (15 percent), and environmental exposures (5 percent).[10] Medical science cannot yet alter genetic influences in a clinically meaningful way, but we do know that lifestyle, social factors, and environment exposures alter the manner in which genes express themselves.

The medical literature is replete with evidence linking poor health to socioeconomic disadvantage. Disease and death rates in all nations are closely linked to

social standing: people in lower socioeconomic groups have more disease and die earlier than those in higher groups. To make this point, Mackenbach et al. examined data on mortality, morbidity, smoking, and obesity in relation to socioeconomic status among 22 countries in Europe.[11] In almost all of the countries, researchers found that rates of death and poorer self-assessments of health were substantially higher in men and women of lower socioeconomic status than those of higher socioeconomic status. Writing in an accompanying editorial, Berkman and Epstein note, "The universal link between social class and mortality seems remarkable, given the differing disease prevalence and risk factors in these countries. Moreover, the relationships between class and mortality are consistent for almost every cause of death, with only a few exceptions, notably certain cancers."[12]

People in lower socioeconomic groups typically have more unhealthy behaviors than those in higher socioeconomic groups. This disparity can be traced, in part, to problems that include lack of access to nutritious food, safe neighborhoods, and recreational facilities, as well as higher levels of physical and psychological stress from being socially disadvantaged. Smoking, still considered by experts to be the leading cause of preventable death worldwide, is now more prevalent among people in lower socioeconomic classes than those in the upper classes.

But even when unhealthy behaviors are removed from the equation, poor people still have worse health and higher death rates than wealthier individuals. No doubt the causes are myriad and complex, and more research is needed to understand the link between class and health. We know, for example, that education, job, income, housing, and neighborhood all have a bearing on well-being. Policies that influence the quality and availability of basic necessities of life (education, housing, jobs, and so on) have important health consequences. These factors must be taken into account when leaders of government and civil society set public policy. These concerns fall directly within the realms of both public health and human rights.

THE MILLENNIUM DEVELOPMENT GOALS AND PUBLIC HEALTH

In September 2000, building on a decade of major United Nations conferences and summits, world leaders came together at United Nations Headquarters in New York to adopt the United Nations Millennium Declaration, committing their nations to a new global partnership to reduce extreme poverty and setting out a series of time-bound targets. This declaration has become known as the Millennium Development Goals (MDGs). In 2001, 192 United Nations member states and at least 23 international organizations agreed to work together to achieve eight global development goals (see table 6.1) by the year 2015. Responding to the world's main development challenges and to the calls of civil society, the MDGs promote poverty reduction, education, maternal health, gender equality, and aim at combating child mortality, AIDS and other diseases. The MDGs acknowledge the broader social factors that contribute to health. That is, 3 of the 8 goals, 8 of the 16 targets, and 18 of the 48 indicators relate directly to health. The other interventions address broader social issues that play a significant role in promoting and protecting health and preventing disease.

Over the past decades, substantial progress has been made in improving global health.[13] For example, over the past 25 years, overall life expectancy at birth has increased by nine years to 64.5 years worldwide. Because of a host of political, economic, medical, and social factors, there are, of course, notable exceptions, such as

Table 6.1
Millennium Development Goals

Goal 1: Eradicate extreme poverty and hunger
Goal 2: Achieve universal primary education
Goal 3: Promote gender equality and empower women
Goal 4: Reduce child mortality
Goal 5: Improve maternal health
Goal 6: Combat HIV/AIDS, malaria and other diseases
Goal 7: Ensure environmental sustainability
Goal 8: Develop a global partnership for development

Source: United Nations 2000.

in Africa. Worldwide, death of children age five years or younger has decreased from 13.5 million per year in 1980 to about 9.7 million in 2005. Mortality from infectious diseases, including HIV/AIDS, tuberculosis, and malaria, also is estimated to decline over the next 20 years. And while overall maternal deaths, approximately 500,000 worldwide per year, have not declined substantially over the last two decades, new initiatives have put this extremely important health issue on the public health and women's rights agendas.

Despite promising progress, much remains to be done. What is needed now is a broad public health approach based on the right to health. The three health MDGs—goals 4, 5, and 6—with the deadline looming in 2015, will not be achieved without additional funding. None of the MDGs is expected to be achieved by the target date in Sub-Saharan Africa. Epidemics of infectious disease increasingly threaten the globe, and preparedness for these outbreaks needs to be greatly improved. Gender-based violence and pernicious discrimination against women and girls still ravage the lives of tens of million of females around the world. Chronic conditions such as heart disease, stroke, diabetes, cancer, and chronic respiratory diseases—for decades the major killer diseases in the developed world—are increasingly prevalent in the developing world. By 2030, chronic noncommunicable diseases will account for more than 75 percent of all deaths worldwide, and yet health systems have not engaged in any substantial way to address these burgeoning health problems.[14] Global environmental changes—including rapid climate change—will have considerable health consequences, including changes in drinking water quality, effects of temperature extremes, and the spread of numerous infectious agents. Finally, we have made dismal progress on addressing the underlying socioeconomic determinants of health and need to make it a primary focus of the global public health agenda.

In 2005, the WHO launched the Commission on the Social Determinants of Health to identify strategies for addressing the causes of health inequalities. As articulated by Beaglehole and Bonita:

> The commission's final report stresses several important needs: to improve daily living conditions, including the circumstances in which people are born, grow, live, work and age; to tackle the inequitable distribution of power, money, and resources (the structural drivers of those conditions) worldwide, nationally and locally; and to measure and

understand the problem and assess the effect of action. The report is a major contribution to global public health with its breadth of vision and strong evidence base. Achievement of the fine goals of the report depends on the willingness of WHO, its member states, other key international agencies, and civil society *to strengthen the social justice approach to health* and give greater attention to intersectoral actions. (emphasis added)[13]

We *must* take action on the social factors that influence health to decrease the health disparities and inequalities that abound around the globe. Public health initiatives that mitigate the social determinants of disease and help individuals adopt healthier lifestyles offer the best opportunity to advance health and well-being in all nations, both developed and developing.

In the end, health is an issue of justice. John Rawls's landmark theory of justice argues that "deliberate inequalities [a]re unjust unless they work to the advantage of the least well off," and require corrective social action.[15] The myriad institutional inequalities that abound in health *do not* work to the advantage of the least well off, thus these injustices need to be addressed. But by whom we might ask?

Justice is one of the four major principles that constitute the most commonly promulgated ethical foundation of medicine. But for far too long, it has taken a back seat in informing health initiatives. Medical professionals are ethically bound to address issues of injustice as they pertain to health, and yet in the 21st century, many health care workers—especially clinicians—do not consider it within their professional obligation to engage in societal issues directly effecting health and well-being. Health care professionals need to redress this imbalance. What then does justice require of the profession? When applied to the domain of health, a commitment to justice becomes a declaration of human rights, including the right to health.

THE LINKS BETWEEN HUMAN RIGHTS AND HEALTH

Health and human rights intersect in a number of important ways. Figure 6.1 explores some of these complex and overlapping linkages. First, *violations* of any number of human rights can, and often do, have serious health consequences (upper circle). The Tuskegee Syphilis Study is a blatant example that involved numerous human rights violations, which resulted in horrendous consequences for the subjects. Other examples include gender-based violence, which can lead to a host of physical and mental health problems in women. Children who are trafficked into sex work suffer untold physical and mental harm. Harmful traditional practices, such as female genital cutting and mutilation, can lead to urinary, gastrointestinal, and reproductive health problems, as well as to myriad mental health issues.

Second, by *respecting, protecting, and fulfilling human rights* (lower left circle), individuals' vulnerabilities to and the impact of ill health can be reduced. A human rights approach recognizes that personal behavior is substantially influenced by social, political, and economic circumstances, and thus stresses that governments are obliged to create healthier environments in which people can undertake that personal responsibility, for example, insuring that lower socioeconomic areas have access to fresh fruits and vegetables rather than mostly fast

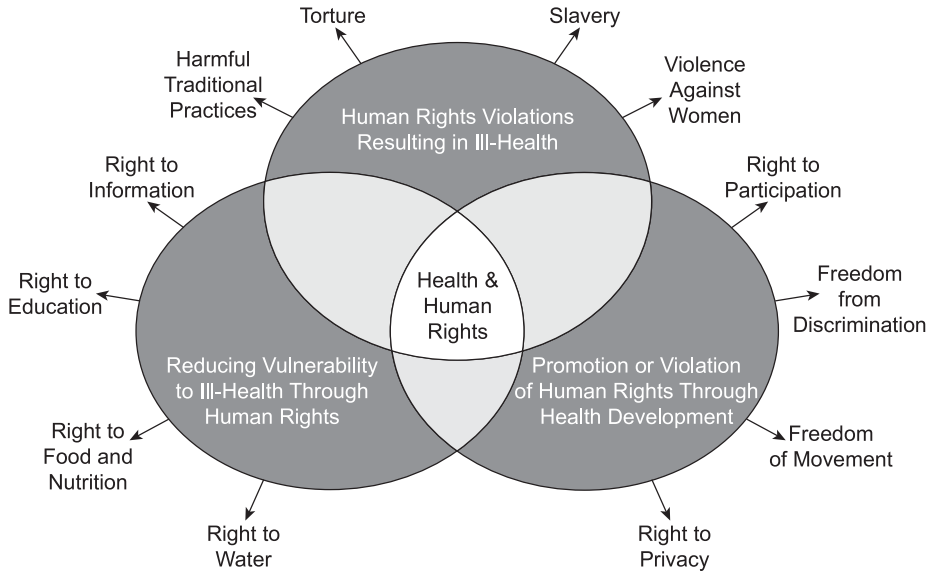

Figure 6.1 Linkages Between Health and Human Rights, from the WHO, July 2002. (*Source:* Atkinson.)

foodchains. When human rights other than the right to health are addressed, health is promoted as well. For example, access to accurate information as well as to an education promotes health and well-being.

Third, *direct efforts at health development* (lower right circle) can either promote or violate various human rights of individuals or groups, which ultimately bears on their well-being. For example, limiting women's access to reproductive health information can have a significant impact on morbidity and mortality, as well as on quality of life. People with mental disabilities often are discriminated against, which has a broad impact on their rights to employment, education, and housing, which in turn have a major impact on their health and well-being. A particular focus on human rights protection and promotion is extremely important to the design, and ultimately the outcome, of health policies and programs.

To be sure, in some situations the exercise of good public health policy necessitates infringing on some individual human rights. For example, sometimes it is appropriate to quarantine an individual to control the spread of a serious infectious disease, and thus the right to freedom of movement can be abridged. International human rights instruments do recognize that situations will arise that justifiably entail restricting certain rights. (Certain rights, however, such as the freedom from slavery or torture, never should be compromised.) To ascertain the appropriateness of restricting rights, human rights law articulates the Siracusa Principles, which state that

> Public health may be invoked as a ground for limiting certain rights in order to allow a state to take measures dealing with a serious threat to the health of the population or

individual members of the population. These measures must be specifically aimed at preventing disease or injury or providing care for the sick and injured.[16]

The Siracusa Principles further stipulate that before rights can be restricted, five criteria must be met:

1. the restriction is provided for and carried out in accordance with the law;
2. the restriction is in the interest of a legitimate objective of general interest;
3. the restriction is strictly necessary in a democratic society to achieve the objective;
4. there are no less intrusive and restrictive means available to reach the same objective; and
5. the restriction is not drafted or imposed arbitrarily, i.e., in an unreasonable or otherwise discriminatory manner.

The restriction of rights should be undertaken only as the last resort, with considerable justification and with safeguards in place. And when it is necessary to curtail individual rights in the service of public good, the rights should be restricted for the shortest possible time.

THE VALUE ADDED BY HUMAN RIGHTS IN PUBLIC HEALTH ARENA

How does a human rights approach to health add value to public health initiatives? First and foremost, it moves health out of the realm of being a commodity or charity into the realm of constituting moral and legal obligations of a society—especially governments—to its people. The human rights approach provides public health with (1) a guiding *framework*; (2) *strategies* for addressing inequities and disparities in health; (3) *tools* for designing, implementing, monitoring, and evaluating public health initiatives; and (4) the *foundation* for holding governments and institutions accountable.

The framework of human rights articulates internationally accepted norms and standards against which public policy can be planned, monitored, and ultimately judged. This, in and of itself, is an important step forward in human moral and legal development. For the first time in human history, we have articulated ideals that describe society's obligations to all citizens in all places at all times. These ideals and obligations are particularly inspiring and empowering to individuals who are poor, disenfranchised, or oppressed, and give them a moral *and* legal foundation on which to base their demands.

Secondly, the human rights approach offers strategies to bring about change. For example, acknowledging that education is a major socioeconomic determinant of health expands the domain of public health initiatives to include promoting and protecting the right to education. By understanding the biological and cultural role gender plays in determining health status, public health officials can use a gendered lens to analyze data, policies, and programs to achieve optimal health among women and men, as well as to reduce the profound impact of gender discrimination on women's well-being.

The human rights approach insists on designing, implementing, monitoring, and evaluating programs and policies in terms of how well they move communities toward positive health outcomes. Benchmarks and indicators can be identified to

track the progressive realization of health rights. The human rights approach especially urges public health officials to consider the impact of programs and policies on vulnerable groups, such as ethnic and religious minorities, internally displaced persons, refugees of war, immigrants, prisoners, victims of sexual trafficking, and the extremely disenfranchised, such as impoverished women and children. For example, data amply demonstrate that

> respect for human rights in the context of HIV/AIDS, mental illness, and physical disability leads to markedly better prevention and treatment. Respect of the dignity and privacy of individuals can facilitate more sensitive and humane care. Stigmatization and discrimination thwart medical and public health efforts to heal people with disease or disability.[17]

Finally, the human rights approach, in articulating the highest standard of health, provides a moral and legal basis for holding governments and other institutions accountable. For as Dr. Helen Potts, chief program officer of Health Programs at Physicians for Human Rights, has said, "the distinctive contribution of a human rights approach is the essential component of accountability by government[s] to the[ir] population[s]."[18] Human rights activists and public health officials, with outcomes data in hand, can press governments to take definitive steps to protect, promote, and fulfill their human rights obligations. This means establishing concrete goals, committing available resources, and executing plans in a timely way to advance health outcomes. At times, legal remedies can be sought in courts of law (for example, individuals with HIV/AIDS have petitioned the courts for access to antiretroviral drugs), while at other times, mobilization and advocacy in civil society or in the broader international arena can drive the public health agenda (for example, the International Campaign to Ban Landmines resulted in the signing of the Ottawa Treaty in December 1997 that bans government use of landmines).

The human rights approach offers an armamentarium to public health that is powerful indeed. It is particularly forceful in addressing issues of social justice and health. It gives a strong voice to the needs of the disenfranchised and vulnerable, who disproportionately suffer the negative impact of the social determinants of disease. As WHO official Ms. Helena Nygren-Krug has put it, "Linking health and human rights could act as a force for mobilizing and empowering the most vulnerable and disadvantaged. Advancing health as a human rights right means making people conscious of both their oppression and the possibility of change."[19]

THE CHALLENGES WE FACE AND THE POSSIBILITY OF CHANGE

It has been only 65 years since the right to health was articulated and subsequently codified in a number of important 20th-century documents. Since then, the concept of human rights has grown in importance around the globe. Human rights now play a central role in all major social, cultural, political, and economic discourses. Health and human rights is a nascent but rapidly growing field within both the human rights arena and the public health arena. Given the brief time this framework has informed modern thinking, the impact that human rights, particularly the right to the highest attainable standard of health, already has made on the world

stage is impressive and inspiring. For example, more than 100 national constitutions worldwide establish health as a human right. And every country in the world has signed on to at least one human rights treaty that enshrines language addressing the goal of attaining the highest attainable standard of health. This is good news indeed, although much remains to be done. Although nations have articulated the right to the highest attainable standard of health, it is the particular duty of health care professionals in the public health arena to progressively *realize* the right to health in a concrete way. Our work has just begun.

How do we move forward? We need three broad initiatives: we need to teach about, advocate for, and operationalize the right to health.

• Teaching the right to health: The core curriculum of health care education, including public health schools, nursing schools, *and* medical schools, should include comprehensive instruction on health and human rights, especially the right to health. Without a deep understanding of the human rights approach, public health and medical personnel lack a powerful framework for driving the public health agenda. Young idealistic students crave this sort of instruction. They search for a way to process what they bear witness to: the deep inequities in society and health care that profoundly affect individual health. They seek a principled way to bring about meaningful change. As Len Rubenstein, former president of Physicians for Human Rights has said,

The reluctance of medical schools to address these issues is understandable: the curriculum is crowded; few faculty have experience in addressing these issues; and most of all there is distinct queasiness about bringing human rights concerns into medical education because it is somehow 'political' and therefore has no place in an education founded in scientific evidence. If the next generation of physicians is to be fully prepared to deal with the social and moral dilemmas they'll face in the coming decades, however, medical education will need to push its way through these barriers.[20] If we fail to bring about such needed education reform, we will fail yet another generation of health care workers who will, in all likelihood, see their idealism driven out by growing cynicism in the face of unattended people and unaddressed injustices the world over.

• Advocating for the right to health: All health professionals, but especially those in public health, have a primary ethical and professional responsibility for the health of the community members they serve. Because the social economic determinants of health and the full realization of human rights play such powerful roles in individual health and well-being, we have a special obligation to advocate for the right to health, and for public policy changes that address institutionalized inequities, including the lack of access to health care for the uninsured, lack of access to clean water and nutritious food, and discrimination in the provision of health services. Health advocacy initiatives need to be based on evidence and professionalism, not on individual political persuasions. The tools of medicine and science can be used to document human rights violations and to analyze the impact of the social determinants on disease. It is the data that must inform advocacy initiatives and subsequently the public debate on critical issues that deeply affects health and well-being. For public health professionals to engage effectively in advocacy work, we need to promote the attitudes of good citizenship in public health and medical education, and we need to teach advocacy skills to all health care professionals. Advocacy work requires a set of skills: the organization of people, goal setting, fact-finding, research, identification of power structures, strategizing, coalition building, taking action,

and analyzing outcomes. While public health professionals already engage in many forms of advocacy, the use of a human rights framework will add even greater moral and legal authority. Advocacy for human rights offers both an inspiring and powerfully pragmatic way to increase the influence of public health professionals on public policy, especially in demanding that governments reduce health inequities and promote the health of their people—and in supporting governments as they take on these challenges.

• Operationalizing the right to health: Even though most countries have ratified at least one treaty that enshrines the right to health as an obligation of their governments to their people, vast health inequities are firmly entrenched around the world, including in prosperous countries such as the United States.[21] Only during the last decade or so have nations, human rights organizations, international organizations, and other members of civil society, including public health officials, begun the task of "operationalizing" the right to health. In essence, it means putting the concept of the right to health into *practice*. What exactly does this mean and how should we accomplish this? In reality, we are still finding our way. We are still learning. Certainly this is a highly complex and necessarily long-term endeavor. After all, the right to the highest attainable standard of health is a right that can be progressively realized only over time, given the resource constraints within any society at any time. Nevertheless, putting the right into practice entails a number of critical activities, including but not limited to the following: committing to a comprehensive governmental plan; working to strengthen health systems; conducting groundbreaking research on using a rights-based approach; advocating for essential shifts in policy; ensuring equity, equality, and nondiscrimination for the vulnerable and marginalized; setting benchmarks and standards; developing a transparent process of operating; ensuring participation by all; promoting gender equality; tracking and analyzing outcomes, especially as they pertain to disadvantaged and marginalized groups; and monitoring the progression of realizing the right to health. We are in the early years of operationalizing the right to health. Public health officials have much to contribute to this critical work.

In *The Future of Public Health*, the Institute of Medicine says: The *mission* of public health is "the fulfillment of society's interest in assuring the conditions in which people can be healthy." The *substance* of public health is "organized community efforts aimed at the prevention of disease and the promotion of health." The *organizational framework* of public health "encompasses both activities undertaken within the formal structure of government and the associated efforts of private and voluntary organizations and individuals."[22] There is one more element we should add to this compelling force for change: human rights provide us with a powerful *ethical and legal framework* for moving forward the global public health agenda. It offers us an inspiring vision, articulates an obligation, lays out an aggressive agenda, provides a framework for analysis and accountability, and ultimately commits us to dignity for all. In the end, the human rights framework offers us a reason to believe in the possibility of change.

REFERENCES

1. Tuskegee University. Research ethics: the Tuskegee syphilis study. http://www.tuskegee.edu/global/story.asp?s=1207598. Accessed December 5, 2009.
2. The full formulation of the right is "the right of everyone to the enjoyment of the highest attainable standard of physical and mental health." In this paper, "the

right to health" and "the right to the highest attainable standard of health" are used as shorthand.
3. Hunt L. *Inventing Human Rights: A History*. New York: W.W. Norton & Company; 2007:15.
4. Other important instruments and treaties that articulate human rights related to health include the International Covenant on Economic, Social and Cultural Rights; the Convention on the Elimination of All Forms of Discrimination Against Women; and the Convention of the Elimination of All Forms of Racial Discrimination; the Convention on the Rights of Persons with Disabilities; and the Convention of the Rights of the Child.
5. World Health Organization. *25 Questions and Answers on Health and Human Rights*. Geneva, Switzerland; 2002. Health and Human Rights Publication Series, Issue No.1.
6. World Health Organization. *25 Questions and Answers on Health and Human Rights*. Geneva, Switzerland; 2002. Health and Human Rights Publication Series, Issue No.1. See also Committee on Economic, Social and Cultural Rights. UN Doc E/C.12/2000/4: General Comment No. 14: the right to the highest attainable standard of health. http://www.ohchr.org/english/bodies/cescr/comments.htm. Published 2000.
7. Winslow CEA. The untilled fields of public health. *Science*. 1920;51(1306):23–33.
8. American Public Health Association. Overview. http://www.apha.org/about/. Accessed November 29, 2009.
9. Theoi Project. Hygeia: Greek goddess of good health. http://www.theoi.com/Ouranios/AsklepiasHygeia.html. Accessed November 29, 2009.
10. Schroeder SA. We can do better: improving the health of the American people. *N Engl J Med*. 2007;357:1221–1228.
11. Mackenbach JP, Stirbu I, Roskam A-JR, et al. Socioeconomic inequalities in health in 22 European countries. *N Engl J Med*. 2008;358:24682481.
12. Berkman L, Epstein A. Beyond health care: socioeconomic status and health. *N Engl J Med*. 2008;358:2509–2510.
13. Beaglehole R, Bonita R. Global public health: a scorecard. *Lancet*. 2008;372:1988–1996.
14. Fuster V, Voute J. MDGs: chronic diseases are not on the agenda. *Lancet*. 2005; 366:1512–1514.
15. Kirch DG, Vernon DJ. The ethical foundation of American medicine: in search of social justice. *JAMA*. 2009;301:1482.
16. UN Commission on Human Rights. *The Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant on Civil and Political Rights*. http://www.unhcr.org/refworld/docid/4672bc122.html. Published September 28, 1984. Accessed November 29, 2009. UN Doc E/CN.4/1985/4.
17. Mann J, Gruskin S, Grodin M, Annas G. *Health and Human Rights: A Reader*. Routledge, 1999. Introduction, paragraph 4.
18. Potts, H. Accountability and the right to the highest attainable standard of health. Human Rights Centre, University of Essex, 2008. (Accessed November 29, 2009 at http://www.essex.ac.uk/human_rights_centre/research/rth/docs/HRC_Accountability_Mar08.pdf.
19. World Health Organization. New WHO publication explores important links between health and human rights, an area drawing increased attention. Press Release WHO, July 2002. (Accessed November 29, 2009 at http://www.who.int/hhr/activities/q_and_a/q_a_press/en/index.html.
20. Rubenstein, LS. Viewpoint: The human rights imperative in medical education. AAMC Reporter. January 2007. (Accessed November 29, 2009 at http://www.aamc.org/newsroom/reporter/jan08/viewpoint.htm.)

21. Physicians for Human Rights. The right to equal treatment: A report by the panel on racial and ethnic disparities in medical care. PHR, September 2003. (Accessed on November 29, 2009 at http://physiciansforhumanrights.org/library/report-equaltreatment-2003.html.
22. Institute of Medicine. The Future of Public Health. National Academies Press, 1988. (Accessed November 29, 2009 at http://www.nap.edu/openbook.php?isbn=0309038308.)

CHAPTER 7

Genetic Testing and Public Health: Ethical Issues

Inmaculada de Melo-Martín, PhD, MS

INTRODUCTION

The publication in 1953 by James Watson and Francis Crick of the three-dimensional molecular structure of DNA resulted in remarkable theoretical and technological achievements during the next decades.[1] The launching of the Human Genome Project, in 1991, which culminated in the publication of the full sequence of the human genome in April 2003, has provided the impetus for an impressive advance in our knowledge of molecular science and our understanding of the genetic etiology of a variety of diseases. New developments in genomic research likely will lead to the identification of the molecular factors that underlie not just rare conditions such Tay-Sachs, Huntington's disease, and heritable breast cancer but also common multifactorial or complex diseases that represent a significant public health burden, such as cardiovascular diseases, asthma, diabetes, hypertension, cancers, obesity, and psychiatric diseases.[2–4]

This new research and knowledge has been the basis of an ever-increasing number of genetic tests that now are used under the auspices of the health care system and even ordered directly by consumers over the Internet.[5–8] Genetic testing initially was used as a resource to identify rare inherited disorders, such as Tay-Sachs and Huntington's disease, which usually affect a small percentage of the population. As more tests are developed, naturally the opportunities are expanding. New tests are being developed that detect genetic alterations that may influence more complex and common conditions, such as breast and ovarian cancer, cardiovascular diseases, colon cancer, or Alzheimer's disease.[6, 7] Currently, more than 1,000 genetic tests are available from testing laboratories. Although genetic testing and screening of a variety of human diseases are now part of medical practice, it is still not completely clear what the impact of these new molecular and genetic tools will be on public health. It is clear, however, that the incorporation of genetic tests and population screening programs as tools to improve public health raise a number of ethical concerns.

The purpose of this chapter is to examine some of these ethical issues. In the next section, I offer a brief overview of the different types of genetic tests that are available. In the second section, I present some of the most significant ethical concerns that arise in relation to the use of genetic tests. I discuss first matters related to the analytic and clinical validity and utility of genetic tests and how these aspects result in ethical quandaries. Next, I focus on the concerns that the use of genetic tests, if such tests prove beneficial for the populations' health, might contribute to furthering existing health inequities. Finally, I discuss ethical issues related to obtaining, or omitting, informed consent and to protecting privacy and confidentiality.

GENETIC TESTING AND SCREENING

Although the terms usually are used interchangeably, genetic testing and screening serve different purposes. Genetic testing targets individuals who might be at an increased risk of having a genetic disorder. They can be used to diagnose individuals with symptoms, to determine disease susceptibility in individuals who are asymptomatic, to ascertain genetic risks for offspring, as well as to guide medical treatment.[9–11] Genetic screening, on the other hand, is directed to populations to identify individuals at a higher risk for a genetic disease or condition. Several genetic screening programs, such as newborn screening for a variety of conditions, carrier screening for hemoglobinopathies, and prenatal screening to detect chromosomal abnormalities, are an established part of public health practice and research in the United States and abroad.[12–14]

Genetic tests, whether for testing or screening, usually are understood as those tests that involve the analysis of a person's chromosomes, DNA, RNA, or proteins to diagnose or rule out particular genetic disorders or to predict the likelihood of suffering such disorders in the future. These tests usually are performed on a sample of blood, but other tissues such as amniotic fluid, embryonic cells, hair, saliva, or skin also can be used. Once the tissue sample is obtained, laboratory technicians look for specific changes in chromosomes, DNA, RNA, or particular proteins that might indicate the existence of a disorder or a susceptibility to developing a particular disease.

As mentioned, genetic tests can be used for diagnostic or predictive aims. Diagnostic testing attempts to confirm whether an individual has a genetic or chromosomal condition. It usually is offered to people who show some signs of suffering a particular genetic disease or who have a family history of the disease. Diagnostic tests can be performed at any time during a person's life span. Predictive testing, on the other hand, is used to identify mutations that might increase an individual's risk of developing a disorder with a genetic basis later in life. Healthy people, with or without a family history of a particular disease, can be candidates for this type of test.

The most widespread use of genetic tests is carried out in newborn babies to identify certain diseases for which early diagnoses and treatment exist.[12] In the United States, for example, all states currently test infants for phenylketonuria (PKU) and hyperthyroidism. Also relatively common is preimplantation genetic diagnosis used in conjunction with in vitro fertilization. The embryos thus produced are then tested to identify genetic abnormalities. This testing usually is offered to

couples with an increased risk of having a baby with a genetic or chromosomal disorder and is used to assess the viability of the embryos. Similarly, prenatal genetic diagnosis seeks to detect genetic or chromosomal abnormalities in developing fetuses. As in the case of preimplantation diagnosis, this test is offered to couples when there is suspicion of an increased risk of a genetic disorder. Finally, carrier testing aims to determine whether an individual carries a copy of a mutated allele that, when present in two copies, might cause a genetic disease. Often this test is offered in the context of reproductive planning to determine the risks of having a child with a particular genetic disorder.

ETHICAL ISSUES RAISED BY GENETIC TESTING AND SCREENING IN PUBLIC HEALTH

Development and implementation of new medical procedures often present ethical concerns. From the involvement of research subjects necessary to assess the safety and efficacy of such procedures, to issues related to the quality of informed consent, to effects on social justice, new biomedical developments require careful consideration of ethical implications. Genetic testing and screening thus are not unusual in this respect. Indeed, many have argued that genetic information actually brings up unique concerns because the nature of genetic data is different from other types of health information in that it is thought to provide definitive life-long health risk information, can reveal information about family members and offspring, and is highly identifiable.[15–17] The recently passed Genetic Information Nondiscrimination Act is an attempt, at least in part, to assuage some of the concerns that the use of genetic information might raise in relation to privacy and the possibility of discrimination.[18] But whether genetic information is, or is not, exceptional, it certainly is the case that the use of genetic tests presents us with important ethical concerns.[19]

THE USEFULNESS OF GENETIC TESTS

It seems clear that, if adequately implemented, both genetic testing and screening can contribute to improving the health of the population. Information on the contribution of genetic factors to disease, identification of people at higher risk for particular conditions, preventative measures, and timely treatments all could result from the suitable use of genetic testing and screening and all would foster public health goals. But, as with any other medical test, how well these goals can be furthered depends on the predictive value of genetic tests. Such assessment needs to consider the analytic validity, clinical validity, and clinical utility of the tests.[20–22]

Analytic validity refers to the ability of a test to accurately measure a particular genetic characteristic in the laboratory. Although for some genetic tests evidence indicates that the analytic validity is high, genetic and genomic technologies are complex and validation data for many of these tests are limited. Moreover, for the test to result in health benefits, clinical validity and utility need to be taken into account.[20, 21] Clinical validity describes the ability of the genetic test to reliably predict the presence or absence of a clinically defined disorder or phenotype of interest. Clinical utility refers to the ability of a test to result in improved health

outcomes. The clinical utility of a genetic test then gives us information about the risks and benefits that result from its use.

Assessing analytic validity and clinical validity and utility of genetic tests is not always easy.[20, 23–25] The clinical validity and utility of genetic tests that detect highly penetrant gene mutations and polymorphisms, that is, the trait the particular mutation produces will almost always be apparent in an individual carrying such an allele and can be quite high as is the case of tests for Huntington's disease or PKU. Clinical validity and utility, however, becomes more questionable for tests that attempt to predict complex diseases, such as cardiovascular diseases, cancers, autoimmune disorders, neurodegenerative diseases, and nutritional disorders—that is, diseases that are responsible for the majority of the disease burden. This is the case because genes with reduced penetrance or variable expressivity are involved in these disorders, that is, the range of signs and symptoms that can occur in different people with the same genetic condition.[24, 26, 27]

Complex diseases or disorders result from mutations occurring simultaneously in several genes. Moreover, alleles contributing to these complex diseases are neither necessary nor sufficient to cause the particular disease. Some people might suffer the disease without having the related mutations and some people might carry the mutations but not have the disease in question. For many of these complex diseases, more than one gene at different loci contributes to the disease and those loci might interact with each other. Furthermore, modifier genes also can interact with mutations involved in the production of some diseases. The effects of interaction between an allele that might predispose toward having a particular disease and a protective allele might be especially difficult to predict with any accuracy. Similarly, epigenetic factors can modify the expression patterns of genes without altering the DNA sequence and the expression of most human diseases also involves the relations of multiple genetic and environmental factors.[28–30] Moreover, the key mutations for a particular disease might vary between different populations and thus the clinical validity of a genetic test might be limited by the particular mutations that are tested as well as in the particular populations in which the test is administered.[20, 21, 22]

Clearly, whether genetic tests have appropriate analytic validity, clinical validity, and utility raises not only scientific issues about the quality of scientific evidence, but also ethical ones. Harms to both the particular individuals using the tests and to society might result in cases in which the clinical validity and utility of these tests are questionable. Individuals obtaining information about future health states through the use of genetic tests of unproven utility might make problematic decisions about their clinical care.[31, 32] Moreover, information of unproven clinical value can result in the use of further medical tests and procedures that can produce anxiety in the patients and expose patients to unnecessary risks.[33] Because of the particular importance that we tend to attach to genetic information, these tests can result in labeling effects that can lead healthy people to see themselves as impaired.[15, 34]

Harms to society also can occur when the clinical validity and utility of these tests is questionable, which might contribute to increased costs to the health care system without appropriate benefits to offset such costs. For instance, the Advisory Committee on Heritable Disorders and Genetic Diseases in Newborns and Children

recently recommended that all states screen newborns for a standard range of 29 diseases. They also included an additional 25 conditions that could be identified in the course of screening as new targets.[35] There is a paucity of studies, however, that have investigated a benefit-risk balance for many of the conditions screened.[36, 37] Thus, we might be providing genetic tests that are relatively expensive when it is unclear whether any clinical benefit is derived from the use of scarce resources. Moreover false positives that result in unnecessary treatments or further testing also contribute to increasing health care costs.

HEALTH DISPARITIES

Genetic testing and screening programs have been espoused as important tools to improve public health outcomes. What effect these tests might have on the health of the population depends not only on whether the tests are able to predict correctly the risks to health as well as on whether treatments are available or preventive measures can be put in place, but also on whether people will have access to these tests, treatments, and preventive measures. Genetic tests' prices range from $200 to more than $3,000 depending on the complexity of the test, the number of individuals tested to obtain a result that is meaningful, or the method of specimen handling.[38] For example, depending on the methodology used and the type of test performed, costs of tests for breast cancer mutations such as BRCA1 and BRCA2, range from about $300 to $3,000.[39] Obviously, under a health care system such as the one in the United States, concerns about the costs of genetic testing are relevant because cost might limit access. If it does so in significant ways, then the benefits for public health of genetic testing would be hard to realize.

Moreover, if some individuals or populations who are already disadvantaged are unable to benefit from the genetic revolution, then we run the risk of furthering injustices against them by increasing health disparities. As a significant amount of evidence shows, lack of access to effective medical technologies can contribute to an increase in heath inequities. Indeed, in spite of the improvements in the overall health of the United States, racial and ethnic minorities continue to receive a lower quality of health services and have higher rates of morbidity and mortality than nonminorities. According to the Centers for Disease Control and Prevention (CDC), for example, in 2001 the age-adjusted death rate for cancer was 25.4 percent higher for African Americans than for white Americans. The infant mortality rate among African Americans was more than twice the rate for white Americans.[40] Similarly, a recent report by the American Heart Association showed that blacks have a 1.3-times greater rate of nonfatal stroke, a 1.8-times greater rate of fatal stroke, and a 1.5-times greater rate of heart disease death than whites. Also, among the adult population over age 20, non-Hispanic blacks, Mexican Americans, American Indians, and Alaska Natives have a higher prevalence of diabetes than non-Hispanic whites.[41]

Lack of access to health care benefits affects minority populations in significant ways. Hispanic and non-Hispanic black individuals are more likely to lack health insurance than non-Hispanic white persons. In 2007, more than 10 percent of non-Hispanic whites were uninsured; the uninsured rates for blacks and Hispanics were 19.5 percent and 32.1 percent, respectively.[42] The fact that the U.S. health

insurance system is primarily based on employer contributions also affects minority populations significantly. This is so because minorities have higher levels of unemployment; for example, in May 2009 the unemployment rate for Hispanics was 12.7, and for African Americans was 14.9, whereas for non-Hispanic white people it was 8.6.[43] Furthermore, more often than not minority populations tend to work full time or part time in service jobs or in temporary jobs that do not provide health insurance coverage.

But, even for those people who do have insurance, access to genetic testing is not guaranteed. At present, state laws with regard to access to, and coverage of, genetic services are limited to newborn screening and childhood genetic diseases. States do not require health insurance coverage of genetic testing for adult onset disorders. Group health insurance plans often do not cover screening tests in the absence of symptoms and thus exclude coverage of genetic testing for many diseases. Where coverage exists for genetic testing, the necessary education and counseling that should accompany such tests often are not covered by insurance.[44] Given the fact that minorities will have more difficulties having access to the possible health benefits associated with genetic testing, then it is not unreasonable to believe that these technologies might contribute to furthering existing health inequities.

INFORMED CONSENT

Shortly after the Nuremberg trials, which presented horrifying accounts of Nazi experimentations on unwilling human subjects, the issue of informed consent began to receive attention.[45] The first sentence of the Nuremberg Codes states that the voluntary consent of human subjects in research is absolutely essential.[46] At Helsinki in 1964, the World Medical Association made consent of patients and subjects a central requirement of ethical research.[47] Since then, virtually all prominent medical and research codes as well as institutional rules of ethics dictate that both physicians and investigators obtain the free informed consent of patients and subjects before performing any substantial intervention. Although obtaining free informed consent serves several goals, such as protecting patients and subjects from harm or encouraging medical responsibility in interactions with patients and subject, the most fundamental aim is to enable autonomous choices.[45, 48]

Legal, regulatory, medical, psychological, and philosophical literature tend to analyze informed consent in terms of the following elements: (1) disclosure, (2) understanding, (3) voluntariness, and (4) competence.[49] According to this understanding of informed consent, one gives free informed consent to an intervention if and only if one is competent to act, receives a thorough disclosure about the procedure, understands the disclosure, acts voluntarily, and consents to the intervention. Disclosure refers to the obligation that professionals have to offer adequate information to patients and subjects. Understanding, which may be the most important component for free informed consent, requires professionals to help patients and subjects to overcome illness, distorted information, irrationality, or other factors that can reduce a patient's grasp of the situation to which he or she has the right to give or refuse consent.

Patients and subjects need to have some basis of understanding the diagnoses, prognoses, nature, and purpose of the intervention as well as the tests' alternatives,

risks, benefits, and recommendations. Voluntary action requires that people are not constrained by manipulation and coercion by other persons. Coercion occurs if and only if one person intentionally uses a credible and serious threat of harm or force to control another. One important form of manipulation in health care is informational manipulation, a deliberate act of handling information that alters patients' understanding of the situation and motivates them to do what the agent of influence plans. The way in which doctors present information by tone of voice, by framing information positively (the therapy is successful most of the time) rather than negatively (the therapy fails in 40 percent of the cases) can influence an individual's perception and, therefore, affect understanding. Finally, the criterion of competence refers to an individual's ability to perform a task. Patients or subjects are competent if they have the ability to understand the material information, to make a judgment about the evidence in light of their values, to intend a certain outcome, and freely to communicate their wishes to the professionals.

The duty to respect patients' autonomy then would seem to require that doctors obtain informed consent from their patients before they perform genetic tests. Indeed, obtaining and documenting informed consent is an essential component of the relationship between physicians and their patients. It allows patients to receive necessary information about their condition, the risks and benefits of performing such tests, the meaning and clinical usefulness of the information provided, and other possible alternatives, as well as information about existing treatments or preventive strategies.

Genetic testing and screening programs raise two different ethical issues in relation to informed consent and the respect for autonomy that such consent attempts to support. The first one relates to the quality of informed consent that is called into question because of the difficulties associated with the disclosure and understanding of genetic information. The second one pertains to genetic screening programs in particular and the mandatory nature of some of such programs, for example, newborn screening policies. In the second case, the concern is that the justification for omitting informed consent is lacking and thus that such policies are not ethically sound. In what follows, I discuss these two different concerns: quality of informed consent and mandatory programs omitting informed consent.

Disclosure of adequate information and understanding of the complexities of genetic information, as well as the psychological and social implications of performing a particular genetic test, and the indirect involvement of third parties, might make adequate informed consent difficult to obtain. As we have seen, the analytic validity as well as the clinical validity and utility of many of the genetic tests now available are difficult to ascertain and thus the quality of disclosure of pertinent information could be compromised. Perhaps if patients were somewhat knowledgeable about genetic information and genetic testing practices, problems related to disclosing and understanding the uncertainties associated with genetic tests would be assuaged.

A variety of studies, however, have revealed that members of the public have important misconceptions about genetic concepts and the use and availability of genetic technologies. Studies have shown that despite widespread efforts to popularize Mendelian genetics throughout the century, public understanding of genetic science is limited.[50–54] Some research has suggested that Mendelian explanations

of inheritance are poorly accepted and understood because they conflict in a number of ways with a widespread lay understanding of inheritance that is derived from the social relationships of kinship. It might be that people do not find the Mendelian concept easy to grasp because it does not fit with what they already believe.[55] People, even when their attitudes toward genetics are positive, often lack adequate information about available genetic technologies, their use, and effectiveness, and how to have access to such technologies.[53, 56, 57]

Studies about public perception of health risks also indicate that people across all different ages, races, and socioeconomic groups, often underestimate or overestimate their risk of disease and, in general, lack adequate knowledge about risk factors.[58–61] This is problematic when people need to make decisions about whether to pursue, or not, particular genetic tests as this would depend in part on their real or perceived risk factors for a disease.

A proper estimation of risks is particularly important to make sense of genetic tests results given that genetic tests provide information about the presumed risks of suffering a particular disease. It is well known, however, that both experts and laypeople have difficulties calculating and understanding probabilities of risk and risk-related information, especially when that information is presented to them quantitatively.[62, 63] For example, people tend to believe that events are more probable when they can recall an incident of its occurrence. Also, people often disregard cumulative probabilities when they are exposed to the same risk factors over a longer time period or when they are exposed simultaneously to several risk factors. Instead, they perceive each hazard as a single, independent exposure.[62, 64]

Individuals' understanding of risk appears to be informed by particular cultural and cognitive biases.[65] People are more often than not unfamiliar with uncertainty in risk assessment and their interpretation of and responses to uncertainty depend on their personal characteristics and values all of which can be affected by the manner in which uncertainty is communicated.[66] Studies also show that recipients of genetic information related to reproductive concerns reduce probabilistic information to a dichotomous interpretation—that is, it either will or will not happen.[67] Finally, evidence suggests that people have difficulties understanding the results of genetic tests or whether the tests predict risk or diagnose disease.[68] Even when people might have sound knowledge of a particular genetic condition, they can have difficulties correctly recalling the results of genetic tests.[69]

Clearly, understanding genetic information and the implications for one's health of genetic tests is not an easy task for most people. But if this is so, then adequate informed consent might be difficult to obtain as one cannot provide such consent if the information necessary to give consent is misunderstood. A possible solution to the problems that result from insufficient or inadequate knowledge is to ensure that patients have access to recommended consultations with genetic counselors and medical geneticists. Certainly, appropriate counseling and guidance on how to interpret the results of genetic tests, on what the implications for someone's health might be, or on what the effect might be for reproductive decision making would go a long way toward limiting problems with informed consent. This possible solution has several problems, however. First, genetic counseling is expensive and many patients might not have health insurance or the financial means to pay such costs. Genetic counseling for breast cancer without the genetic testing, for instance, costs on

average, more than $200, whereas counseling, testing, and disclosure of results exceeds $3,000.

Second, although the numbers of individuals graduating from counseling programs is increasing, the number of medical geneticists is not. Furthermore, it is not clear that this increase in trained genetic counselors is occurring at the rate necessary to ensure adequate and appropriate levels of support for genetic services in the future.[70]

Third, because of insufficient numbers of professionals specially trained to deal with genetic aspects of health, disease, treatment, and prevention, it is becoming apparent that primary care providers, whether physicians or nurses, will be the ones providing the necessary genetic services to their patients. For these services to be effective, however, primary care providers should have time to provide thorough counseling or to engage in lengthy discussions with patients. They should have knowledge of important genetic disorders, patterns of inheritance, genetic testing procedures, and their availability, as well as the existence of possible therapies and treatments. Practitioners also should have skills in understanding and communicating risk information. Unfortunately, evidence shows that primary care providers have inadequate knowledge of genetics, the nature of inherited disorders, and screening techniques and availability.[71–74] Moreover, primary care providers tend to view genetics as peripheral to everyday clinical concerns. They lack the skills to collect a genetic family history and lack confidence in offering advice and believe that guidelines are lacking to determine when referrals for genetic testing and services are appropriate.[75–77] Given these problems then, it is not at all clear that patients can obtain an adequate disclosure of appropriate information to provide a valid informed consent.

A different ethical concern related to the importance of obtaining informed consent for genetic tests is related to the issue of performing mandatory screening programs. As mentioned earlier, screening programs are directed at particular populations—for example, newborns or pregnant women over a particular age. Some of these screening programs are voluntary and people need to give informed consent for the tests to be performed. At times, however, such programs are mandatory; such is the case, for instance, of newborn screening.

In general, justification for mandatory screening programs for a particular disease or disorder rests on the premise that such a condition if not controlled or detected would adversely affect the individual or even the health of members of the community. This is normally the justification offered for mandatory screening of infectious diseases. In the case of genetic conditions, however, such justification is not available. Nonetheless, most U.S. states have laws making newborn screening mandatory.[78] The original rationale for mandatory policies was the urgent need for early diagnosis of some of the conditions screened—in particular for PKU—and the great benefit to the health of the babies of providing timely treatment. The recent expansion of newborn screening programs to include conditions for which no treatments exist, however, makes omitting parental informed consent controversial.[37, 79]

Justification for screening infants for conditions for which there is no available treatment, or conditions that do not need immediate treatment in the newborn period, or for which the benefits of treatment are not significant cannot be grounded on benefits to the infant but rather on possible benefits to parents or to society. But

given the fact that evidence for the benefits and risks of screening of many of these conditions is not robust, arguments for possible benefits to parents or society might not be particularly compelling either. Moreover, these arguments tend to disregard the effects of false positives on the parents and the child as well as the costs to the health care system. Even if one were to agree that benefits to parents or to society are justified, I argue that this is not sufficient justification to mandate genetic screening. Arguably, screening for conditions that do not present a clear benefit to the newborn should be done on a voluntary basis and with the informed consent of the parents.

PRIVACY AND CONFIDENTIALITY

One of the primary central concerns in relation to genetic tests pertains to the confidentiality of genetic data. Clearly people see the need to secure health-related information as a way to protect their privacy, and they agree that necessary steps need to be taken by institutions and states to protect the confidentiality of such information.[80] But, as mentioned earlier, access to genetic data tends to be seen as more problematic than access to other types of health information because it is thought of as more unique, with more predictive power, immutable through the life span of the person, and with implications for other family members.[15]

There is a valid concern that unauthorized access to this data by insurance companies and employers can result in unjust discrimination against people. Insurance companies could use genetic information to make decisions about health coverage and life and disability insurance, and employers could use such information in making decisions about hiring and promotion. It is unclear whether genetic information has been used to unfairly discriminate among individuals, but it is clear that people are afraid it could be used in this manner.[81, 82] This concern is even more prevalent given the increasing use of electronic health records.[83] Indeed, it was precisely because it was thought that current legal protections for genetic health information were not sufficient that Congress passed the Genetic Information Nondiscrimination Act (GINA) of 2008. This act is intended to extend important protection against discrimination in health insurance and employment. In particular, GINA forbids health insurers from collecting and using an individual's genetic information when making determinations about eligibility and premiums. Similarly, GINA prohibits employers from using such information in making employment decisions.[84]

A variety of state laws to limit access to and the use of genetic information have been enacted. These laws protect such information against unauthorized disclosure. GINA now provides for a consistent national standard and clarifies ambiguities in existing federal law.[85] But at the same time that the GINA is intended to enhance protection against genetic discrimination in health insurance and employment, it also is expected to ease the public fears about genetic discrimination and thus facilitate genetic testing. Physicians may be more willing to offer genetic testing knowing that the results cannot be used for employment or insurance eligibility decisions.[18] Similarly, having confidence that their information will be kept confidential, individuals might be less reluctant to participate in research that collects genetic data, a growing and promising area of public health epidemiology. Of course, it is too soon

to determine what the effect of GINA will be on industry practice and public opinion, whether it will protect against unjust discrimination, and whether it will promote participation in research and encourage responsible genetic testing. GINA, however, does not address issues related to access to other insurance products such as life, disability, long-term care, and mortgage insurance.[18]

A significant ethical concern about GINA is that a strict protection of privacy and confidentiality may endanger legitimate public health needs such as health surveillance and epidemiological research. Public health goals might legitimately call for research that addresses individual aspects of disease and disease prevention with the ultimate goal of improving health outcomes. Protection against discrimination might have the effect of preventing, rather than enhancing, these types of research. Indeed, a significant criticism against the privacy rule included in the Health Insurance Portability and Accountability Act (HIPAA) enacted in 1996, which was designed to safeguard access to, and use of, protected health information, is that it has slowed the research process and complicated basic medical care.[86, 87] Restrictions on legitimate genetic research might result in part from the fact that privacy and confidentiality are best protected by the requirement of informed consent. It often is difficult, however, to ascertain how researchers might use genetic data. Moreover, requirements of consent for genetic research might result in a diminished ability to recruit participants or even create serious obstacles in accessing stored tissue and genetic data sets.[88] Hence, the need to protect privacy and confidentiality, particularly through informed consent requirements, can have a deleterious effect on legitimate public health needs.

CONCLUSION

The so-called genetic revolution has been received by scientists, medical professionals, public policy makers, public health officials, and the general public with great hope. Indeed, the mapping of the human genome has been presented as the first step in solving the medical problems that afflict humans.[89] But even when one tempers the rhetorical excesses about the promises and dangers of the genetic revolution, it seems reasonable to believe that genetic tests could provide significant potential public health benefits. These benefits, however, will not be realized unless we are quite aware of the ethical concerns that these new medical technologies raise and provide assurance that such concerns are addressed adequately. Attention needs to be given to the increase in offerings of genetic tests that lack evidence of analytic validity as well as tests for which scarce evidences exists of clinical validity and utility.

The spread of genetic tests with unproven clinical validity and utility are likely to cause harm to individual patients and to society in general. Moreover, it is important to consider the social context in which new medical technologies are implemented; otherwise, we run the risk of increasing existing health care inequities. Attention to educating the public about genetics, providing a better understanding of risks of genetic testing, and ensuring that patients have access to trained medical professionals also are necessary to ensure that patients can give an adequate informed consent. Similarly, mandatory genetic screening programs need to have a clear justification to omit informed consent. Ensuring that people are not unjustly

discriminated against because of their genetic or health status will require a careful attention to issues of privacy and confidentiality; yet, concerns about privacy will need to be balanced against the legitimate public health needs. Focusing on these ethical concerns when making public-policy decisions about implementation of genetic testing and screening is necessary if we want to use these medical technologies in ways that will advance the public's health.

REFERENCES

1. Watson J, Crick F. A structure for deoxyribose nucleic acid. *Nature*. 1953;171:737–738.
2. Khoury MJ, et al. Do we need genomic research for the prevention of common diseases with environmental causes? *Am J Epidemiol*. 2005;161(9):799–805.
3. Brand A, Brand H, Schulte IN, den Bäumen T. The impact of genetics and genomics on public health. *Eur J Hum Genet*. 2008;16(1):5–13.
4. Agurs-Collins T, et al. Public health genomics: translating obesity genomics research into population health benefits. *Obesity*. 2008;16(Suppl 3):S85–S94.
5. Khoury MJ, et al. A decade of public health genomics in the United States: Centers for Disease Control and Prevention 1997–2007. *Public Health Genomics*. 2008:Epub ahead of print.
6. Moyer VA, et al. United States Preventive Services Task Force: expanding newborn screening: process, policy, and priorities. *Hastings Cent Rep*. 2008;38(3):32–39.
7. Hogarth S, Javitt G, Melzer D. The current landscape for direct-to-consumer genetic testing: legal, ethical, and policy issues. *Annu Rev Genomics Hum Genet*. 2008;9:161–182.
8. Geransar R, Einsiedel E. Evaluating online direct-to-consumer marketing of genetic tests: informed choices or buyers beware? *Genet Test*. 2008;12(1):13–23.
9. Ozanne EM, et al. Identification and management of women at high risk for hereditary breast/ovarian cancer syndrome. *Breast J*. 2009;15(2):155–162.
10. Squitieri F, Cannella M, Frati L. Molecular medicine: predicting and preventing Huntington's disease. *Neurol Sci*. 2008;29(4):205–207.
11. Cram D, Pope A. Preimplantation genetic diagnosis: current and future perspectives. *J Law Med*. 2007;15(1):36–44.
12. Fernhoff PM. Newborn screening for genetic disorders. *Pediatr Clin North Am*. 2009;56(3):505–513.
13. Buckley F, Buckley SJ. Costs of prenatal genetic screening. *Lancet*. 2008; 372(9652):1805.
14. Langlois S, et al. Carrier screening for thalassemia and hemoglobinopathies in Canada. *J Obstet Gynaecol Can*. 2008;30(10):950–971.
15. McGuire AL, et al. Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: points to consider. *Gen Med*. 2008;10:495–499.
16. Rothstein MA. Genetic exceptionalism and legislative pragmatism. *J Law Med Ethics*. 2007;35:59–65.
17. Roche PA, Annas GJ. Protecting genetic privacy. *Nature Rev Genet*. 2001;2:392–396.
18. McGuire AL, Majumder MA. Two cheers for GINA? *Genome Med*. 2009;1(1):6.
19. Evans JP, Burke W. Genetic exceptionalism: too much of a good thing? *Genet Med*. 2008;10(7):500–501.
20. Burke W. Clinical validity and clinical utility of genetic tests. *Curr Protoc Hum Genet*. 2009;9:9–15.

21. Teutsch SM, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: methods of the EGAPP Working Group. *Genet Med*. 2009;11(1):3–14.
22. Zimmern RL, Kroese M. The evaluation of genetic tests. *J Public Health (Oxf)*. 2007;29(3):246–250.
23. Rogowski WH, Grosse SD, Khoury MJ. Challenges of translating genetic tests into clinical and public health practice. *Nat Rev Genet*. 2009:Epub ahead of print.
24. Janssens AC, van Duijn CM. Genome-based prediction of common diseases: methodological considerations for future research. *Genome Med*. 2009;1(2):20.
25. Lavine G. Questions linger over utility of genetic testing. *Am J Health Syst Pharm*. 2008;65(9):793–795.
26. Shen Y, Wu BL. Microarray-based genomic DNA profiling technologies in clinical molecular diagnostics. *Clin Chem*. 2009;55(4):659–669.
27. Buchanan AV, Weiss KM, Fullerton SM. Dissecting complex disease: the quest for the Philosopher's Stone? *Int J Epidemiol*. 2006;35(3):562–571.
28. Barros SP, Offenbacher S. Epigenetics: connecting environment and genotype to phenotype and disease. *J Dent Res*. 2009;88(5):400–408.
29. Shames DS, Minna JD, Gazdar AF. DNA methylation in health, disease, and cancer. *Curr Mol Med*. 2007;7(1):85–102.
30. Moss TJ, Wallrath LL. Connections between epigenetic gene silencing and human disease. *Mutat Res.* 2007;618(1–2):163–174.
31. Henrikson NB, Bowen D, Burke W. Does genomic risk information motivate people to change their behavior? *Genome Med*. 2009;1(4):37.
32. Hunter DJ, Khoury MJ, Drazen JM. Letting the genome out of the bottle: will we get our wish? *N Engl J Med*. 2008;358(2):105–107.
33. McGuire AL, Burke W. An unwelcome side effect of direct-to-consumer personal genome testing: raiding the medical commons. *JAMA*. 2008;300(22):2669–2671.
34. Moses T. Self-labeling and its effects among adolescents diagnosed with mental disorders. *Soc Sci Med*. 2009;68(3):570–578.
35. American College of Medical Genetics. *Newborn Screening: Toward a Uniform Screening Panel and System*. Washington, DC: Health Resources and Services Administration; 2005. ftp://ftp.hrsa.gov/mchb/genetics/screeningnoappendices.pdf.
36. Wilfond BS, Parad RB, Fost N. Balancing benefits and risks for cystic fibrosis newborn screening: implications for policy decisions. *J Pediatr*. 2005;147(3 Suppl):S109–113.
37. Baily MA, Murray TH. Ethics, evidence, and cost in newborn screening. *Hastings Center Report*. 2008;38:23–31.
38. Human Genome Project Information. Gene testing. http://www.ornl.gov/sci/techresources/Human_Genome/medicine/genetest.shtml. Published September 19, 2008.
39. Lawrence WF, et al. Cost of genetic counseling and testing for BRCA1 and BRCA2 breast cancer susceptibility mutations. *Cancer Epidemiol Biomarkers Prev*. 2001;10:475–481.
40. Centers for Disease Control and Prevention. Fact sheet: racial/ethnic health disparities. http://www.cdc.gov/od/oc/media/pressrel/fs040402.htm. Published April 2, 2004. Accessed June 26, 2009.
41. American Heart Association. *Heart Disease and Stroke Statistics: 2009 Update*. Dallas, TX: American Heart Association; 2009. http://www.americanheart.org/downloadable/heart/1240250946756LS-1982%20Heart%20and%20Stroke%20Update.042009.pdf.

42. DeNavas-Walt C, Proctor BD, Smith JC. *US Census Bureau, Current Population Reports, P60-231, Income, Poverty, and Health Insurance Coverage in the United States: 2005*. Washington, DC: US Government Printing Office; 2005.
43. US Department of Labor, Bureau of Labor Statistics. *The Employment Situation: May 2009*. Washington, DC: Bureau of Labor Statistics; 2009. http://www.bls.gov/news.release/pdf/empsit.pdf.
44. Rothstein MA, Hoffman S. Genetic testing, genetic medicine, and managed care. *Wake Forest Law Rev*. 1999;34(3):849–888.
45. Faden R, Beauchamp T. *A History and Theory of Informed Consent*. New York: Oxford University Press; 1986.
46. *Nuremberg Code. Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol. 2, Nuremberg, October 1946–April 1949*. Washington, DC: US Government Printing Office; 1949:181–182.
47. World Medical Association. World Medical Association declaration of Helsinki: ethical principles for medical research involving human subjects. Helsinki, Finland. June 1964. http://www.wma.net/e/policy/b3.htm. Accessed June 15, 2009.
48. Beauchamp T, Childress JF. *Principles of Biomedical Ethics*. 6th ed. New York: Oxford University Press; 2008.
49. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: US Government Printing Office; 1978.
50. Molster C, Charles T, Samanek A, O'Leary P. Australian study on public knowledge of human genetics and health. *Public Health Genomics*. 2009;12(2):84–91.
51. Shaw A, Hurst JA. "What is this genetics, anyway?" understandings of genetics, illness causality and inheritance among British Pakistani users of genetic services. *J Genet Couns*. 2008;17(4):373–383
52. Frazier L, Calvin AO, Mudd GT, Cohen MZ. Understanding of genetics among older adults. *J Nurs Scholarsh*. 2006;38(2):126–132.
53. Catz DS, et al. Attitudes about genetics in underserved, culturally diverse populations. *Community Genet*. 2005;8(3):161–172.
54. Lanie AD, et al. Exploring the public understanding of basic genetic concepts. *J Genet Couns*. 2004;13(4):305–320.
55. Richards MP. Lay understanding of Mendelian genetics. *Endeavour*. 1998;22(3):93–94.
56. Morren M, Rijken M, Baanders AN, Bensing J. Perceived genetic knowledge, attitudes towards genetic testing, and the relationship between these among patients with a chronic disease. *Patient Educ Couns*. 2007;65(2):197–204.
57. MacDonald DJ, Sarna L, Uman GC, Grant M, Weitzel JN. Health beliefs of women with and without breast cancer seeking genetic cancer risk assessment. *Cancer Nurs*. 2005;28(5):372–379.
58. Katapodi MC, Dodd MJ, Lee KA, Facione NC. Underestimation of breast cancer risk: influence on screening behavior. *Oncol Nurs Forum*. 2009;36(3):306–314.
59. Dearborn JL, McCullough LD. Perception of risk and knowledge of risk factors in women at high risk for stroke. *Stroke*. 2009;40(4):1181–1186.
60. Burke-Doe A, Hudson A, Werth H, Riordan DG. Knowledge of osteoporosis risk factors and prevalence of risk factors for osteoporosis, falls, and fracture in functionally independent older adults. *J Geriatr Phys Ther*. 2008;31(1):11–17.
61. Nicol MB, Thrift AG. Knowledge of risk factors and warning signs of stroke. *Vasc Health Risk Manag*. 2005;1(2):137–147.

62. Visschers VH, Meertens RM, Passchier WW, de Vries NN. Probability information in risk communication: a review of the research literature. *Risk Anal*. 2009; 29(2):267–287.
63. Kahneman D, Slovic P, Tversky A, eds. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press; 1974.
64. Knäuper B, et al. Motivation influenced the underestimation of cumulative risk. *Pers Soc Psychol Bull*. 2005;31:1511–1523.
65. Wildavsky A, Dake K. Theories of risk perception: who fears what and why? *Daedalus.* 1990;114:41–60.
66. Politi MC, Han PK, Col NF. Communicating the uncertainty of harms and benefits of medical interventions. *Med Decis Making*. 2007;27(5):681–695.
67. O'Doherty K, Suthers GK. Risky communication: pitfalls in counseling about risk, and how to avoid them. *J Genet Couns*. 2007;16(4):409–417.
68. Vuckovic N, Harris EL, Valanis B, Stewart B. Consumer knowledge and opinions of genetic testing for breast cancer risk. *Am J Obstet Gynecol*. 2003;189(4 Suppl):S48–S53.
69. Gordon C, Walpole I, Zubrick SR, Bower C. Population screening for cystic fibrosis: knowledge and emotional consequences 18 months later. *Am J Med Genet A*. 2003;120A(2):199–208.
70. Goodwin K. Genetic specialists trailing research boom. *State Legis*. 2009;35(2):33.
71. McInerney JD. Genetics education for health professionals: a context. *J Genet Couns*. 2008;17(2):145–151.
72. Telner DE, Carroll JC, Talbot Y. Genetics education in medical school: a qualitative study exploring educational experiences and needs. *Med Teach*. 2008;30(2): 192–198.
73. Harvey EK, Fogel CE, Peyrot M, Christensen KD, Terry SF, McInerney JD. Providers' knowledge of genetics: a survey of 5915 individuals and families with genetic conditions. *Genet Med*. 2007;9(5):259–267.
74. Maradiegue A, Edwards QT, Seibert D, Macri C, Sitzer L. Knowledge, perceptions, and attitudes of advanced practice nursing students regarding medical genetics. *J Am Acad Nurse Pract*. 2005;17(11):472–479.
75. Guttmacher AE, Porteous ME, McInerney JD. Educating health-care professionals about genetics and genomics. *Nat Rev Genet.* 2007;8(2):151–157.
76. Guttmacher AE, Collins FS, Carmona RH. The family history: more important than ever. *N Engl J Med.* 2004;351(22):2333–2336.
77. Emery J, Hayflick S. The challenge of integrating genetic medicine into primary care. *BMJ*. 2001;322(7293):1027–1030.
78. National Conference of State Legislatures. Newborn genetic and metabolic disease screening. http://www.ncsl.org/IssuesResearch/Health/NewbornGeneticandMetabolicScreeningLaws/tabid/14416/Default.aspx. Published 2009.
79. President's Council on Bioethics. *The Changing Moral Focus of Newborn Screening: An Ethical Analysis by the President's Council on Bioethics*. Washington, DC: US Government Printing Office; 2008.
80. Page SA, Mitchell I. Patients' opinions on privacy, consent and the disclosure of health information for medical research. *Chronic Dis Can*. 2006;27(2):60–67.
81. Lowstuter KJ, et al. Influence of genetic discrimination perceptions and knowledge on cancer genetics referral practice among clinicians. *Genet Med.* 2008;10(9): 691–698.
82. Oster E, et al. Fear of health insurance loss among individuals at risk for Huntington disease. *Am J Med Genet A*. 2008;146A(16):2070–2077.

83. Ray P, Wimalasiri J. The need for technical solutions for maintaining the privacy of EHR. C*onf Proc IEEE Eng Med Biol Soc* 2006;1:4686–4689.
84. Genetic Information Nondiscrimination Act of 2008. http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_bills&docid=f:h493enr.txt.pdf. Published May 2008.
85. Tan MH. Advancing civil rights, the next generation: the Genetic Information Nondiscrimination Act of 2008 and beyond. *Health Matrix Clevel*. 2009;19(1):63–119.
86. Institute of Medicine. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. Washington, DC: National Academies Press; 2009.
87. Kumar S, Henseler A, Haukaas D. HIPAA's effects on US healthcare. *Int J Health Care Qual Assur*. 2009;22(2):183–197
88. Gostin LO, Nass S. Reforming the HIPAA privacy rule: safeguarding privacy and promoting research. *JAMA*. 2009;301(13):1373–1375.
89. Collins F, McKusick V. Implications of the Human Genome Project for medical science. *JAMA*. 2001;285(5):540–544.

CHAPTER 8

Palliative Care

Roma Tickoo, MD, MPH, and Paul Glare, MD

INTRODUCTION

As chronic diseases, including cancer, surpass infectious diseases as the primary causes of death, and as individuals, thanks to advances in treatment of these diseases, are living longer with their diseases, providing timely access to consistently high-quality end-of-life care has become an important international issue. The World Health Organization (WHO) defines palliative care as "an approach which improves the quality of life of patients and their families facing life threatening illness, through the prevention and relief of suffering by means of early identification and impeccable assessment and treatment of pain and other problems, physical psychosocial and spiritual."[1] Indeed, the WHO has recognized palliative care as an important clinical and humanitarian need, particularly in places where a high proportion of patients present in advanced stages of disease usually with little chance of cure. It estimates that the quality of life of at least 100 million people globally would be improved if knowledge of 21st-century palliative care was accessible to everyone.[2]

The Center to Advance Palliative Care reported in 2008 that nearly 90 million Americans were living with serious and life-limiting illness.[3] This number is expected to more than double over the next 25 years as the population ages. Globally, the WHO has estimated that the number of global cancer deaths, for example, is projected to increase 45 percent from 2007 to 2030 (7.9 million to 11.5 million) primarily because of the increase in the aging population. Although this estimate takes into account expected declines in death rates for some cancers in developed countries, new cases of cancer in the same period are estimated to increase from 11.3 million in 2007 to 15.5 million in 2030.[4] This chapter provides an overview of the palliative care and hospice movement and explains why this is an important medical and public health issue.

WHY PALLIATIVE CARE?

Many patients and caregivers express a helpless feeling when facing a life-limiting illness. Often they have to make difficult decisions regarding the course of care, including but not limited to quality versus quantity of life. These decisions require careful communication and consideration. The Canadian Australian ethicist and academic Margaret Somerville has said "the ethical and legal tone of a society can be judged by how it treats its weakest, neediest and most vulnerable members."[5] Clearly addressing end-of-life care is an important medical, ethical, and public health issue. Yet, despite the significant advances in medical care, far less had been done to address the medical, social, and psychological needs of the dying patient. This situation is not limited to one country; rather, it is a universal issue that begs for attention. Studies had shown that most people living with a terminally fatal illness experience inadequately treated symptoms, fragmented care, poor communication with their doctors, and enormous strains on their family caregivers.[6–9] Palliative care was developed in response to these inadequacies.

Pain management is an integral component of palliative care. Some have gone so far as to call pain relief "a basic human right."[10] A growing number of statements and initiatives on the necessity for pain management reflect a "call to arms" based on three propositions. First, pain, whether acute or chronic, is inadequately addressed for a variety of cultural, educational, political, religious, and logistical reasons. Second, inadequately treated pain has major physiological, psychological, economic, and social ramifications for patients, their families, and for society. Third, it is within the capacity of all nations to significantly improve the treatment and management of pain. Indeed there is an emerging international consensus that "unreasonable failure" to treat pain is not only poor medicine, but also unethical practice and an abrogation of a fundamental human right.

Pain causes terrible suffering, yet there are inexpensive, safe, and effective medications that generally are straightforward to administer. Furthermore, international law obliges countries to make adequate pain medications available. Over the last 20 years, the WHO and the International Narcotics Control Board, the body that monitors the implementation of the United Nations drug conventions, repeatedly have reminded nations of their obligations. Unfortunately, however, little progress has been made in many countries.

Under international human rights law, governments must take steps to ensure that individuals have adequate access to treatment for their pain. At a minimum, states must ensure availability of morphine, the mainstay medication for the treatment of moderate to severe pain. Morphine is considered an essential medicine that should be available to all persons who need it. It is relatively inexpensive and widely available. Failure to make essential medicines available or to take reasonable steps to make pain management and palliative care services available might result in a violation of the right to health. In some cases, failure to ensure that patients have access to treatment for severe pain even could be construed as a violation of the prohibition of "cruel, inhuman, or degrading treatments" under Article 5 of the Universal Declaration of Human Rights.[10]

Unrelieved pain is not the only ethical imperative driving the need for access to palliative care. Many deficiencies in broader end-of-life care also demand a response.

For instance, in the early 1990s the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment (SUPPORT) provided for the first time hard data on the problems with the care of the dying in modern hospitals.[11] Not only was pain relief shown to be inadequate, but patients' preferences for their terminal care also were ignored and their advanced directives ineffective. SUPPORT provided the impetus for the Institute of Medicine (IOM) to examine the systemwide deficiencies in end-of-life care as described in its 1997 report, *Approaching Death*.[12] The third of the seven recommendations made in this report called for systems of care to change to support the actions of individuals.

In the current context of ever-increasing patient autonomy, the sad state of terminal care has driven calls for physician-assisted suicide and voluntary euthanasia to be legalized. Fortunately not everybody sees this as an acceptable solution. In their 2002 book, *The Case against Assisted Suicide: For the Right to End-of-Life Care,* Foley and Hendin highlight the challenges to create a culture that identifies the care of the seriously ill and dying as a public health issue.[13] Neither autonomy nor compassion—the major justifications for assisted suicide—provides an adequate basis for legalizing these practices. In particular, patient autonomy is illusory when physicians do not know how to assess and treat suffering, making the only "choice" for patients either continued agony or hastened death. But, as the authors emphasize, many changes are required to American health care to ensure that patients really do have a choice. Reforming the Medicare system is advocated to allow functional disability and severity of illness to be the criteria for eligibility for the Hospice Benefit, not, as is the case, a prognosis of six months. An open public discussion with state and local officials as well as with professionals to address what we as a society can do to promote improved care of patients with serious life-threatening illness is needed.

WHAT IS HOSPICE AND PALLIATIVE CARE?

Hospice and palliative care is a philosophy of care that focuses on the palliation of a terminally ill patient. For much of its history, hospice and palliative care focused on caring for terminally ill cancer patients. Often referred to as "palliative" or "supportive" care, hospice care emphasized the management of pain and discomfort and the emotional support of the patient and family. Over the past two decades, however, noncancer diseases, such as congestive heart failure, emphysema, and Alzheimer's disease have accounted for an increasing proportion of hospice referrals. Care is rendered either on an inpatient basis or in the home setting. Palliative care concerns itself with the relief of suffering in patients with a life-limiting illness that is not amenable to cure. Palliative care is a philosophy of care that differs from mainstream health care in that it does not concern itself with *the cure of diseases* rather with *the needs of the dying.* That is, palliative medicine, a relatively new medical subspecialty,[14] focuses on improving the quality of remaining life for terminally ill patients and their family. For the purposes of this chapter, unless otherwise specified, the terms palliative care and hospice will be used interchangeably, according to the British and WHO concepts.

Palliative care, characterized by interdisciplinary teams working to meet the multidimensional needs of the dying, slowly has been integrated into the health care delivery system. This area of medicine focuses on comprehensive symptom management,

coordination of care, as well as on intensive communication with the patient and the family regarding the emotional aspects of dealing with a terminal illness. The presumption of palliative care is that patients and their loved ones would benefit by a coordinated approach to end-of-life care, including pain management and a focus on patient-physician-family communication. An important outcome is that the patient and the family would be gain a better understanding of the prognosis of the disease and theoretically would be more satisfied with the care received.

The U.S. National Cancer Control Network defines palliative care as both a philosophy of care and an organized highly structured system for delivering care to persons with life-threatening or debilitating illnesses.[15] Palliative care has been defined by the WHO as an approach that improves the quality of life of patients and their families facing the problems associated with life-threatening illness, through the prevention and relief of suffering by means of early identification and impeccable assessment, treatment of pain and other problems, including physical, psychosocial, and spiritual.[1] According to the WHO, palliative care provides relief from pain and other distressing symptoms; integrates the physical, emotional, and spiritual aspects of patient care; offers a support system to help the family cope during the patient's illness; uses a team approach to address the needs of patients and families, including grief counseling and support; enhances quality of life and can positively influence the course of illness; and can be offered with other therapies that are intended to prolong life—for example, chemotherapy or radiation therapy.

In the United States, there is a unique distinction between palliative care and hospice care that primarily is driven by reimbursement issues.[16] Both place a strong emphasis on reducing pain and discomfort. Whereas palliative care is provided as part of acute care for patients with a life-threatening illness who still wish to pursue curative treatments, hospice care is specifically developed to enable patients who have acknowledged that they are dying to forego aggressive, curative treatments in lieu of receiving compassionate terminal care in their home or in a hospice facility. In other countries, especially Britain, the word "hospice" refers to an inpatient unit geographically separate from a hospital where dying patients are admitted to receive palliative or terminal care.

THE ORIGINS OF HOSPICE AND PALLIATIVE CARE MOVEMENT

The word "hospice" traces back to medieval times, when it was used to describe charitable institutions that welcomed weary travelers and pilgrims.[17] In the mid-19th century, the word was revived and adopted for residential facilities established for the care of the dying poor.[17] Such buildings existed in Britain, Australia and France, and continued to operate up until the end of World War II. In one sense, these facilities were really "proto-hospices" because they were rooted in the religious and philanthropic ideologies of the Victorian era, providing nursing and spiritual care but limited medical service,[18] but doing little to impact on the care of the dying generally.[17]

After World War II, there was a revolution in theory and practice in the care of the dying. Modern palliative care began in Britain in the 1950s and 1960s, primarily in response to the observation that resources for the care of patients dying from cancer in National Health Service (NHS) hospitals were minimal. The NHS focus was

on acute care and rehabilitation and, at the same time, it had an ideological rejection of charity as a source for the provision of health care.[18–19] But several other issues were driving the development of hospice care and palliative care, as summarized by Clark.[18] These include such things as demographic changes in postwar Britain, with an aging in population and social debate about how willing families were to care for the dying in the newly constructed postwar housing. The 1960 report *Peace at Last* was highly critical of the care provided by the old "homes of the dying."[20] An awakening medical interest in pain, geriatrics, chronic illness, and the care of the dying subsequently led to the palliative care movement.[21]

Cecily Saunders, a physician who was also qualified as a nurse and social worker, is generally acknowledged to have been the driving force behind the development of the modern palliative care movement in Britain in the 1950s and 1960s.[18] Her commitment to care for the dying was driven by a personal vocation, continuing the Christian influence in modern hospice that had driven the proto-hospices decades earlier. Her work combined medical innovations in pain relief and symptom management with wider concerns of the practical and social needs of the patient as well as their spiritual needs as they approached death. The residential facility remained the focus for Dr. Saunders for delivering this new kind of comprehensive care for the dying. These residential facilities helped the palliative care concept to spread with the "bricks and mortar" of a new building being an obvious focal point for community fundraising. There was a rapid spread of these kinds of facilities across Britain over the next two decades, and by 1985, there were more than 100 such inpatient hospices. In the 1990s, the principles and practices of the inpatient hospice unit were modified for other settings, which not only included taking hospice principles into the home by means of a community team, but also into the NHS hospital via the hospital inpatient unit and consult service. By 2006, Britain had 193 hospices or other specialist inpatient units providing 2,774 beds, 20 percent of which were operated by the NHS, 314 hospital palliative care teams, 295 home care teams, and 294 day care services.[22]

In Britain, palliative care and hospice form a continuum of care. Although there has been a rapid development of palliative care, much of it occurred with little regard as to whether needs were being met by these new services,[23] although 21st-century government policy very much drives service development in Britain. (We refer to the National Institute for Clinical Excellence's guidance on cancer services: improving supportive and palliative care for adults with cancer).[24]

As the hospice concept took hold and spread around the world, it evolved to suit local conditions. The basic principles can be interpreted in widely differing cultures and with few resources other than the family values of the developing world.[17] In many countries, especially those belonging to the British Commonwealth like Australia and India, inpatient hospices similar to the British ones were established and hospital and community palliative care teams were created as well. At a national level, Australia and India are both global examples of excellent palliative care programs.

The history of Australian palliative care can be traced back to the 19th century with the arrival of the Sisters of Charity nuns who came to Sydney from Ireland to care for their compatriot female convicts who were dying in prison from tuberculosis (TB) and other infectious diseases. The Sisters of Charity founded St. Joseph's

hospital in Auburn, which continued their mission of the care of the dying. Soon thereafter they founded Australia's first hospice, Sacred Heart Hospice, in Darlinghurst (Sydney) in 1890. Although run by nursing nuns, medical input was provided by family physicians.

There were two main drivers to the development of modern palliative care in Australia in the latter part of the 20th century: clinical and political. Visiting medical advocates of the palliative care movement had an impact on local physicians who became interested in advances in pain management. Politically, in 1972, Australia adopted socialized and nationalized health care. As the cost of hospital beds increased, the focus shifted reducing the length of stay of hospitalized patients, including palliative care.[25] The Australian government made community-based palliative care a priority, aiming to reduce the number of patients dying in hospital. To meet the palliative care needs of the Australian population (approximately 20 million citizens of whom some 100,000 die from diseases amenable to palliative care),[26] the federal government gives $180 million to the states to pay for palliative care services. To improve Australian palliative care further, in 2000, the Australian federal and state governments joined together with service providers and nongovernmental organizations (NGOs) interested in palliative care to formulate a framework for improving awareness of and access to quality palliative care programs.[27] An additional $60 million was made available to fund new initiatives to fill gaps in current services during the first three-year period of the program. The key platforms in the National Palliative Care strategy included support for patients, families, and caregivers throughout the community; increased access to palliative care medicines in the community; education, training, and support for the workforce; research and quality improvement for palliative care services; and methods for ensuring that these policies are being achieved or are under development as part of the program.

The evolution of modern hospice and palliative care in India began in 1986 when Dr. Louis DeSouza, a surgical oncologist, founded the first Western-style Indian Hospice, Shanti Avedna, in Mumbai, which was inspired by St. Christopher's Hospice in London. In 1993, the NGO Pain and Palliative Care Society (PPS) was founded in Calicut in the state of Kerala. Furthermore, a network of community initiatives called Neighborhood Network in Palliative Care (NNPC) was established. More than 60 units now cover a population of more than 12 million. This is probably the largest community-owned palliative care network in the world. A volunteer force is the main strength of the NNPC program. Trained volunteers work with health care professionals to provide psychological and spiritual care.

In 1996, the "CanSupport" initiative was established in Delhi. This initiative was based on the personal experience of a cancer survivor, Harmala Gupta, and its mission was to offer support to other cancer patients in India. Gupta's vision and efforts date to 1991 when she started the first cancer support group in India. She along with other cancer survivors and caregivers visited cancer clinics to provide information, assistance, and an empathetic listening ear to the dying and their family.

PALLIATIVE CARE MOVEMENT IN THE UNITED STATES

Although Calvary Hospital in New York had been established in 1899 in the image of one of the early European hospices,[17] in the United States, hospice came to

represent home-based rather than institution-based care of the dying. This trans-Atlantic variance has been attributed to the consumerism movement in the 1970s that was concerned with a rediscovery of "the natural," a rising interest in thanatological subjects, and a reaction against the medicalization of death.[18] Primarily inspired by the work of Dr. Cecily Saunders, hospice in the United States was a grassroots movement outside the medical mainstream. U.S. hospice care initially was designed for cancer patients with a functional family support system and a home at which they could be cared for away from the high-tech hospital environment.

The first U.S. hospice program, the Connecticut Hospice, opened in 1974 funded by a three-year grant from the National Cancer Institute. A decade later, findings from the National Hospice Organization Study provided the impetus for the expansion of the hospice movement. The study showed that hospice was more effective and less expensive than conventional terminal care.[28] The results of the NHO study led to the recognition of hospice care by the U.S. government. The Medicare Hospice Benefit legislation legitimized end-of-life care making hospice equivalent to other reimbursable medical services. For the first time, care of the dying was institutionalized and publicly supported. Signing on for hospice care enabled a patient to receive services not normally paid for, including provision of all medical drugs and devices necessary for palliative care. Other services provided free under the benefit include professional nursing care, personal assistance with activities of daily living, various forms of rehabilitation therapy, dietary counseling, psychological and spiritual counseling for both patient and family, volunteer services, respite care, and bereavement services for the family for up to one year after the patient's death.[29] This care is provided by a multidisciplinary team under the management of a physician who may be the patient's primary physician or else a hospice physician. To qualify for this range of services, however, patients had to have a prognosis of less than six months and agree to forgo life-prolonging therapies.[30] Although this stipulation did not pose a problem when dealing with terminally ill cancer patients, patients with noncancer diagnoses, now accounting for a large part of hospice care, often lived much longer than six months. These requirements eventually led to a separation of hospice care from palliative care in the United States that is not found in other countries.

Although the hospice movement has expanded substantially, home hospice care is acknowledged to be underutilized. The reasons for this are manifold: Physicians find it difficult to prognosticate that the patient has less than six months to live,[31] or believe that hospice referral means admitting medical defeat. Patients and families have difficulty giving up hope of recovery, forgoing active treatment, or accepting the inevitability death.[32] Many find it challenging to make the abrupt cognitive and emotional shifts in transitioning from acute care to hospice care as they enter a completely new system of care that is geographically and ideologically isolated from traditional and familiar health care systems that frequently separates patients from their primary care providers. That being said, substantial progress has been made since Meier et al. identified these kind of issues and the wide range of changes needed in education and training, medication availability, and health care system reorganization needed to improve palliative care in the United States.[33] But, much work remains to be done.

Palliative care in the United States is a broad term covering all forms of prevention and treatment and suffering regardless of the diagnosis, treatment, or prognosis

in a patient with a life-threatening illness even if it is potentially curable, whereas hospice is more narrowly defined as the subset of palliative care especially catering to those near death. American palliative care is a more recent development that aims to bring many of the principles of hospice, especially symptom control and attention to psychosocial and spiritual issues, to hospitalized patients while they are still receiving active treatment but without the comprehensive range of services that hospice provides,[16, 34] and has really only emerged since the 1990s.

PALLIATIVE CARE AND PUBLIC HEALTH: CURRENT STATUS AND FUTURE DIRECTIONS

Although hospice and palliative care largely began on both sides of the Atlantic as a grassroots movement, over the past 10 to 15 years, it is increasingly being viewed as a public health issue.[1–2, 23, 35–39] Advances in the treatment and management of acute and chronic diseases has led to a dramatic increase in life expectancy and a concomitant increase in the incidence and prevalence of cases of progressive, life-limiting illness. This, combined with ethical imperative to improve pain and relieve suffering in these patients, is behind the push for an expansion in palliative care services.

Using the language of the IOM's 1988 report on the future of public health,[40] one could make the cases that the physical, psychosocial and spiritual needs of dying patients and their families also falls within the "mission" and "substance" of public health and that delivery of palliative care services should fall within public health's "organizational framework." As such, there is a need for a standardized means to assess and monitor pain and suffering among those with advanced progressive terminal disease. There is a need for the formulation of palliative care policy in collaboration with community and government leaders, and there is a need to ensure that all in need have access to appropriate and cost-effective palliative care. Certainly there also must be protocols for the evaluation of effectiveness of palliative care.

The United States has done reasonably well in making palliative care part of public health. In 1983, the United States was the first nation to legislate the hospice piece of palliative care as part of mainstream medicine. The Medicare Hospice Benefit acknowledged the importance of offering hospice care to Medicare beneficiaries. Since then, there has been growing enrollment in hospice care; between 2000 and 2004, the percentage of Medicare decedents enrolling in home hospice programs increased from by almost 50 percent from 500,000 to 800,000. In 2005, more than 1.2 million Americas received hospice care.[29] In 2006, Medicare spent close to $10 billion on hospice (up from $2 billion eight years earlier), but at the same time it saved approximately $2,500 per decedent, and this number could have been closer to $7,000 with more timely referrals.[41] Evaluation studies show consistently high family satisfaction rates.[42–43] Although impressive, much more needs to be done in this area. It is estimated that at least two-thirds of deaths are amenable to palliative care,[26] yet only one-third die on hospice, thus dramatically illustrating the underuse of palliative or hospice care. Many more evaluations and cost-effectiveness studies of palliative care need to be undertaken.[44]

The development of palliative care programs, which bring hospice principles to patients ineligible for the hospice benefit because they have a "better" prognosis, has

been slow. Although over the last five years palliative care programs have been implemented in U.S. hospitals,[16, 34] wide geographic variation exists making unequal access to such care a continued reality. The American Hospital Association (AHA) tracks hospital palliative care programs using its Annual Survey Database™. Along with the other 850 elements of the survey, the AHA queries all hospitals (AHA members and nonmembers) as to the presence of a palliative care program, defined as

> an organized program providing specialized medical care, drugs or therapies for the management of acute or chronic pain and/or the control of symptoms administered by specially trained physicians and other clinicians, and supportive care services such as counseling on advance directives, spiritual care, and social services to patients with an advanced disease and their families.

These results have been reported by the Center to Advance Palliative Care (CAPC) in 2001, 2005, and 2008,[45–47] in the form of a "report card" scored for the percentage of hospitals in each state with at least 50 beds reporting a palliative care program. An A is awarded to states with palliative care programs in more than 80 percent of its hospitals, a B for palliative care offered in 61 to 80 percent hospitals, a C for palliative care offered in 41 to 60 percent hospitals, a D for palliative care offered in 21 to 40 percent hospitals, and an F if less than 20 percent hospitals in a state offer palliative care.

In its 2008 report, CAPC, utilizing the AHA data from 2006, found that more than 50 percent of hospitals nationwide had a palliative care program (i.e., an overall national grade of C), indicating a need for significant improvement in access to hospital palliative care.[48] Although half of the 50 states received a grade of A or B (only three states scored an A: Vermont, New Hampshire, and Montana), almost 40 percent get a grade of C, and the remainder received unacceptable grades of D and F (see appendix 8.A). While much remains to be done, just 10 years ago almost no palliative care programs were offered in U.S. hospitals. In the last five years alone, access to palliative care in American hospitals has more than doubled.

The CAPC report utilized data from the Association of American Medical Colleges, the U.S. Census, and the Dartmouth Atlas of Healthcare to examine geographic variations at the state level. Patient access to board-certified palliative medicine physicians, medical student access to clinical training in palliative medicine, and physician access to specialty-level training in palliative medicine are quantified. Findings showed that despite high access in some states, overall wide geographic variation still remains a barrier to care for many patients and families. In general, the availability of palliative care is lower in the southern states compared with the northern states. African Americans and other minorities tend to underutilize palliative care, even when they have access to it.[49] The reasons are largely speculative at this stage, but they likely are multifactorial and include historical, social, cultural, ethical, economic, legal, health policy, and medical ones. Additionally, access is lower in public and sole community-provider hospitals. Particularly worrying for the future is the fact that not all medical schools are presently affiliated with a hospital providing palliative care.

An equally serious problem is that patients and their families often are not offered palliative care as an option, and even when such a service exists and is

offered, its value is discounted.[50] A recent telephone survey of 1,200 Australians representative of the general population found they most of the respondents had little understanding of palliative care.[51] Admittedly 80 percent of respondents were somewhat aware of palliative care at some level, but generally not well enough to be able to explain the concept to someone else. Respondents who knew more about palliative care had learned about it through during the care of a terminally ill family member or friend; 90 percent of them reported a positive view about palliative care and wished that they had had known about it or had their caregiver suggest it earlier. This study shows that if palliative care is not offered as an option, or discounted in value when it is offered, then it is unlikely that patients and families will request it.[50] Although no data are available for the United States, it would be most surprising if results were different from the findings from the Australia study. Raising community awareness of palliative care therefore is an important medical and public health worldwide.

Much more is known about clinicians' knowledge, skills, and attitudes toward palliative care. Multiple surveys have been done over the past 15 years, which repeatedly document deficiencies in both the medical management and psychosocial care of pain, end-of-life issues, and palliative care.[33, 52, 53] This body of data was summarized by the IOM in its 1997 report, *Approaching Death*.[12] Data from the Britain, Europe, and the United States show that up to 80 percent of dying patients are in pain, often severe, and that other symptoms like fatigue, nausea, anorexia, and breathlessness also occurred in 40 to 80 percent of cases.[54–56] Other findings also showed that undertreatment of these symptoms persist even when effective treatments are available.[57–59] Clearly, education and training of professionals must be a fundamental strategy of any public health approach to palliative care. Presently, health care professional initiatives in the area of palliative care include the American Medical Association's Educational Program in End of Life Care (EPEC) for physicians and for nurses (ELNEC) and social workers; the American College of Physicians and American Board of Internal Medicine have initiatives to enhance physicians' competency in end-of-life care and faculty development programs. Each seeks to adopt a consistent set of core principles.

In the United States, improving clinicians' knowledge, skills, and attitudes toward palliative care can succeed only if there are concomitant changes in the health care delivery system. In what would be the first federal legislation addressing comprehensive palliative care, Senators Rockefeller (D-W.Va.) and Collins (R-Maine) introduced a bill, Advance Planning and Compassionate Care Act of 2009, whose main points include workforce (physician and nurse practitioner) adequacy and loan forgiveness; National Service Corps; curricular changes; development of provider reimbursement for conversations about goals of care, and in particular support for completion of orders for life-sustaining treatment in appropriate patient populations; ensuring access to concurrent curative and hospice care for children; incentives for hospital and nursing home delivery of quality palliative care; establishment of a National Center on Palliative Care within the National Institutes of Health (NIH) as a mechanism to ensure adequate attention to the evidence necessary to deliver the highest quality of care; and an ongoing National Mortality Followback Survey to ensure a process of continuous improvement in the quality of care we deliver to this most vulnerable and needy of patient populations.

Organizational and structural characteristics of a health care delivery system can promote or inhibit palliative care for persons who are facing death.[33] Some of these organizational issues include the following:

- The difference between the United States and most other Western countries is that U.S. health care is an industry not a system,[60] and the government's role has been to regulate rather than make policy on the delivery of services.
- Cost-containment in medicine may inhibit palliative care as its patients often have long hospital stays in a highly labor-intensive, expensive clinical setting. Such patients require frequent adjustments in pain medication as well as careful management of dyspnea, delirium, or agitation. Provision of such care often leads to conflict with utilization review requirements under the current U.S. health care reimbursement system.
- Laws regulating opioid prescribing patterns can inhibit access by requiring the use of multiple-copy prescription forms, limitations on the number of tablets per prescription, and regulations for reporting on "habitual users" that stigmatize patients who need opioids and use them appropriately. Insurance companies the restrict usage of some drugs and pharmacies that refuse to stock them also contribute to this inhibition.
- The risk management approach adhered to by many hospitals views in-hospital death as an indicator of substandard delivery of health care, thus creating an incentive for medical staff.

A number of ongoing international initiatives are advocating for the integration of palliative care into national health programs. The WHO has been providing leadership in cancer pain management and palliative care for more than 20 years and advocates a public health model of policy, context, and outcomes that is based on principles of drug availability, education, and implementation.[61] In 2005, the World Health Assembly urged member states to ensure the medical availability of morphine and other opioid analgesic, and requested the WHO director general to explore mechanisms for funding cancer prevention, control, and palliative care, especially in developing countries, and to examine with the International Narcotics Control Board how to facilitate the adequate treatment of pain using opioids. The WHO has established an office for controlled substances as essential medicines, and works with such groups as the Pain and Policy Studies Group at the University of Wisconsin to develop methods and resources to achieve these changes.[62] The International Union against Cancer (UICC) supports the Global Access to Pain Relief Initiative and also included pain relief and palliative care among its immediate targets and 2020 goals in its Draft World Cancer Declaration released in 2008.

In 2006, the International Association for Hospice and Palliative Care developed a list of 34 essential medicines,[63] 14 of which already are included on the WHO Model Drugs list. Other initiatives focusing on cancer pain relief and palliative care include the International Atomic Energy Agency's Global Cancer Control Alliance to Advance Country Strategy and Action Plans; the Open Society Institute's Pain Policy Fellowship; the International Children's Palliative Care Network global advocacy agenda, and the Palliative Care as a Human Rights Initiative.[64] The International Observatory on End of Life Care (IOELC) aims to provide research-based information on palliative care in the global context. The IOELC's useful Web site (www.eolc-observatory.net) provides information on what is happening on palliative care in Eastern Europe and Central Asia.[65]

Some notable developments in the United States over the past decade include the following:

- Establishing pain as the "5th vital sign" by the Veterans Administration Health Care system in 1998;[66]
- Establishing of an International Classification of Diseases (ICD)-9 code for palliative care;[67]
- Centers for Medicare and Medicaid Services' recognition of palliative care as a subspecialty in 2004, with its own code, effective October 1, 2009;[68]
- The IOM's Report, *Improving Palliative Care in Cancer*, which highlighted the need for palliative care, especially for cancer patients;[69]
- National Quality Forum's 2008 recommendations on operational characteristics of quality palliative care programs;[70]
- The establishment of CAPC, and the recommendations associated with its Report Card;[48]
- State initiatives, such as the New York attorney general's 1998 initiative to overcome barriers to good quality end-of-life care, which included ending the regulation requiring triplicate prescribing for controlled substances and including palliative care in the curricula of the 14 New York state medical schools; and
- The American Society of Clinical Oncology's 2009 special report calling for the full integration of palliative care into comprehensive cancer care by 2020.[71]

QUALITY EFFECTIVENESS AND PALLIATIVE CARE

Palliative care services should be assessed and measured to ensure that high-quality services are being provided.[72] According to the IOM,[73] health care services, including palliative care, need to be safe, effective, patient centered, timely, efficient, and equitable. In 2006, the National Quality Forum (NQF), America's major public-private partnership organization charged with advancing the quality of health care, identified a clear set of 38 practices associated with quality palliative care, and these are presented under the IOM's quality headings.[70]

Safe

Palliative care should avoid injuries to patients from care that is intended to help them. The management of pain and other physical symptoms and psychosocial distress should adhere to opioid standards, order sets, and adverse event reporting standards. At present, the evidence base for many palliative care interventions is weak. For example, a systematic review of the WHO analgesic "ladder" approach to cancer pain management, which is based on escalating the potency of analgesics from nonopioids like aspirin or acetaminophen to strong opioids like morphine depending on the severity of pain,[74] found that despite the fact that the WHO has been promulgating the ladder for more than 20 years, the evidence for its effectiveness is meager.[75] The review found that only eight studies involving some 500 patients had been done; they all had methodological weaknesses and the results could not be combined into a meta-analysis. Some of the methodological weakness included small sample size, short-term follow-up, and variations in measuring pain. Similarly, a systematic review of morphine for cancer pain showed that despite the

widespread use of morphine, there are no large-scale, well-designed studies.[76] The same applies to systematic reviews of more complex palliative care interventions.[77–78]

Effective

The NQF recommends that palliative care services should undertake various continuous improvement activities, both intermittent and continuous, for pain/nonpain symptoms, psychosocial and spiritual distress, and communication between providers and patient and surrogates. There should be continuing education in palliative care for all health care professionals as well as training and support for specialist palliative care professionals. Five of the NQF preferred practices are aimed at measuring clinical outcomes, including measuring pain and other symptoms both physical (for example, nausea, constipation, dyspnea) and psychological using standardized scales such as the Memorial Symptom Assessment Schedule,[79] as well as assessing patient satisfaction outcomes (including ability of staff to discuss hospice as an option). To be able to achieve effectiveness in all of these domains requires an interdisciplinary team of appropriately trained or certified staff, including physicians, nurses, social workers, pharmacists, and chaplains. The team should be knowledgeable in the assessment and management of symptoms, trained to assess and manage the patient's and family's emotional state, adept at discussing goals of treatment and offering advance care planning, and able to identify when the patient is transitioning to the dying phase.

Efficient

Palliative care services should avoid waste and harmful delays for those who receive and give care. With regard to avoiding waste, the palliative care service needs to be interdisciplinary and should avoid duplication. With regard to cost-savings for the health care system, the study by Morrison et al. showed that, on average, US$300 per day can be saved in directive costs through adopting palliative care, primarily via reducing laboratory and Intensive Care Unit (ICU) costs.[80] To avoid harmful delays, there needs to be a smooth transition between acute care and palliative care, and it needs to be bidirectional. The NQF requires that to the extent that existing privacy laws permit, care plans developed for the patients should be disseminated both internally and externally to all professionals involved in the patient care, especially upon transfer between care settings.

Patient-Centered

Hospice should be introduced as an option for patients with a life-limiting illness as a less aggressive approach to care. As the patient declines, end-of-life care should be modified to reflect the change in status. Care needs to be respective of and responsive to individual patient and family preferences, needs, and values and ensure that patient values guide all clinical decisions. Cultural assessment has now become an important part of high-quality, appropriate, contemporary palliative care services. The cultural assessment should include, but not be limited to, locus of decision making (individual versus family), preferences regarding disclosure of information, truth telling and decision making, dietary preferences, language, family communication, desire for

support measures such as palliative care, perspectives on death, burial customs, and grieving. Professional translator services and culturally sensitive materials in the patients own language should be available. Regular patient and family conferences should be held to provide information and to discuss changes in disease status, prognosis, goals of care, advanced cancer planning and to offer support both to the patient and the family.

The patient-centered care plan needs to assess and manage the patient and family's psychological reactions to the situation (including stress, anticipatory grief, and coping); a grief and bereavement care plan; a social care plan to address the social practical and legal needs of the patient, family, and other informal caregivers; an assessment of religious spiritual and existential concerns. To formulate the best care plan, the patient and family need education on the disease, its natural history, prognosis, and the burden or benefits of any interventions. This includes educating the family on the signs and symptoms of an imminent death, when timely and as appropriate. Any treatments offered must be appropriate to patient and family. In the case of pediatric palliative care patients, the child's preference should be elicited if possible and respected.

With regard to the documentation of the goals of care, this may be in the form of an advanced directive or living will, but it is now preferable to convert them into medical orders such as the Medical Orders for Life Sustaining Therapy form, which are transferrable across care settings.[81] Initiatives such as Web-based registries or electronic personal health care records may be needed to achieve transfer of this information between care settings and to protect patient privacy and be compliant with HIPAA regulations.

Another important process component of good quality palliative care is the care of the imminently dying and his or her families. Fulfilling patient and family preferences regarding preference of place of death, ensuring that adequate doses of analgesics and sedatives are ordered to relieve pain, treating the body after death with respect, and implementing a bereavement care plan to facilitate effective grieving by the family are all important factors in any high-quality palliative care program.

Timely

Timely care means reducing wait time. Ensuring that a patient can access palliative care sooner rather than later is a very important component of palliative care. Generally speaking, in the case of cancer patients, the offer of palliative care services or referral to hospice is uniformly late, usually in the last week or two of life. Educational efforts to raise the level of awareness of palliative care are fundamental. If patients and families do not know their options, they will not ask about palliative care.

Equitable

Palliative care should not vary in quality because of personal characteristics, such as gender, ethnicity, geographic location, and socioeconomic status. U.S. health care has been differentiated from European-style health care by that fact that

the ethical principle of distributive justice (that is, access to care) is trumped by the principle of autonomy (the right to choose, or in the case of end-of-life care, the right to refuse treatment). As health care reform is being debated and formulated, it is important to recognize that palliative care available to all in need must be an integrated component of any reform package. Health care providers must make a concerted effort to incorporate hospice principles into hospitals and nursing homes where the majority of Americans spend their last days of life.

One should not get the impression that little has been done; progress indeed has been made. As of 2008, of the more than 5,000 hospitals in the United States, approximately 1,314 now have palliative care programs, with 2,678 physicians and 15,133 nurses certified in palliative care. But more needs to be done. There is a strong bias within the U.S. health care system toward curative medicine, including financial incentives that encourage provision of aggressive treatment. Even if palliative care were universally available in the health care system, the 15 percent of the population lacking insurance (racial and ethnic minorities being overrepresented) could not access it. Such inequality must be addressed and changes made to ensure accessibility and equality in the provision of palliative care.

CONCLUSION

Relief of symptoms and other suffering at end of life has emerged as an important goal of medicine in the 21st century, and this domain of medical practice is now known as palliative care. Modern palliative care emerged as a grassroots movement in Australia, Britain, India, and the United States, but it is only within the last 10 to 15 years that palliative care has been integrated better into the medical system. The United States is a comparative newcomer in this area of medicine and can learn much from the organizational framework for developing palliative care services from countries like Britain, Australia, and India. In particular, these countries can provide guidance on the policy development piece regarding how governments might work with the private sector in the design and implementation of a high-quality, ethical palliative care program in America.

In the 21st century, patients and practitioners have a better appreciation of the need for and importance of treating pain and suffering at the end of life. Not only is the medical management important, but also the economic consequences of treating end-of-life patients aggressively in hospital is extremely costly. The issue will not diminish in scope given the aging of the population with its concomitant burden of chronic diseases. High-quality palliative care can do much not only to alleviate and relieve suffering, but also to contain costs.

Every one of us, be it for oneself or for one's loved ones, will at some point have to confront the inevitability of end of life. How one prepares for the eventuality of death is a personal and individual matter. What is necessary and important, however, is that end-of-life choices be made clear and available. Providing for end-of-life care is emotionally difficult, thus making it even more imperative that all patients have the option of timely access to palliative care services that are both appropriate and cost-effective. While impressive strides have been made, much more work needs to be done.

Appendix 8.A.
Results of CAPC Report Card 2008

- States Receiving A Grades
  States that are top performers (programs in 81 to 100 percent of hospitals):
  Vermont (100 percent); Montana (88 percent); New Hampshire (85 percent)
- States Receiving B Grades
  States that are on their way (programs in 61 to 80 percent of hospitals):
  District of Columbia (80 percent); South Dakota (78 percent); Minnesota (75 percent); Missouri (73 percent); New Jersey (72 percent); Oregon (72 percent); Iowa (70 percent); Maine (69 percent); Michigan (69 percent); North Carolina (69 percent); Ohio (68 percent); Colorado (67 percent); Maryland (67 percent); West Virginia (67 percent); North Dakota (67 percent); Washington (65 percent); Wisconsin (64 percent); Virginia (63 percent); Kansas (61 percent)
- States Receiving C Grades
  States in the middle (programs in 41 to 60 percent of hospitals):
  Alaska (60 percent); Delaware (60 percent); Rhode Island (60 percent); Illinois (58 percent) New York (58 percent); California (56 percent); Idaho (56 percent); Nebraska (56 percent); Utah (56 percent); Connecticut (54 percent); Indiana (54 percent); Pennsylvania (54 percent); Arizona (50 percent); Massachusetts (50 percent); Florida (49 percent); Hawaii (45 percent); Tennessee (45 percent); Arkansas (41 percent)
- States Receiving D Grades
  States that need significant improvement (programs in 21 to 40 percent of hospitals):
  Georgia (38 percent); Kentucky (37 percent); New Mexico (33 percent); Texas (33 percent); South Carolina (30 percent); Louisiana (27 percent); Wyoming (25 percent); Nevada (23 percent)
- States Receiving Failing Grades
  States with little or no access (programs in 0 to 20 percent of hospitals):
  Oklahoma (19 percent); Alabama (16 percent); Mississippi (10 percent)

Note: CAPC = Center to Advance Palliative Care

REFERENCES

1. Sepulveda C, Marlin A, Yoshida T, et al. Palliative care: the World Health Organization's global perspective. *J Pain Symptom Manage*. 2002;24:91–6.
2. Stjernsward J. Palliative care: the public health strategy. *J Public Health Policy*. 2007;28:42–55.
3. Morrison RS, Dietrich J, Meier DE. *America's Care of Serious Illness: A State-By-State Report Card on Access to Palliative Care in our Nation's Hospitals*. New York: Centre to Advance Palliative Care and the National Palliative Care Research Center; 2008:36.
4. World Health Organization. Ask the expert Online Q&A for April 1, 2008. Are the number of cancer cases increasing or decreasing in the world? http://www.who.int/.
5. Somerville M. *The Ethical Canary: Science, Society and the Human Spirit*. Toronto, Canada: Penguin Canada; 2000:254.
6. Teno JM, Clarridge BR, Casey V, et al. Family perspectives on end-of-life care at the last place of care. *JAMA*. 2004;291:88–93.
7. Covinsky KE, Goldman L, Cook EF, et al. The impact of serious illness on patients' families: Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment (SUPPORT) investigators. *JAMA*. 1994;27(2):1839–1844.

8. Addington-Hall J, McCarthy M. Dying from cancer: results of a national population-based investigation. *Palliat Med*. 1995;9:295–305.
9. Toscani F, Di Giulio P, Brunelli C, et al. How people die in hospital general wards: a descriptive study. *J Pain Symptom Manage*. 2005;30:33–40.
10. Brennan F, Carr DB, Cousins M. Pain management: a fundamental human right. *Anesth Analg*. 2007;105:205–221.
11. The SUPPORT Principal Investigators. A controlled trial to improve care for seriously ill hospitalized patients: the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT). *JAMA*. 1995; 274:1591–1598.
12. Institute of Medicine. *Approaching Death*. Washington, DC: National Academies Press; 1997.
13. Foley K, Hendin H. *The Case against Suicide*. Baltimore: Johns Hopkins University Press; 2002.
14. von Gunten CF, Lupu D. Development of a medical subspecialty in palliative medicine: progress report. *J Palliat Med*. 2004;7:209–219.
15. Levy MH, Back A, Benedetti C, et al. NCCN clinical practice guidelines in oncology: palliative care. *J Natl Compr Canc Netw*. 2009;7:436–473.
16. Meier DE. Palliative care in hospitals. *J Hosp Med*. 2006;1:21–28.
17. Saunders CM. Foreword. In: Doyle D, Hanks G, Cherny N, et al., eds. *Oxford Textbook of Palliative Medicine*. Oxford: Oxford University Press; 2004:xvii–xx.
18. Clark D, Seymour J. *Reflections on Palliative Care: Sociological and Policy Perspectives*. Buckingham, UK: Open University Press; 1999:67–78.
19. Clark D. *Cicely Saunders: Founder of the Hospice Movement. Selected Letters 1959–1999*. Oxford: Oxford University Press; 2002.
20. Bean WB. Peace at the last: a survey of terminal care in the United Kingdom: a report to the Calouste Gulbenkian Foundation 1960. *Arch Intern Med*. 1961;107:631–632.
21. Hinton JM. The physical and mental distress of the dying. *Q J Med*. 1963;32:1–21.
22. *Palliative Care Explained*. London,UK: The National Council for Palliative Care; 2009. http://www.ncpc.org.uk/palliative_care.html. Accessed September 29, 2009.
23. Franks PJ, Salisbury C, Bosanquet N, et al. The level of need for palliative care: a systematic review of the literature. *Palliat Med*. 2000;14:93–104.
24. Guidance on Cancer Services. *Improving Supportive and Palliative Care for Adults with Cancer*. London, UK: National Institute for Clinical Excellence; 2004.
25. Eagar K, Garrett P, Lin V. *Health Planning: Australian Perspectives*. Crows Nest, NSW, Australia: Allen & Unwin; 2001.
26. Rosenwax LK, McNamara B, Blackmore AM, et al. Estimating the size of a potential palliative care population. *Palliat Med*. 2005;19:556–562.
27. The National Palliative Care Strategy. *A National Framework for Palliative Care Service Development: October 2000*. Canberra, Australia: Australian Government; 2000.
28. Greer DS, Mor V, Morris JN, et al. An alternative in terminal care: results of the National Hospice Study. *J Chronic Dis*. 1986;39:9–26.
29. Gazelle G. Understanding hospice: an underutilized option for life's final chapter. *N Engl J Med*. 2007;357:321–324.
30. Lynn J. Perspectives on care at the close of life: serving patients who may die soon and their families: the role of hospice and other services. *JAMA*. 2001;285:925–932.
31. Christakis NA, Iwashyna TJ. Attitude and self-reported practice regarding prognostication in a national sample of internists. *Arch Intern Med*. 1998;158:2389–2395.

32. Wright AA, Katz IT. Letting go of the rope: aggressive treatment, hospice care, and open access. *N Engl J Med*. 2007;357:324–327.
33. Meier DE, Morrison RS, Cassel CK. Improving palliative care. *Ann Intern Med*. 1997;127:225–230.
34. von Gunten CF. Secondary and tertiary palliative care in US hospitals. *JAMA*. 2002;287:875–881.
35. Higginson IJ, Koffman J. Public health and palliative care. *Clin Geriatr Med*. 2005;21:viii,45–55.
36. Higginson I. Palliative care: a review of past changes and future trends. *J Public Health Med*. 1993;15:3–8.
37. Gomez-Batiste X, Porta-Sales J, Paz S, et al. Program development: palliative medicine and public health services. In: Walsh D, ed. *Palliative Medicine*. Philadelphia, PA: Saunders; 2008:198–203.
38. Ingleton C, Skilbeck J, Clark D. Needs assessment for palliative care: three projects compared. *Palliat Med*. 2001;15:398–404.
39. Currow DC, Abernethy AP, Fazekas BS. Specialist palliative care needs of whole populations: a feasibility study using a novel approach. *Palliat Med*. 2004; 18:239–247.
40. Institute of Medicine. *The Future of Public Health*. Washington, DC: National Academies Press; 1988:40–55.
41. Taylor DH Jr. The effect of hospice on Medicare and informal care costs: the US Experience. *J Pain Symptom Manage*. 2009;38:110–114.
42. Miceli PJ, Wojciechowski SL. Impacting family satisfaction with hospice care. *Caring*. 2003;22:14–18.
43. Connor SR, Teno J, Spence C, et al. Family evaluation of hospice care: results from voluntary submission of data via website. *J Pain Symptom Manage*. 2005;30: 9–17.
44. Higginson IJ, Foley KM. Palliative care: no longer a luxury but a necessity? *J Pain Symptom Manage*. 2009;38:1–3.
45. Pan CX, Morrison RS, Meier DE, et al. How prevalent are hospital-based palliative care programs? status report and future directions. *J Palliat Med*. 2001;4:315–324.
46. Morrison RS, Maroney-Galin C, Kralovec PD, et al. The growth of palliative care programs in United States hospitals. *J Palliat Med*. 2005;8:1127–1134.
47. Goldsmith B, Dietrich J, Du Q, et al. Variability in access to hospital palliative care in the United States. *J Palliat Med*. 2008;11:1094–1102.
48. Center to Advance Palliative Care. *America's Care of Serious Illness: A State-By-State Report Card on Access to Palliative Care in our Nation's Hospitals*. New York: Center to Advance Palliative Care; 2008.
49. Crawley L, Payne R, Bolden J, et al. Palliative and end-of-life care in the African American community. *JAMA*. 2000;284:2518–2521.
50. Fins JJ. *A Palliative Ethic of Care: Clinical Wisdom at Life's End.* Sudbury, MA: Jones and Bartlett Publishers; 2006.
51. Australian Government, Department of Health and Aging. *Community Attitudes towards Palliative Care: Integrated Report*. Canberra, Australia: Australian Government: 2006.
52. Von Roenn JH, Cleeland CS, Gonin R, et al. Physician attitudes and practice in cancer pain management: a survey from the Eastern Cooperative Oncology Group. *Ann Intern Med*. 1993;119:121–126.
53. Brickner L, Scannell K, Marquet S, et al. Barriers to hospice care and referrals: survey of physicians' knowledge, attitudes, and perceptions in a health maintenance organization. *J Palliat Med*. 2004;7:411–418.

54. Edmonds PM, Stuttaford JM, Penny J, et al. Do hospital palliative care teams improve symptom control? Use of a modified STAS as an evaluation tool. *Palliat Med*. 1998;12:345–351.
55. Komurcu S, Nelson KA, Walsh D, et al. Common symptoms in advanced cancer. *Semin Oncol*. 2000;27:24–33.
56. Schuit KW, Sleijfer DT, Meijler WJ, et al. Symptoms and functional status of patients with disseminated cancer visiting outpatient departments. *J Pain Symptom Manage*. 1998;16:290–297.
57. Cleeland CS, Gonin R, Hatfield AK, et al. Pain and its treatment in outpatients with metastatic cancer. *N Engl J Med*. 1994;330:592–596.
58. Reuben DB, Mor V. Nausea and vomiting in terminal cancer patients. *Arch Intern Med*. 1986;146:2021–2023.
59. Breitbart W, Rosenfeld BD, Passik SD, et al. The undertreatment of pain in ambulatory AIDS patients. *Pain*. 1996;65:243–249.
60. von Gunten CF. Financing palliative care. *Clin Geriatr Med*. 2004;20:viii,767–781.
61. Stjernsward J, Foley KM, Ferris FD. Integrating palliative care into national policies. *J Pain Symptom Manage*. 2007;33:514–520.
62. Joranson DE, Ryan KM. Ensuring opioid availability: methods and resources. *J Pain Symptom Manage*. 2007;33:527–532.
63. De Lima L, Krakauer EL, Lorenz K, et al. Ensuring palliative medicine availability: the development of the IAHPC list of essential medicines for palliative care. *J Pain Symptom Manage*. 2007;33:521–526.
64. Brennan F. Palliative care as an international human right. *J Pain Symptom Manage*. 2007;33:494–499.
65. Bath PA, Clark D, Wright M. The International Observatory on End of Life Care (IOELC): an information resource for palliative care. *Health Inform J*. 2004;10(2)121–126.
66. Kerns RD. *VHA Directive 2003-021: Pain Management*. Washington DC: Veterans Health Administration; 2003.
67. Cassel CK, Vladeck BC. ICD-9 code for palliative or terminal care. *N Engl J Med* 1996;335:1232–1234.
68. Medicare Learning Network. *MLN Matters Number MM6311: Adding a New Specialty Code for Hospice and Palliative Medicine*. Center for Medicare and Medicaid Services; 2009. www.cms.gov/mlnmattersarticles/downloads/MM6311.pdf
69. Foley KM, Gelband H. *Improving Palliative Care for Cancer: Summary and Recommendations*. Washington, DC: National Academies Press; 2001.
70. Weissman DE, Meier DE. Operational features for hospital palliative care programs: consensus recommendations. *J Palliat Med*. 2008;11:1189–1194.
71. Ferris FD, Bruera E, Cherny N, et al. Palliative cancer care a decade later: accomplishments, the need, next steps: from the American Society of Clinical Oncology. *J Clin Oncol*. 2009;27:3052–3058.
72. Ferris FD, Gomez-Batiste X, Furst CJ, et al. Implementing quality palliative care. *J Pain Symptom Manage*. 2007;33:533–541.
73. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press; 2001.
74. World Health Organization. *Cancer Pain Relief and Palliative Care*. Geneva, Switzerland: World Health Organization; 1990.
75. Jadad AR, Browman GP. The WHO analgesic ladder for cancer pain management. Stepping up the quality of its evaluation. *JAMA*. 1995;274:1870–1873.
76. Wiffen PJ, McQuay HJ. Oral morphine for cancer pain. *Cochrane Database Syst Rev*. 2007;4:CD003868. doi:10.1002/14651858.

77. Zimmermann C, Riechelmann R, Krzyzanowska M, et al. Effectiveness of specialized palliative care: a systematic review. *JAMA*. 2008;299:1698–1709.
78. Lorenz KA, Lynn J, Dy SM, et al. Evidence for improving palliative care at the end of life: a systematic review. *Ann Intern Med*. 2008;148:147–159.
79. Chang VT, Hwang SS, Kasimis B, et al. Shorter symptom assessment instruments: the Condensed Memorial Symptom Assessment Scale (CMSAS). *Cancer Invest*. 2004;22:526–536.
80. Morrison RS, Penrod JD, Cassel JB, et al. Cost savings associated with US hospital palliative care consultation programs. *Arch Intern Med*. 2008;168:1783–1790.
81. Hickman SE, Nelson CA, Moss AH, et al. Use of the Physician Orders for Life-Sustaining Treatment (POLST) paradigm program in the hospice setting. *J Palliat Med*. 2009;12:133–141.

SECTION 4

PUBLIC HEALTH PRACTICE AND EDUCATION

This page intentionally left blank

CHAPTER 9

The Role of Epidemiology and Biostatistics in Health News Reporting

Paula Trushin and Heejung Bang, PhD

Medicine is a science of uncertainty and an art of probability.

—Sir William Osler

How often does a health news story make you pause and wonder what to do next? Headlines such as "Highly active compound found in coffee may prevent colon cancer,"[1] "The surprising ingredient causing weight gain,"[2] or "Three cups of brewed coffee a day 'triples risk of hallucinations,'"[3] grab the reader's attention, but are oversimplified, sensationalized,[4] and—in most cases—speak with undue authority. Barry Kramer, associate director for disease prevention at the National Institutes of Health, calls such reporting "the cure of the week or the killer of the week, the danger of the week. It's like treating people to an endless regimen of whiplash."[5] And Donald Berry, chair of the Department of Biostatistics at the M. D. Anderson Cancer Center, says that he has "seen so many contradictory studies with coffee that I've come to ignore them all."[5]

THE ROLE OF UNCERTAINTY IN SCIENCE

Science is a way of thinking and trying to understand the natural world. Advances in science are largely propelled by the scientific method, a rational and systematic mode of inquiry (based largely on common sense)[6] whereby natural phenomena can be studied through observation and experimentation. Statistics, which may be considered a tool of the scientific method, provides a theoretical basis and mathematical techniques for collecting, organizing, and analyzing data quantitatively, taking into account chance, variation, and error (defined as the difference between the computed or measured value of a factor and its actual or theoretically correct value). When used appropriately, statistics makes it possible to describe an aspect of reality and draw inferences about the relationships among variables (for a definition of "variable," see "The Meaning of Words"). Inherent in this process is

the notion of "scientific uncertainty": a constant state of doubt or limited knowledge. Any meaningful attempt to understand health-related information requires an appreciation of the essential role that uncertainty plays in science, spurring the development of new ideas and leading to discovery. Statistics can help manage uncertainty by quantifying the likelihood that a particular event occurred by chance.

Like all human endeavors, the scientific method is not infallible, but it can be self-correcting over time as long as its practitioners remain open to doubt and questions about their work. Thus, disagreement and—yes—controversy are fundamental to the advancement of science.[7]

Sometimes, however, scientific controversies are manufactured for nonscientific reasons, whether political, economic, or religious. This may be done by overstating (sensationalizing) research findings; ignoring findings that do not support a particular viewpoint; casting doubt on empirical evidence, despite its strength and consistency; or arguing that legitimate differences of opinion are being suppressed.[8–12] Manufactured controversies have serious consequences. Not only can they lead to public mistrust of science and make it difficult for lay persons to appreciate legitimate differences of scientific opinion, but they also can cause public alarm, as was the case with stories linking the MMR (measles, mumps, and rubella) vaccine to autism.[10, 11, 13] Inappropriate use of scientific findings may compromise clinical practice recommendations, resulting in misdirected prevention and treatment strategies.[14]

Because the mass media influence personal health care choices, public health policies, and medical decision making, it is incumbent on everyone to develop an understanding of what statistics mean and how they are used (and sometimes misused) in health news reporting,[10, 15, 16]

This chapter will focus on statistical thinking—a broad and flexible mode of reasoning about data, variation, and chance.[17] Since there are virtually no "facts" in science, only varying degrees of certainty depending on the strength of the available evidence, statistical thinking begins with an acceptance of chance and uncertainty.[18–23] It also entails an understanding of basic principles and concepts, which we shall endeavor to explain; an appreciation of context; an ability to detect logical and factual flaws in information and ideas;[24] and the realization that science is fluid, with new empirical evidence being accumulated every day. To grasp these fundamentals, one does not have to be scientifically trained. (This chapter will not discuss the Bayesian paradigm or Bayesian statistics, which is a way of quantifying probability by factoring in prior knowledge and additional evidence.)

THE MEANING OF WORDS

A number of familiar words have scientific definitions that differ from the vernacular. They include theory, sample, uncertainty, error, bias, odds, parameter, and significance. Unless explained, words such as these hinder public understanding of science (the "homonymic obstacle"[25]). Members of the public cannot be expected to know which words have both a vernacular and a scientific definition. Therefore, scientists and journalists have a responsibility to make these distinctions clear.

Throughout this chapter, words with a distinct scientific or statistical meaning will be defined as they occur. Let us begin, however, by explaining certain key terms.

For the scientist, *a theory* does not mean a conjectural and unsubstantiated statement (as in the saying that something is "just a theory"). A scientific theory is supported by observation and experimentation, has been verified to some degree by different investigators,[26] and is used to explain and predict a natural phenomenon. If a theory is contradicted by new observations, it must be revised or replaced, but this does not mean that the underlying phenomenon has changed. As Stephen Jay Gould pointed out, "Einstein's theory of gravitation replaced Newton's in [the 20th] century, but apples didn't suspend themselves in midair pending the outcome."[27] In contrast, *a hypothesis* is a conjectural explanation for an observed phenomenon (see "Study Designs").

Risk is the estimable probability, or quantifiable likelihood, that a particular (generally negative) event will occur. In medicine, some risks, including a number of genetic disorders, are biologically determined. Many major diseases are multifactorial; however, they depend on a combination of environmental (for example, chemicals, pathogens, physical injury, social forces, lifestyle and behavior) and biological factors. For example, we can scarcely say that there is "a" cause of coronary heart disease (CHD) or lung cancer. In such cases, statistics are used to estimate the probability that a given risk factor (or set of risk factors), or an intervention, will increase, or decrease, risk.

Exposure refers to contact with a physical, chemical, or biologic influence (including a therapeutic intervention).

A *variable* is any factor or characteristic that varies, either from person to person or within a person. To be studied scientifically, a variable must be classifiable or measurable, for example, age, sex, body weight, or blood chemistry.

ASSOCIATION, CAUSATION, CONFOUNDING, AND BIAS

Association and causation, which refer to the relationship between two variables, are fundamental concepts in scientific research. An association is a correspondence or coincidence between two variables in cases in which one does not necessarily affect or alter the other.[28] In contrast, causation is defined as an act or process whereby one factor (an "independent variable") creates or produces an outcome (a "dependent variable"). An observed association (or correlation) between two events (or variables) does not mean causation. For instance, if children with larger feet are better spellers, does this mean that foot size causes orthographic proficiency? Undoubtedly not, since age is a more likely explanation for both.[29]

This introduces the notion of confounding, which refers to a confusion of effects.[30] A confounder is generally understood to be a variable, possibly unidentified or unmeasured, that (1) is associated with the exposure, either causally or noncausally; (2) has a cause-related effect on the outcome; and (3) is not a causal factor on the pathway between exposure and outcome.[31–33] For example, some observational evidence suggests an association between alcohol use and an increased risk for lung cancer. A possible confounder in this case would be cigarette smoking, which is associated both with drinking and (causally) with lung cancer. By failing to identify confounders, we may end up concluding erroneously that an association between two variables represents a cause-and-effect relationship (however unlikely it may seem in some cases).

Confounding can distort the results of a study. In extreme situations, it can even invert the direction of an association, so that an exposure appears to decrease, rather than increase, the risk for an event (or vice versa).[32]

When designing a study, researchers must try to identify and, if possible, control the variables that can confound the relationship under investigation or lead to a spurious interpretation of that relationship. Failure to do this could result in *bias.* In science, bias does not mean a prejudiced outlook. It refers to unrecognized error that systematically and consistently pervades the entire study, distorting the results. The best possible way to correct bias would be to pinpoint and eliminate any potential source(s) in the study design phase.

Frequently, the media contribute to the confusion between association and causation by oversimplifying or sensationalizing the results of scientific research.[4] This can mislead and misinform the reader.

RISK

In all fields of endeavor, history abounds with remarkable developments that have transformed human life. One such development is "mastery of risk"—the notion that the probability of future events can, to some degree, be understood, quantified, and predicted [34] by statistical means, instead of entrusted to the gods and the prophecies of oracles. There are two standard ways of quantifying risk: absolute and relative. Let us imagine a study in which 150 subjects are given drug X, an experimental treatment to prevent major cardiovascular events, and 250 subjects receive placebo (an inert substance disguised as medication). Table 9.1 illustrates how measures of absolute and relative risk would be calculated in such a scenario.[35]

Absolute Risk

Absolute risk, also called the *event rate*, is a percentage representing the number of people in a group who experience an event, such as a heart attack, in relation to the total number of people in the group. It also can be calculated by using the number of actual events in relation to the total number of possible events (that is, the total number of *at-risk* people in the group).

To assess the effectiveness of a drug or other intervention, clinical studies commonly compare the event rate with treatment (the event rate in the experimental group) versus the event rate in a similar group of untreated (control) subjects. There are two basic methods of comparison: subtraction or division.

The *absolute risk reduction* is determined by *subtracting* the experimental event rate from the control event rate: 40 percent − 10 percent = 30 percent (see table 9.1). If the intervention increases the risk for an event, however, the *absolute risk increase* is calculated in the reverse manner, also illustrated in table 9.1: 50 percent − 40 percent = 10 percent. If the event rates in both groups are the same, there will be no difference between them. Thus, a value of zero means that the intervention has no effect.

Relative Risk

Relative Risk, or *risk ratio*, represents comparison by division. To calculate the relative risk, we divide the experimental event rate by the control event rate. Based

Table 9.1
Sample Calculations of Predictive Measures (risk or odds)

Subjects	Risk Reduction (Intervention Beneficial)		Risk Increase (Intervention Harmful)	
	Experimental Group	Control Group	Experimental Group	Control Group
With Events (E)	15	100	75	100
With No Events (N)	135	150	75	150
Total (T)	150	250	150	250
Absolute Risk (Event Rate) E/T	15/150 = 10% (EER)	100/250 = 40% (CER)	75/150 = 50% (EER)	100/250 = 40% (CER[a])
Odds E/N	15/135 = 11%	100/150 = 67%	75/75 = 1	100/150 = 67%

Predictive Measure	Equation	Intervention Beneficial (Experimental Group)	Intervention Harmful (Experimental Group)
Absolute Risk Reduction	**CER – EER = ARR**	0.40 – 0.10 = 0.30 (30%)	NA
Absolute Risk Increase	**EER – CER = ARI**	NA	0.50 – 0.40 = 0.10 (10%)
Number Needed to Treat	**1/ARR**	1.0/0.30 = 3.33	NA
Relative Risk Reduction	**ARR/CER = RRR**	0.30/0.40 = 0.75 (75%)	NA
Relative Risk Increase	**ARI/CER = RRI**	NA	0.10/0.40 = 0.25 (25%)
Relative Risk (Risk Ratio)	**EER/CER = RR**	0.10/0.40 = 0.25 (25%)	0.50/0.40 = 1.25 (125%)
Odds	**E/N**	15/135 = 11%	75/75 = 1
Odds Ratio	**E/N (EXP) / E/N (CONT)**	0.11/0.67 = 0.164	1.0/0.67 = 1.5

a. The control event rate is also the baseline risk, because it represents the percentage of untreated people in the target population who are likely to experience an event.

Source: Adapted from Number needed to treat. Wikipedia, the free encyclopedia. Available at: http://en.wikipedia.org/wiki/Number_needed_to_treat. Retrieved September 28, 2009.[35]

Note: ARI = absolute risk increase; ARR = absolute risk reduction; CER = control event rate (also called the baseline risk); EER = experimental event rate; EXP = experimental group; CONT = control group; RR = relative risk; RRI = relative risk increase; RRR = relative risk reduction.

on the data in table 9.1, the relative risks for the beneficial and the harmful interventions are 0.25 and 1.25, respectively. In a clinical study, the control event rate is also called the baseline risk, because it indicates the percentage of *untreated* people in the population from which the study participants were drawn who are likely to experience an event. If the experimental and the control event rates are the same, the ratio will equal one. Thus, for the relative risk (and the odds ratio, discussed below), a value of one means that the exposure has no effect. The *relative risk reduction* is a ratio of the absolute risk difference to the baseline risk.

Relative differences are generally greater than absolute differences; thus, they may magnify the effect of a treatment or exposure. Moreover, they are nonspecific, since any baseline risk can be reduced by the same relative amount. In table 9.2, for example, absolute risk reductions of 2 percent and 24 percent both translate into a relative risk of 33 percent and a relative risk reduction of 67 percent. The potential misuse of relative numbers is illustrated by a CBS News headline announcing that "Trial AIDS Vaccine Cuts Infection Risk 31 Percent."[36] In this article, the lead paragraph describes the result as a "watershed event," but the story fails to explain that 31 percent is a relative number, and that the absolute risk reduction was just 0.3 percent. When the raw data (number of events and total number of subjects in the treatment and control groups) are given in an article, the reader can easily calculate both the experimental and the control event rates, as well as the absolute risk reduction (by simple subtraction). The calculation takes just a few minutes, yet it will contribute substantially to a more responsible understanding of the study results.

Odds Ratio

In statistics, odds refer to the probability that something will occur in relation to the probability that it will not. If the underlying risk is small (that is, the event of interest is rare), the odds ratio will approximate the relative risk. As the baseline (absolute) risk increases, however, the difference between the relative risk and the odds ratio grows larger. In most studies with a binary exposure status (exposed versus not exposed or treated versus untreated) and outcome (event versus nonevent), the data are customarily analyzed by using either the relative risk or the odds ratio. One type of observational study, however, must use the odds ratio only. This is the case-control study, which will be discussed in the following section.

Number Needed to Treat

Number needed to treat (NNT) refers to the number of patients who must be treated to prevent one (additional) event. The number needed to treat is the reciprocal of the absolute risk reduction (1 divided by the absolute risk reduction). If the absolute risk is based on the number of events, rather than the number of people with events, the recurrence of multiple events in some subjects may increase the difference between the treatment and the control groups, thereby lowering the number needed to treat and exaggerating the benefit of the intervention.

Table 9.2
Absolute versus Relative Risk

Risk for Disease					
Group I Experimental Event Rate	Group II Control Event Rate (Baseline Risk)	Absolute Risk Difference (B – A = C)	Relative Risk (A/B)	Relative Risk Reduction (C/B = D)	Absolute Risk Difference (B × D)
1.0% (10/1,000)	3.0% (30/1,000)	3.0% − 1.0% = 2.0%	33.0%	67.0%	2.0%
12.0% (240/1,000)	36.0% (360/1,000)	36.0% − 12.0% = 24.0%	33.0%	67.0%	24.0%
27.0% (270/1,000)	81.0% (810/1,000)	81.0% − 27.0% = 54.0%	33.0%	67.0%	54.0%

Note: A = experimental event rate; B = control event rate (also called "baseline risk"); C = absolute risk difference (reduction or increase); D = relative risk reduction.

STUDY DESIGNS: OBSERVATIONAL VERSUS EXPERIMENTAL

The purpose of research is to develop or contribute to generalizable knowledge.[37] When conducted in human subjects, it is known as clinical research. The two basic types of clinical research are *observational and experimental*.

Observational Studies

In observational studies (also called epidemiologic or nonexperimental studies), the researcher does not attempt to affect the outcome or control other variables that may influence the results. This type of research can be either descriptive, providing information on health-related events in populations, or analytic, in which researchers attempt to quantify the relationship between an exposure and an outcome. There are several observational research designs, distinguished chiefly by the method used to select the study sample (participants). The three basic designs—cross-sectional, cohort, and case-control—cover most observational studies, including genetic studies. Each has its advantages and disadvantages, and the decision about which one to use depends on multiple considerations.

Observational studies also can be forward looking (prospective) or backward looking (retrospective). In the former, disease-free individuals exposed and unexposed to a suspected risk factor are compared over time to determine whether an outcome develops. In the latter, subjects are identified after the outcome has occurred, and various means (for example, medical records, subject interviews) then are used to look back and see if an association between an exposure, such as cigarette smoking, and an outcome, such as lung cancer, can be identified. Because there is no need to wait for outcomes to develop, retrospective study designs are generally less costly and time-consuming than prospective designs, and they also can be useful as pilot studies in anticipation of prospective research.[38] But retrospective studies can be hampered by incomplete and unreliable data, since medical records may not be up to date and subjects may have faulty or selective memories (known as "recall bias"). With few exceptions, therefore, prospective studies are methodologically preferable to their retrospective counterparts.[39]

Because of confounding and various biases, which are part of the natural settings in which observational studies take place, evidence from this type of research generally is limited or not reliable as a basis for causal inference. As an example, let us imagine a prospective cohort study showing a lower risk for CHD among women taking antioxidant vitamins compared with those who do not. This result may be influenced by what we call the healthy user effect or bias, which means a tendency among antioxidant users to be particularly health conscious overall, with habits that include a low-fat diet, regular exercise, and proper medical care.[39] Other factors, including income level and years of education, also may confound (provide an alternative explanation for) the results if each is independently associated with antioxidant use and lower CHD risk.[40]

A case-control study is a form of observational research in which subjects are selected on the basis of outcome, rather than exposure. Consequently, subjects with the disease (cases) exceed the percentage in the population from which the study sample is drawn.[41] This means that the sample is not representative. Because a

representative sample is fundamental to the prediction of risk, the data from case-control studies may only be analyzed in terms of the odds ratio.

To illustrate, let us look at a case-control investigation of hip/femur fracture (the outcome) and stroke (the exposure).[42] This study reported an odds ratio of 2.0, indicating that those who sustained a fracture were twice as likely to have suffered a stroke than their uninjured counterparts, but it cannot do the reverse and predict the risk for fracture in stroke patients.[41] Nevertheless, the following headline from *US News and World Report* implies a prospective study of risk prediction (similar headlines appeared in many other news outlets): "Stroke Doubles Risk of Hip, Thigh Fractures."[43] A more accurate (although less compelling) headline could be: "Odds of Having Had a Stroke Are Greater in Patients with Hip, Thigh Fractures."

While case-control studies are appropriate in certain situations (for example, the study of rare diseases or diseases in which there is a long lag time between exposure and outcome), their results should not be interpreted as predictive of risk for a future event. Even if the observational study described above were prospective, the headline would still be misleading because it implies that X causes Y, a claim that generally requires experimental evidence. When evaluating a health news story about clinical research, we should ask whether the study was observational or experimental and, if observational, what study design was used (see appendix 9.A).

Experimental Studies

The fundamental difference between observational and experimental research is that, in the latter, investigators can control factors that might influence the results, with the goal of being able to infer causation, quantify risk, and make probabilistic statements that could apply to like circumstances. Experimental research involving human subjects typically is conducted by means of randomized controlled trials (RCTs), a rigorous form of scientific investigation that eliminates or, via the coin-toss mechanism, controls for multiple variables that could affect the result; consequently, the finding of a strong association between an exposure and an outcome generally implies a causal relationship.[44] In the hierarchy of clinical evidence (see table 9.3),[45, 46] the results of RCTs generally are considered to be the most robust (least susceptible to error).

Each experimental study (as well as some observational studies) begins with a prespecified, well-defined hypothesis, which is a conjectural explanation for a phenomenon stated in terms that can be systematically evaluated. For example: *the antioxidant vitamins C, E, and beta-carotene reduce the risk for myocardial infarction in women between the ages of 40 to 60.* Hypotheses are generated through the application of inductive reasoning to data from observational studies, whereas hypothesis-testing studies follow the principles of reasoning by deduction. They begin with a research protocol that prespecifies all features of the study, thus ensuring that the hypothesis, methodology, endpoints (outcomes to be measured), and statistical analysis are not determined *post hoc* based on what the data show or on a desire to present the data in a favorable light. The study endpoints are classified as "primary" (the main question to be addressed) and "secondary" (all other questions).

In an experimental study, random selection of subjects helps ensure that the sample (study subjects) will be representative of the population from which it was

Table 9.3
"Viewpoints" Described by Austin Bradford Hill for Considering Whether an Observed Association May Represent a Cause-and-Effect Relationship (1965)

Condition	Explanation
Strength of the association	The lung cancer death rate in cigarette smokers was nine to ten times higher than in non-smokers, and the rate in heavy cigarette smokers was twenty to thirty times as great. With an association of this strength, the presence of an undetected cause (that is, a confounder) was unlikely.
Consistency	The association is repeatedly observed by different persons in different places, circumstances, and times.
Specificity	A single cause is linked to a single effect. However, a lack of specificity does not negate the presence of a causal relationship, since lung cancer and other diseases can result from multiple causes.
Temporality	The putative cause must precede the observed effect.
Biological gradient	The severity of the effect increases monotonically (that is, linearly) with the extent of the exposure. Also known as a dose-response curve.
Plausibility	The suspected cause is biologically plausible, based on available knowledge. This is not a necessary condition, however, because (1) current biological knowledge might not be sufficient to provide an explanation, or (2) research that conflicts with established knowledge may prompt a reconsideration of accepted beliefs and theories.
Coherence	The observed association should be compatible with existing knowledge and theories. As with plausibility, however, an association that conflicts with established knowledge is not automatically false; it may prompt a reconsideration of accepted ideas.
Experiment	Is experimental or semi-experimental[a] evidence available? Such evidence may reveal "[t]he strongest support for the causation hypothesis."
Analogy	The association may bear comparison with another circumstance. For example, if exposure to a different environmental toxin produces a similar outcome, the current situation might be analogous.

a. Semi-experimental evidence. In this context, a before-and-after time trend analysis.[56]

Source: Hill AB. The environment and disease: association or causation? *Proc R Soc Med.* 1965;58:295-300.[55]

Note:

Hill expressed his ambivalence about the usefulness of the above by saying that "[n]one of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*." Actually, "temporality" (that is, the cause must precede the effect) is a *sine qua non* for causality, whereas the other conditions are neither necessary nor sufficient for determining whether an observed association is causal.[57] In a 2005 paper, Rothman and Greenland concluded that there are no causal criteria in epidemiology, and that "[c]ausal inference in epidemiology is better viewed as an exercise in measurement of an effect rather than as a criterion-guided process for deciding whether an effect is present or not."[58]

drawn, so that the findings will be generalizable. For a variety of reasons, however, a representative sample may not be obtained in practice. This can introduce selection bias, potentially providing rival explanations for the outcome or making the findings less generalizable. To protect against this threat, the experimental and control groups should be as equivalent as possible with respect to "critical demographic and clinical variables" (baseline characteristics), both known and unknown, that can influence the results.[47] Such an objective may be nearly achievable only through random allocation, in which each subject has an equal chance of being assigned to one group or the other. For the same reason that a series of coin tosses is likely to yield a balanced number of heads and tails, random allocation produces groups that, on average, are balanced with regard to all variables except for the exposure (for example, treatment) being tested.[1] Thus, we can be reasonably confident that any between-group difference in outcomes is due to the exposure. Even if the groups are not perfectly balanced, randomization minimizes the possibility that subjects with a particular characteristic will be systematically assigned to the exposure or the control.[48]

Throughout the study, it is important for both groups to have equivalent experiences except for the intervention being tested, an objective that may be best achieved by concealing information on treatment allocation from subjects and investigators (a mechanism often employed in RCTs). Known as *blinding*, this helps prevent the study from being influenced by the expectations of those who receive the intervention or assess its effects. In some cases, such as a surgical procedure or an experimental drug with a high rate of adverse events, blinding may not be feasible or sustainable. It still would be possible, however, to blind the investigators during the data analysis, with the intention of protecting the integrity of the results.

Replication

Regardless of the type of study, other researchers must be able to replicate the results in independently conducted investigations. Replication establishes the credibility of the findings, extends their generalizability to other populations and contexts, contributes to the "big picture" or more complete evidence regarding an intervention, and provides support for the development of theories.[49, 50] When the findings of an observational study are successfully replicated in a hypothesis-testing study, this means that a conclusion reached deductively agrees with the results of observation and inductive inference on which the hypothesis was based.[51] Thus, in a sense, the research process is brought full circle.

If a study cannot be replicated, this sometimes may be due to methodological, rather than biological, factors, including chance, error, bias, or insufficient statistical power (that is, a sample size too small to detect the presence of an effect), either in the initial study or in the follow-up investigation.[39, 50, 52] Therefore, it is essential to examine the totality of the evidence before deciding whether to abandon a particular line of research. For example, although most placebo-controlled trials have been unable to confirm observational results suggesting that antioxidants

1. Every published RCT includes a table showing the distribution of baseline characteristics in the study groups. In almost all cases, the groups are expected to be balanced with regard to variables that could influence the study outcome, except for random deviation.

reduce the risk for coronary events, the totality of laboratory, animal, and epidemiologic evidence seems to indicate that oxidative stress plays an important role in atherosclerosis. Consequently, investigators are trying to understand why placebo-controlled studies of vitamins C, E, and beta-carotene have failed to show a benefit.[53]

In another example, multiple observational studies conducted over a 20-year period found that hormone replacement therapy (HRT) reduced a woman's risk for CHD. But in 2002, the Women's Health Initiative (WHI), an RCT enrolling more than 16,000 postmenopausal women, reported a greater risk for CHD among women using HRT compared with those receiving a placebo.[39] Researchers have identified multiple sources of potential bias that could explain this discrepancy. In some earlier observational studies, the healthy user effect may have accounted for the putative beneficial outcome, while in the WHI, there was a high dropout rate and a substantial rate of premature unblinding among clinicians and participants, largely because of a need to manage vaginal bleeding. Unblinding could lead to behavioral changes (for example, participants in the control arm may lose their enthusiasm and withdraw from the trial) and detection bias, which occurs when subjects known to be using the experimental medication are monitored more closely than those whose treatment allocation remains concealed. This can result in the detection of outcomes that otherwise might have gone unrecognized.[39]

Causal Inference from Observational Studies

Although well-designed RCTs are the preferred means of inferring causation and replicating results, they may not be feasible because of ethical or practical considerations. Researchers cannot knowingly randomize subjects to a harmful exposure (for example, a carcinogen) or to an investigational treatment that may be less effective than an available treatment for the target disorder. Additionally, in some cases, the time or sample size needed to test a given intervention might make an RCT prohibitively expensive or otherwise impractical (although the potential effects of some exposures or interventions, such as environmental hazards or dietary change, may only be understood over the long term). Conversely, with rare diseases, there may be too few available subjects to produce credible results. Given any of these situations, observational studies may provide a more appropriate or realistic source of evidence, despite their limitations.

The case of cigarette smoking illustrates how nonexperimental (observational) studies may, at times, permit causal inference. In the first half of the 20th century, observational evidence showed that smokers were at least nine times more likely than nonsmokers to die of lung cancer.[54] The strength of this association in multiple studies led researchers to rule out confounding variables (for example, genetic, environmental) or limitations in study design as an alternative explanation and to conclude that there was a strong causal relationship between smoking and lung cancer, despite the lack of RCT data. In this context, Austin Bradford Hill, a British epidemiologist, published a landmark paper in 1965 describing considerations that may permit us to infer causation based on an observed association between two variables (see table 9.3).[55–58] As Hill noted, "[w]hat [these 'viewpoints'] can do, with greater or less strength, is to help us to make up our minds on the fundamental

Table 9.4
Evidence Types

Grade	Type of Evidence
I	Systematic reviews and meta-analyses of randomized controlled trials (with minimal or no publication bias)
II	Randomized controlled trial(s)
III	Controlled trials without randomization
IV	Cohort studies, preferably conducted by more than one center or research group
V	Case-control studies, preferably conducted by more than one center or research group
VI	Cross-sectional surveys
VII	Case series (reports on a series of patients with outcome of interest)
VIII	Opinions of respected authorities based on clinical experience; narrative reviews; descriptive studies and case reports; or reports of expert committees
IX	Anecdotal information

Source:
Note: VI to IX grades generate hypotheses; V to I grades infer causation.

question—is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?"[55]

THE IMPORTANCE OF MEASURABLE ENDPOINTS

In research, the endpoint (outcome) must be both classifiable and quantifiable. An outcome that directly measures how a patient feels, functions, or survives is called a clinical endpoint. Examples include death, loss of vision, heart attack, and onset of symptoms (for example, pain).[59] Because clinical endpoint studies often are long and costly, however, there is great interest in defining surrogate endpoints that can be quantified earlier and more conveniently.[59] A surrogate endpoint can be a laboratory measure (for example, cholesterol levels) or physical sign used as a substitute for a clinically meaningful endpoint. Changes induced by a therapy on a valid surrogate endpoint are expected to predict changes in a true clinical endpoint.[60]

While valid surrogate endpoints have the potential to reduce the size, duration, and cost of clinical trials, they are difficult to identify. Often, this is because scientists do not have a complete understanding of the causal pathway leading to the disease. For example, ventricular arrhythmias are associated with an almost fourfold increase in risk for cardiovascular death, and so it was thought that suppression of arrhythmia would predict a reduction in mortality. In the 1990s, however, three antiarrhythmic drugs approved by the Food and Drug Administration (FDA) were found to increase sudden cardiac death among subjects in placebo-controlled follow-up studies.[59] In addition, statistical methods for validating a surrogate endpoint usually require trials with sample sizes much larger than those needed to establish the clinical efficacy of a drug for FDA approval.[59] Consequently, if a biologically plausible surrogate endpoint has failed to predict a change in a clinical endpoint,

one possible explanation may be that the trial was too small, suggesting that a trial with greater power (a larger sample) might be able to detect the presence of an effect.

When considering the results of a study, we must ask whether the effect of an intervention is measured in terms of a clinical endpoint or a surrogate endpoint. If the latter, further inquiry is needed to assess whether the surrogate endpoint reliably predicts the clinical outcome of ultimate interest.

INTERNAL AND EXTERNAL VALIDITY

"Validity" means how well a study measures what it proposes to measure. There are two types of validity: internal and external. Internal validity is the degree to which researchers can be reasonably confident about the finding within a particular study, for example, the outcome is the result of the variable being tested. External validity means that the results can be generalized beyond the study. There are several "threats" to internal validity, each of which may provide an alternative explanation for the observed effect of an intervention. They include the presence of confounders, selection bias (faulty samples or lack of balance between exposure and control groups that was not controlled in the statistical analysis), events outside the study that may affect the subjects' behavior, maturation (natural changes in the course of a disease or in the subjects themselves, such as growing older), instrumentation (any change in the instrument or the researchers used to measure treatment effect), experimental mortality and attrition (not death, but withdrawal of subjects from the study, particularly if it is disproportionate, for example, more men, more women, more higher risk subjects, and so on), repeated testing (improved performance due to test-taking practice, as in psychological research), contamination (use of the experimental intervention by the control subjects), and regression (a statistical phenomenon in which a variable that is extreme when first measured will be closer to the mean when measured again, regardless of the intervention being tested).

Researchers also must be mindful of threats to external validity: for example, conditions in which the observed outcome is due to the experimental situation itself, and not to the exposure or intervention being tested. If this occurs, the results may apply only to the participants in an experimental study and may not be generalizable. Threats to external validity include the experimenter effect (intervention works because of the person implementing it), and the Hawthorne effect. Named after the location of a study conducted in the 1920s, this refers to a change in the subjects' behavior simply because they know that they are being observed.

TOOLS FOR INTERPRETING THE DATA

Location versus Spread

To understand the outcome of a study, the numerical data must be summarized in a way that is both succinct and meaningful. There are two important summary statistics used to describe the data: measures of location (or central tendency) and measures of spread. The term "measure of location" refers to the location of the central, or "typical," value for a set of data. The three commonly used measures of

location are the mean (average), median (middle value or 50th percentile), and mode (the value that occurs most frequently). Measures of spread (or variability around the central value) include range (highest value minus lowest value), interquartile range (the 75th percentile minus the 25th percentile), and standard deviation (a standardized measure of spread in relation to the mean). The decision about which measures of location and spread to use often depends on the presence of outliers (extreme values) or the skewness (asymmetry) of the distribution and on the type of data—for example, continuous, such as weight or height, or categorical, meaning data that are separable into mutually exclusive categories such as ethnicity or age-group.

Statistical Significance

"Significance" has a definition specific to science, especially statistical science. In common parlance, significance means importance or consequence, whereas statistical significance assesses the likelihood, or probability, that a result occurred by chance.

Evaluations of statistical significance attempt to determine the likelihood that a difference between exposure and control groups will be found when no difference actually exists (the absence of a between-group difference is called the "null hypothesis," and it is the default assumption unless contradicted by the data). The number most popularly used to express statistical significance is the "*p*-value" (probability value). A *p*-value is the probability of obtaining a statistical result at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. The lower the *p*-value, the less likely it is that the result occurred by chance; thus, the result is deemed more significant from a statistical perspective. By convention, a *p*-value less than 0.05 often is considered to be a basis for rejecting the null hypothesis.[61] But a *p*-value greater than 0.05 does not mean that there is no difference between exposure and control; it may merely indicate that a difference has not been satisfactorily demonstrated. Given this uncertainty and potential for error, it is more appropriate to regard the *p*-value as a tool for quantifying the relationship between a given statistical result and the null hypothesis, rather than as a basis for summarily rejecting the latter.[61]

Just as juries sometimes convict an innocent person of a crime or let a guilty person go free, statistical analysis may produce two basic kinds of errors: a Type I error, or false positive (that is, the probability of finding a difference between exposure and control when no difference actually exists), and its converse, a Type II error, or false negative (that is, the probability of failing to detect a between-group difference when one exists). Type I and Type II errors often are denoted by α and β, respectively.

The α-level refers to a significance level agreed upon before the data are collected. This differs from the *p*-value, which tells us how extreme the data are. An α-level of 5 percent, a commonly accepted standard (although it is sometimes more stringent), means a 5 percent chance (5 times out of 100) of finding a between-group difference when none exists—in other words, a 5 percent chance that the result was a random occurrence and that the null hypothesis has been mistakenly rejected, resulting in a Type I error. When a significance test yields a *p*-value that is the same

or lower than the predetermined α-level for a given study, researchers may conclude that the result would be unlikely to occur by chance if the null hypothesis were true. This is commonly termed "statistically significant."

In medicine, a Type I error often is considered relatively more serious, and thus more important to control, because the use of a treatment that we think is effective—but is actually no different from placebo or standard treatment—could result in great potential harm. It also can be argued that a Type II error (failing to recognize a meaningful clinical difference) may have similarly serious consequences. For instance, consider the possible implications of not identifying smoking as a risk factor for lung cancer, mistakenly diagnosing an HIV-infected patient as disease free, or erroneously concluding that Cox-2 inhibitors for pain relief do not increase the risk for cardiovascular events compared with other nonsteroidal anti-inflammatory drugs.[62]

The *p*-value is the most widely used statistical tool for drawing inferences from clinical studies,[63] yet it is subject to misuse or even abuse, in part because the demands of career advancement place researchers under pressure to establish the significance of their findings.[64] While this measure of significance has been dubbed "the almighty *p*-value,"[65] many scientists regret what some call the "cult of statistical significance,"[66] stating that it treats statistical models as if they were the reality, rather than a tool for the investigation of natural phenomena.[67] They also argue that emphasizing the *p*-value imposes a false dichotomy on the data (significant versus nonsignificant), leads to automated decision making instead of scientific reasoning, and fosters the view that those who draw different conclusions from the data may be motivated by "nonscientific" considerations.[61, 67, 68] This can inhibit the constructive controversy essential to scientific progress.

Although there is virtually no doubt about the usefulness of the *p*-value as a one-dimensional summary measure in decision making, careful judgment is needed to understand this powerful number adequately and use it appropriately.

Confidence Intervals

The confidence interval, which is calculated from a sample data set, presents a range of values that is likely to include the true effect (see figure 9.1). Unlike *p*-values, confidence intervals show us the results in terms of actual data (for example, mean values or units of measure).[69] They usually are expressed as 95 percent confidence intervals. In other words, we can feel 95 percent confident that the true effect (which is fixed) falls within this interval (which can vary), assuming the sample is properly selected.[70] If a confidence interval crosses the line of no effect ("0" for absolute risk or units of measure; "1" for ratios), it is customary to interpret this as a statistically nonsignificant result.[71] In such a case, however, we should try to understand whether the exposure and control really are equivalent, or whether a small sample size made the study incapable of detecting a meaningful difference.[71] Conversely, larger studies yield more precise confidence intervals, which can magnify the statistical significance of even minor between-group differences (for example, by producing a smaller *p*-value).

By presenting two numbers (an upper and a lower limit), confidence intervals broaden our focus to include an estimation of the size and strength of an intervention's effects. This takes us beyond statistical significance (a single number) and may better contribute to an assessment of the clinical significance of an intervention.[69]

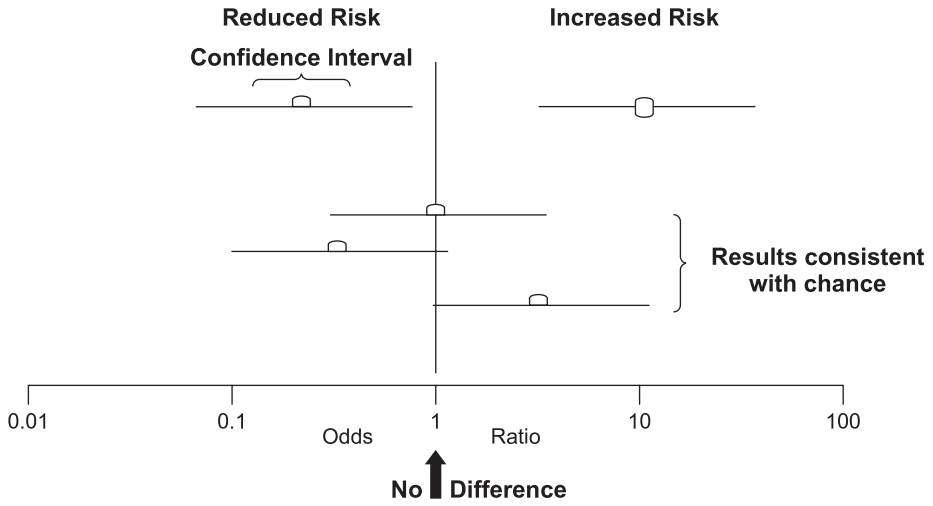

Figure 9.1 Concept of the Confidence Interval.

The margin of error, often mentioned in news reports of opinion polls, is a variant of the confidence interval. For example, if 63 percent (+/− 3 percent) of those surveyed support candidate X for mayor, this means that the true result probably lies within 3 percentage points above or below the reported result (that is, between 60 percent and 66 percent). Thus, the margin of error is usually defined as the radius of the confidence interval. Importantly, the margin of error accounts for random sampling error only, and not for error or bias from other sources, including poorly worded questions or false answers from those surveyed.

Clinical Significance

When scientists and reporters speak about the "significance" of a result, they often mean its statistical—not its clinical—significance. Unfortunately, the distinction may not be fully explained in the published research.

The concept of clinical significance was first proposed in 1984 as a way to answer the question of whether a therapy or treatment is effective.[72] Unlike statistical significance, clinical significance is not a "yes or no" concept. It involves standards of effectiveness set by a range of interested parties, including patients, physicians, and researchers.[73] For example, if a large study finds that regular screening for a particular type of cancer extends life by three days on average, the result may be statistically—but not clinically—significant.

Nuanced discussions of clinical significance are difficult to find in the mass media. Instead, relative risk reduction and the word "significant" are frequently used to emphasize the potential implications of a study, making a treatment, test, or procedure seem more effective than it is.[4]

ADDITIONAL USES OF THE DATA

Multiple Analyses

Once a study (either randomized or observational) is concluded, additional hypotheses may be tested based on the findings. These post hoc analyses often explore whether the effects of an intervention are consistent in subgroups of patients

defined by baseline characteristics such as age, sex, ethnicity, or comorbid conditions.[74] For example, a drug that produces no overall benefit might be found to have a favorable effect in a particular subgroup, implying that a narrowly defined set of patients may benefit from the intervention.[44] Investigators also use multiple analyses to assess the effect of an exposure or intervention on different outcome variables.

Multiple analyses are susceptible to Type I error (false positives), because the greater the number of explorations, the greater the probability that a statistically significant result will occur purely by chance.[75] Consider the statistically strong association between reserpine, an antihypertensive drug, and breast cancer, which was reported in 1974 by the Boston Collaborative Drug Surveillance Program, a large epidemiologic study.[76] This result suggested that reserpine might have a carcinogenic effect, possibly explainable by drug-induced prolactin secretion.[77] The association was not consistently replicated, however, and 30 years later, the lead investigator bravely confessed that it was "really a chance finding" due to thousands of comparisons involving hundreds of drug exposures. True to the spirit of science, he recognized the benefit of acknowledging and learning from our mistakes.[76–78]

In addition to subgroups and outcome variables, multiple analyses commonly look at different cutpoints or categorizations. For example, if three glasses of wine a day, or seven cups of coffee, are associated with a reduced risk for some disease, we should consider the possibility that such observations or relationships might be the result of (hidden) multiple analyses. Determining whether this is so may entail some detective work on the part of those who want to understand the numbers correctly.

Statistical methods can control for Type I error in multiple analyses, typically by lowering the p-value threshold for statistical significance, so that the standard becomes more stringent (for example, using $p<0.025$ as the threshold for two analyses, rather than the standard $p<0.05$ suitable for a single analysis). This makes it harder to conclude that there is a between-group difference when none exists. Ideally, however, investigators should prespecify, or at least limit, the number of subgroup analyses or hypotheses to be tested, and reporters should distinguish between "significant results of well-planned, powerful, sharply focused studies" and those from "fishing expeditions" [79] (that is, data dredging, which means the search for statistically significant patterns or relationships in large sets of data). This can lead to unexpected findings that are also subject to Type I error.

In addition to the matter of statistical significance, the biological plausibility of post hoc arguments requires some skepticism, since the human imagination seems capable of developing a rationale for most findings, however unanticipated.[80] Thus, while multiple analyses may be useful and sometimes lead to new lines of research, the results generally are subject to future validation before they can be accepted.

Systematic Reviews and Meta-Analyses

One way of appraising current knowledge on a particular research topic and identifying areas for future study is to perform a *systematic review* of the medical literature.

Systematic reviews begin with a detailed plan for selecting the studies to be analyzed (including a comprehensive literature search of multiple databases), assessing their methodological quality, and extracting and summarizing the data, with strategies to minimize bias at each step.[81, 82] There are two kinds of systematic reviews: qualitative and quantitative. The latter, also called a *meta-analysis*, uses statistical methods to synthesize the data from several independent studies and estimate a common pooled effect, if pooling is justified.[82] Not all systematic reviews lead to a meta-analysis, but every meta-analysis should be based on a systematic review.

A meta-analysis also can evaluate the diversity (heterogeneity) of results among studies, determine whether the pooled effect of the intervention agrees with the effect in individual studies, estimate the effect with greater precision, and test hypotheses that the individual studies did not address or addressed inadequately, possibly because they were underpowered.[82, 83] Some meta-analyses use individual patient data from each study, whereas others use study-level results (for example, risk difference or odds ratio) as the unit of measure. In general, the former are considered more statistically robust and superior.[84]

Despite their utility, reviews have a number of limitations. By including only published studies, they may be biased toward favorable outcomes (see "Publication Bias"), producing an inflated appraisal or estimate of a treatment effect or an association between variables.[83] Moreover, many investigators are hesitant to perform meta-analyses that combine data from studies with different hypotheses and characteristics.[82] And perhaps most important, our confidence in the conclusions of a review depends on a basic assumption: that none of the trials is systematically flawed. If there is bias in any of the trials, it will be reinforced in a review or perpetuated and entrenched in a meta-analysis.[83] This can be particularly problematic for meta-analyses of observational studies, since the latter generally are much more susceptible to bias than RCTs.

While systematic reviews and meta-analyses make an important contribution to the medical literature, they must be evaluated carefully. In particular, caution is needed when considering meta-analyses as a basis for clinical recommendations.[82] In an era of advanced statistical methods, sophisticated computational tools, and powerful search engines, meta-analyses have become increasingly popular in medicine. But unfortunately, they are not always carefully and appropriately performed (sometimes because of shortcomings in the available data). As one team of researchers commented, by way of Mark Twain, "'There are two things you should never watch being made. The law, and sausage.'" Perhaps meta-analysis should be added to the list."[85]

Nevertheless, pooling of data in a high-quality systematic review with meta-analysis can provide a more precise estimate of the effect of an intervention or risk factor, with conclusions that are more credible than those of individual studies (see table 9.4).[46]

ASSESSING RISKS AND BENEFITS

The points covered in the foregoing discussion bring us to the fundamental question that health-related research seeks to address: how can physicians, policy

planners, and patients quantify and weigh the risks and benefits of an intervention or other exposure?

When different ways of presenting the potential benefit of drug therapy to patients are compared, relative risk reduction is the most likely to prompt acceptance of treatment.[86, 87] But many physicians consider absolute risk reduction more useful for decision making—not only because relative numbers tend to magnify the treatment effect, but also because different baseline risk levels can be reduced by the same relative amount. In table 9.2, for example, absolute risk reductions of 2.0 percent, 24.0 percent, and 54.0 percent all translate into a relative risk of 33 percent and a relative risk reduction of 67 percent. Therefore, reports of relative risk reduction can be misleading unless accompanied by information on baseline risk. While absolute risk reduction may be the preferred metric for individual patient decisions, however, relative risk reduction is useful for statistical inferences because it is inclined to be more stable across patient groups.[88, 89]

Absolute numbers have limitations as well, because they do not generally account for the effects of time, age, or other variables that influence risk. A debate about how to measure risk is currently taking place over Tysabri, a powerful multiple sclerosis drug associated with a three-year risk of 0.019 percent for progressive multifocal leukoencephalopathy (PML), a neurodegenerative disease. Because the number of PML cases per 1,000 patients is slightly larger with each year of therapy, the collective event rate adjusted statistically for the effect of time is close to 1 percent, a threshold that some doctors and patients might find unacceptably high.[90] For this issue, an in-depth discussion of risks and benefits and their implications would be critical.

The Tysabri example, in which the rate of adverse events increases with duration of therapy (a consideration for all chronic conditions) or when an intervention is used in large numbers of patients, illustrates one reason why postmarketing surveillance is important. In other cases, the anticipated benefit may take decades to emerge. For example, a vaccine used in girls and young women to prevent infection with the genital human papillomavirus, which can cause cervical cancer, has been available for three years. While reports of adverse events include some that are serious, it is too soon to know whether immunization will reduce the future incidence of cancer.[91, 92] With other interventions, an effect (for example, reduced mortality) that emerges in preapproval studies may be lost over time in longer-term postmarketing trials, altering the trade-off between risks and benefits.

OTHER ISSUES THAT MAY AFFECT THE REPORTING OF RESULTS

Publication Bias

Publication bias refers to the tendency of researchers to submit, and editors to accept, studies for publication based on the strength and direction of the findings.[93] Specifically, if the findings are positive and statistically significant, or if they do not challenge accepted theory or practice, the likelihood of publication may be greater than otherwise. One meta-analysis comparing published and unpublished

studies of 12 antidepressants, for example, found a bias toward the publication of positive results. Moreover, the published studies tended to report an inflated effect size.[94] This problem, which is not unique to antidepressants, highlights the need to think critically about scientific data and to avoid hyperbole in the coverage of results.

Transparency

Transparency is defined as unlimited perpetual access to scientific research (raw data, published results) and to information about the competing interests (financial and nonfinancial) of scientists and study sponsors that may influence any phase of research, from design through publication.[95] The issue of transparency is widely discussed by experts, many of whom argue that it is the best assurance of progress, quality, and benefit to patients.[96, 97] Transparency raises many complex questions, including how to protect the private data of subjects enrolled in clinical trials and safeguard intellectual property rights. Efforts to address these issues are ongoing. One important step toward achieving transparency is the Public Library of Science (PLoS), an online resource founded in 2000 by Nobel Laureate Harold Varmus and colleagues Patrick Brown and Michael Eisen. Guided by the principle that science is a public resource, PLoS and other initiatives are dedicated to making research results freely available to scientists, physicians, patients, and students around the world.[98]

CONCLUSION

More than a century ago, Benjamin Disraeli observed that "there are three kinds of lies: lies, damned lies, and statistics."

While statistical science, like all human endeavors, is subject to error or misuse, it is the most effective way yet devised of predicting the probability (risk) of events under a given set of circumstances. Thus, when we hear conflicting reports about the risks or benefits of a drug or a food product, for example, it is important to examine the validity of the study and the strength of the statistical conclusion(s).

To do this, one does not have to be a trained scientist. The scientific method is based largely on common sense, and statistical thinking involves concepts that are accessible to all: an acceptance of chance and uncertainty, an appreciation of context, an ability to detect logical and factual flaws in information and ideas, and the realization that science is a fluid process whereby new empirical evidence is accumulated every day. The true spirit of science also requires a healthy skepticism, which means suspended judgment and the use of reason to evaluate the validity of research results. Science thrives on these qualities, because they lead to a search for knowledge and help ensure that the scientific method remains self-correcting.

Physicians, policy planners, patients, and the general public depend on the media for health-related information. As such, it is imperative that research results be reported accurately and responsibly to ensure that the reader (professional or lay person) will be able to assess the risks, the benefits, and the effectiveness of a drug or other product or exposure. If the study was designed well, and if the numbers were properly generated in response to the right questions, lies and damned lies most certainly would be revealed by statistics.[99]

Appendix 9.A.
Assessment of Study Design: What to Ask

Question	Chapter Section
Is the study • observational (epidemiologic)? • experimental?	**Study Designs: Observational versus Experimental**
If observational, is it • retrospective? • prospective?	
What target population does the study sample represent? • risk level (e.g., high, moderate, or low) • other characteristics	
What is the sample size?	
Is the study controlled? For example, • a randomized controlled trial (RCT)? • another type of experimental controlled trial? • an observational study with a control group (e.g., case-control)?	
Are the comparison groups balanced with respect to variables/characteristics that could influence or provide an alternative explanation for the results?	
If experimental, is the study • double blind (neither investigators nor subjects know the treatment allocation)? • single blind (investigators know the treatment allocation, but subjects do not)? • open label (no blinding is used)? Why? • blinded during the endpoint analysis only?	
How might each of the above potentially affect the conduct and outcome of the study?	
In a blinded study, was the blinding • sustained? • breached? Why? What effect could this have?	
How many events occurred in the • active treatment/exposure group? • control group?	
Have the study results been replicated or validated? If so, by • observational studies? • experimental studies?	
Did other studies report different or opposite results?	
Do the results indicate • association? • causation (see Austin Bradford Hill viewpoints, table 9.3, for guidance regarding observational studies)	**Association, Causation, Confounding, and Bias**
Is the finding plausible? • implausible (e.g., likely a random occurrence)?	

What are the • potential sources of bias or confounding? • alternative explanations for the results?	**Association, Causation, Confounding, and Bias**
Are the data skewed or do they have any peculiar properties?	**Tools for Interpreting the Data: Location versus Spread**
Is the "typical" value (measure of location) reported as • mean? • median? • mode?	
What does this suggest about the variability of the data?	
Is the spread (variability) reported as • range? • interquartile range? • standard deviation?	
Does the study use • clinical endpoints? • surrogate endpoints?	**The Importance of Measurable Endpoints**
Are the endpoints reported in terms of • absolute risk reduction? • relative risk reduction? • odds ratio?	**Risk**
Are the adverse effects reported in terms of • absolute risk? • relative risk?	
If the relative risk reduction is reported, • is the baseline risk also presented (or can you calculate it based on the number of events and the total number of subjects in the control group)? • is the event rate in the experimental group also presented (or can you calculate it based on the number of events and the total number of subjects in the experimental group)? • can you calculate the absolute risk reduction and the number needed to treat?	
Are results reported for • primary endpoints? • secondary endpoints? • unexpected findings (may indicate data-dredging[a])? • multiple analyses?	**Tools for Interpreting the Data: Significance**
Are the results • statistically significant (what is the *p*-value)? • clinically significant?	
What does the sample size suggest about the potential for • statistical significance (e.g., is it too small to detect a between-group difference)? • clinical significance (e.g., is it large enough so that an effect with little or no clinical significance may be statistically significant)?	
Is the 95 percent confidence interval reported? Is it • narrow?	

- wide?
- across the "no effect" line?

What does the width/direction of the confidence interval suggest about
- the relation of the estimated effect to the "true" effect (i.e., is the result precise)?
- the statistical significance of the effect?

With multiple analyses, are they — **Additional Uses of the Data: Multiple Analyses**
- prespecified (e.g., in the protocol)?
- post hoc?

Are the results of multiple analyses
- biologically/clinically plausible?
- statistically significant (with a *p*-value adjusted for multiple analyses or the number of analyses reported)?

Do the benefits of the intervention outweigh the risks? — **Assessing Risks and Benefits**

What is the baseline risk (control event rate)? What does it say about the patients who may benefit from the intervention? Are they at
- high risk (i.e., events are fairly common in untreated individuals)?
- low risk (i.e., events are rare in untreated individuals)?
- moderate risk?

Which of the following measures of risk are reported:
- baseline risk?
- absolute risk reduction?
- relative risk?
- relative risk reduction?

If the absolute risk reduction is not reported, can it be calculated (baseline risk × relative risk reduction)?

Is this a postmarketing study? If so, do the results alter previous risk-benefit assessments concerning the intervention?

What are the potential threats to — **Internal and External Validity**
- internal validity?
- external validity?

Are the results reported in language that — **Additional Questions (not discussed in chapter)**
- conveys a misleading or exaggerated impression of their importance?
- reinforces the therapeutic misconception?[b]
- supports a particular viewpoint?

Have the results been reported in
- a peer-reviewed professional journal?[c]
- an opinion piece by an expert (e.g., in a professional journal, a mass media publication)?
- a book?
- a presentation at a scientific meeting?
- an expert interview?

Is information accessible to journalists and the public on

- who funded the study?
- the researchers' industry affiliations and other potential conflicts of interest (financial and nonfinancial)?

Are titles or headlines

- overstated and misleading?
- accurate in terms of the information reported?

Do editorials, commentaries, or other opinion pieces

- use multiple sources?
- present diverse viewpoints in a fair manner?
- place the information in context?

Does the research involve

- an approved use of the drug?[d]
- an off-label use?[d]

Do reports about laboratory or animal studies make it clear that the results may not translate into benefit for humans?

---

a. Data-dredging: the search for statistically significant patterns or relationships in large sets of data. Data-dredging has a negative connotation and is sometimes called a "fishing expedition" (see Multiple Analyses).

b. Therapeutic misconception: "the defining purpose of clinical research is to produce "generalizable knowledge, regardless of whether the subjects enrolled in the trial may potentially benefit from the intervention under study or from other aspects of the clinical trial."[37] While ethical principles and stringent safeguards (e.g., informed consent) protect the rights of research subjects and prevent intentional harm, the therapeutic misconception exists when study participants do not understand the distinction between the purpose of clinical practice (to treat the patient) and clinical research (to gain knowledge).[37]

c. Reporting the results of scientific research: Publication in a peer-reviewed professional journal is the most important way of making research results known. While this does not guarantee the accuracy or validity of the results, professional journals are a more reliable source of information than other forms of dissemination.

d. Approved use: In the United States, medical products and devices are approved and regulated by the Food and Drug Administration (FDA). Although FDA-approved prescription drugs are designated for specific purposes, they may sometimes be beneficial for diseases or conditions not described in the authorized labeling. Doctors must use sound medical evidence and clinical judgment when prescribing a drug for an off-label purpose. Drug marketing/advertising for off-label use is never permitted.

REFERENCES

1. ScienceDaily. Highly active compound found in coffee may *prevent* colon cancer. http://www.sciencedaily.co*m/releases/*2003/10/031015031251.htm. Published October 15, 2003. Accessed October 19, 2009.
2. Furtado M. The surprising ingredient causing weight gain. YAHOO!© HEALTH. http://health.yahoo.com. Published July 27, 2009. Accessed July 30, 2009.
3. Devlin K. Three cups of brewed coffee a day 'triples risk of hallucinations.' Telegraph.co.uk. http://www.telegraph.co.uk/science/science-news. Published January 14, 2009. Accessed September 27, 2009.

4. Beck M. Correlation, causation and what Welsh rugby means for the Pope. *The Wall Street Journal*. http://blogs.wsj.com/health/2009/01/27/. Published January 27, 2009. Accessed October 16, 2009.
5. von Bubnoff A. Numbers can lie: good today, but how about tomorrow? *Los Angeles Times*. Health page. http://articles.latimes.com/2007. Published September 17, 2007. Accessed February 2, 2009.
6. Angell M. *Science on Trial. The Clash of Medical Evidence and the Law in the Breast Implant Case*. London: W. W. Norton & Company; 1996.
7. Narasimhan MG. Controversy in science. *J Biosci.* 2001;26:299–304.
8. Stocking AH, Holstein LW. Manufacturing doubt: journalists' roles and the construction of ignorance in a scientific controversy. *Public Understand Sci.* 2009;18:23–42.
9. Ceccarelli L. Manufactured controversy: Science Studies Network (SSNet) Seminar; October 20, 2008; University of Washington, Seattle, WA. http://depts.washington.edu/ssnet/handout_102008.pdf. Accessed October 28, 2009.
10. Wilby P. The media's addiction to controversy can seriously damage your health. *The Guardian*. http://www.guardian.co.uk. Published August 13, 2008. Accessed April 29, 2009. Correction published on August 21, 2008. *The Guardian*.
11. Ransohoff DF, Ransohoff RM. Sensationalism in the media: when scientists and journalists may be complicit collaborators. *EFF Clin Pract.* 2001;4:185–188.
12. Greer K. Journalists forced to sensationalize science to attract readers. *Tennessee Journalist*. http://tnjn.com/2008/mar/27/journalists-forced-to-sensatio/. Published March 27, 2008. Accessed September 28, 2009.
13. Shuchman M, Wilkes MS. Medical scientists and health news reporting: a case of miscommunication. *Ann Intern Med.* 1997;126:976–982.
14. Rothman SM, Rothman DJ. Marketing HPV vaccine. Implications for adolescent health and medical professionalism. *JAMA*. 2009;302:781–786.
15. Jackson T. News skews health priorities, study claims. *BMJ.* 2003;327:688.
16. Sharma V, Dowd D, Swanson DS, Slaughter AJ, Simon SD. Influence of the news media on diagnostic testing in the emergency department. *Arch Pediatr Adolesc Med.* 2003;157:257–260.
17. Schield M. Statistical literacy and liberal education at Augsburg College. *Peer Rev* 2004. http://www.aacu-edu.org/peerreview/index.cfm. Accessed May 14, 2009.
18. University of California Museum of Paleontology. "Misconceptions about science:" understanding science. http://undsci.berkeley.edu/teaching/misconceptions.php. Retrieved May 19, 2009.
19. Martin A, Faraone SV, Henderson SW, et al. Mission statement: advancing the science of pediatric mental health and promoting the care of youth and their families. *J Am Acad Child Adolesc Psychiatry.* 2008;47:1.
20. Norton JD. Science and uncertainty. *Synthese.* 1994;99:3–22.
21. Moore DS. Statistics among the liberal arts. *J Am Stat Assoc.* 1998;93:1253–1259.
22. Campbell SK. *Flaws and Fallacies in Statistical Thinking*. Mineola, NY: Dover Publications, Inc.; 2002.
23. Feynman RP. Uncertainty in science. In: *The Meaning of It All: Thoughts of a Citizen-Scientist*. New York: Basic Books: 2005.
24. Wild CJ, Pfannkuch M. Statistical thinking in empirical enquiry. *Intl Stat Rev.* 1999;67:223–265.
25. Pellegrin P. Physics. In: Brunschwig J, Lloyd GER, eds. *The Greek Pursuit of Knowledge*. Cambridge, MA: The Belknap Press of Harvard University Press; 2000: xiii, 282–300.
26. Kuhn JE, Greenfield MVH, Wojtys EM. A statistics primer: hypothesis testing. *Am J Sports Med.* 1996;24:702–703.

27. Gould SJ. Evolution as fact and theory (abridged). In: Vetter HF, ed. *Speak Out Against the New Right.* Boston, MA: Beacon Press; 1982.
28. Parascandola M, Weed DL. Causation in epidemiology. *J Epidemiol Community Health.* 2001;55:905–912.
29. Shugan SM. Causality, unintended consequences and deducing shared causes. *Market Sci.* 2007;26:731–741.
30. Vandenbroucke JP. The history of confounding. *Soz-Präventivmed.* 2002;47: 216–224.
31. Kestenbaum B. *Epidemiology and Biostatistics. An Introduction to Clinical Research*. New York: Springer Publications; 2009.
32. Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. 2nd ed. Sudbury, MA: Jones and Bartlett Publishers; 2006.
33. Mayhall CG. *Hospital Epidemiology and Infection Control*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2004.
34. Bernstein PL. *Against the Gods. The Remarkable Story of Risk*. New York: John Wiley & Sons, Inc.; 1998.
35. Wikipedia: The Free Encyclopedia. Number needed to treat. http://en.wikipedia.org/wiki/Number_needed_to_treat. Accessed September 28, 2009.
36. Trial AIDS vaccine cuts infection risk 31 percent. CBS News. http://www.cbsnews.com/stories/2009/09/24/health/main5334317.shtml. Published September 24, 2009. Accessed October 18, 2009.
37. Henderson GE, Churchill LR, Davis AM, et al. Clinical trials and medical care: defining the therapeutic misconception. *PloS Med.* 2007;4(11):xxe324. doi:10.1371/journal.pmed.0040324. http://www.ncbi.nlm.nih.gov/. Accessed October 25, 2009.
38. Garbe E, Suissa S. Issues to debate on the Women's Health Initiative (WHI) study. Hormone replacement therapy and acute coronary outcomes: methodological issues between randomized and observational studies. *Hum Reproduction.* 2004;19:8–13.
39. Hess DR. Retrospective studies and chart reviews. *Respir Care.* 2004;49:1171–1174.
40. Humphrey L, Chan BKS, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med.* 2002;137:273–284.
41. Simon SD. Understanding the odds ratio and the relative risk. *J Andrology.* 2001;22:533–536.
42. Pouwels S, Lalmohamed A, Leufkens B, et al. Risk of hip/femur fracture after stroke: a population-based case-control study. *Stroke.* 2009;40:3281–3285.
43. Edelson E. Stroke doubles risk of hip, thigh fractures. Preventive measures urgently needed, researcher says. *U.S. News & World Report*. http://health.usnews.com/articles/health/healthday/2009/08/06/stroke-doubles-risk-of-hip-thigh-fractures.html. Published August 6, 2009. Accessed August 23, 2009.
44. Moyé LA. *Multiple Analyses in Clinical Trials: Fundamentals for Investigators*. New York: Springer Publications; 2003.
45. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000;342:1887–1892.
46. Mahid SS, Hornung CA, Minor KS, Turina M, Galandiuk S. Systematic reviews and meta-analysis for the surgeon scientist. *Br J Surg.* 2006;93:1315–1324.
47. Salmond SS. Randomized controlled trials: methodological concepts and critique. *Orthop Nurs.* 2008;27:116–122.
48. Thisted RA. What is a p-value? Department of Statistics, University of Chicago. http://galton.uchicago.edu/~thisted/Distribute/pvalue.pdf. Published May 25, 1998. Retrieved September 30, 2009.

49. Burns N, Grove SK. *Understanding Nursing Research. Building an Evidence-Based Practice*. 3rd ed. Philadelphia, PA: WB Saunders; 2002.
50. Ioannidis JPA. Large scale evidence and replication: insights from rheumatology and beyond. *Ann Rheum Dis.* 2005;64:345–346.
51. Boyer CB. *The History of the Calculus and Its Conceptual Development*. New York: Dover Publications, Inc.; 1949.
52. Boffetta P, McLaughlin JK, La Vecchia C, Lipworth L, Blot WJ. False-positive results in cancer epidemiology: a plea for epistemological modesty. *J Natl Cancer Inst.* 2008;100:988–995.
53. Steinhubl SR. Why have antioxidants failed in clinical trials? *Am J Cardiol.* 2008;101[suppl]:14DS-19D.
54. Parascandola M. Two approaches to etiology: the debate over smoking and lung cancer in the 1950s. *Endeavour.* 2004;28:81–86.
55. Hill AB. The environment and disease: association or causation? *Proc R Soc Med.* 1965;58:295–300. http://www.edwardtufte.com/tufte/hill.
56. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins: 2008.
57. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott-Raven; 1998.
58. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health.* 2005;95:S144–S150.
59. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996;125:605–613.
60. Temple RJ. A regulatory authority's opinion about surrogate endpoints. In: Nimmo WS, Tucker GT, eds. *Clinical Measurement in Drug Evaluation*. New York: J. Wiley; 1995. Quoted in Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996;125:605–613.
61. Weinberg CR. It's time to rehabilitate the p-value. *Epidemiology.* 2001;12:288–290.
62. Graham DJ. Cox-2 inhibitors, other NSAIDs, and cardiovascular risk: the seduction of common sense. *JAMA.* 2006;296:1653–1656.
63. Lesaffre E. Use and misuse of the p-value. *Bull NYU Hosp Joint Dis.* 2008;66:146–149.
64. Drotar D. Thoughts on establishing research significance and preserving scientific integrity. *J Pediatr Psychol.* 2008;33:1–5.
65. Peck CC. The almighty p-value *or* the significance of "significance." *Present Concepts Intern Med.* 1971;4(11):1021–1024.
66. Ziliak ST, McCloskey DN. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor: University of Michigan Press; 2008.
67. Goodman SN. Of p-values and Bayes: a modest proposal. *Epidemiology.* 2001;12:295–297.
68. Editors. The value of p. *Epidemiology*. 2001;12:286.
69. du Prel J-B, Hommel G, Röhrig B, Blettner M. Confidence interval or p-value. *Dtsch Arztebl Int.* 2009;106:335–339.
70. Curran-Everett D. Explorations in statistics: confidence intervals. *Adv Physiol Educ.* 2009;33:87–90.
71. Simon SD. Interpreting negative studies. *J Andrology.* 2001;22:13–16.
72. Jacobson NS, Follette WC, Revenstorf D. Psychotherapy outcome research: methods for reporting variability and evaluating clinical significance. *Behav Ther.* 1984;15:336–352.
73. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol.* 1991;59:12–19.

74. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med.* 2007;357:2189–2194.
75. Young SS, Bang H, Oktay K. Cereal-induced gender selection? Most likely a multiple testing false positive. *Proc R Soc B.* 2009;276:1211–1222.
76. Lawson DH. Detection of drug-induced disease. *Br J Clin Pharmacol.* 1979;7:13–18.
77. Ross RK, Paganini-Hill A, Krailo MD, Gerkins VR, Henderson BE. Effects of reserpine on prolactin levels and incidence of breast cancer in postmenopausal women. *Cancer Res.* 1984;44:3106–3108.
78. Shapiro S. Looking to the 21st century: have we learned from our mistakes, or are we doomed to compound them? *Pharmacoepidemiol Drug Safety.* 2004;13:257–265.
79. Garcia LV. Escaping the Bonferroni iron claw in ecological studies. *Oikos.* 2004;105:657–663.
80. Ware JH. The National Emphysema Treatment Trial—how strong is the evidence? [Perspective] *N Engl J Med.* 2003;348:2055–2056.
81. Jadad A. *Randomised Controlled Trials: A User's Guide*. London: BMJ Books; 1998.
82. White PF, Watcha MF. Has the use of meta-analysis enhanced our understanding of therapies for postoperative nausea and vomiting? *Anesth Analg.* 1999;88:1200–1202.
83. Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ.* 1997;315:617–619.
84. Teramukai S, Matsuyama Y, Mizuno S, Sakamoto J. Individual patient-level and study-level meta-analysis for investigating modifiers of treatment effect. *Jpn J Clin Oncol*. 2004;34:717–721.
85. Hamer RM, Simpson PM. SAS® Tools for Meta-Analysis. http://www2.sas.com/proceedings/sugi27/p250-27.pdf. Accessed September 27, 2009. SAS® Paper No.250–27.
86. Hux JE, Naylor CD. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? *Med Decis Making.* 1995;15:152–157.
87. Carling CLL, Kristoffersen DT, Montori VM, et al. The effect of alternative summary statistics for communicating risk reduction on decisions about taking statins: a randomized trial. *PLoS Med.* 2009;6(8):e1000134. doi:10.1371/journal.pmed.1000134. www.plosmedicine.org. Accessed September 23, 2009.
88. Crewdson J. Statistics misleading, some doctors say: relative numbers skew benefit to look much larger. *Chicago Tribune*. March 15, 2002: 21.
89. Sorensen L, Gyrd-Hansen D, Kristiansen IS, Nexoe J. Laypersons understanding of relative risk reductions: randomized cross-sectional study. *BMC Med Inform Dec Making.* 2008;8:1–7. http://www.biomedcentral.com.
90. Winstein K. Medicine's dangerous guessing game. *Wall Street Journal*. http://online.wsj.com. Published September 9, 2009. Accessed September 9, 2009.
91. Slade BA, Leidel L, Vellozzi C, et al. Postlicensure safety surveillance for quadrivalent human papillomavirus recombinant vaccine. *JAMA.* 2009;302:750–757.
92. Haug C. The risks and benefits of HPV vaccination. *JAMA.* 2009;302:795–796.
93. World Health Organization. Publication bias in clinical trials due to statistical significance or direction of trial results. RHL: The WHO Reproductive Health Library. http://apps.who.int/rhl/education/MR000006_butlerpa_com/en/index.html. Accessed September 23, 2009.
94. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med.* 2008;358:252–260.

95. PLoS Medicine Editors. An unbiased scientific record should be everyone's agenda. *PLoS Med.* 2009;6(2):e1000038. doi:10.1371/journal.pmed.1000038. http://www.plosmedicine.org/. Accessed September 23, 2009.
96. Freedman DA, Pettiti DB. Hormone replacement therapy does not save lives. Comments on the Women's Health Initiative. *Biometrics.* 2005;62:918–920.
97. Public Library of Science. PLoS core principles. http://www.plos.org/about/principles.html. Accessed September 23, 2009.
98. Eisen MB, Brown PO, Varmus HE. PLOS Medicine: a medical journal for the internet age. *PLOS Med.* 2004;1(1):e31. http://www.ncbi.nlm.nih.gov. Accessed October 25, 2009.
99. Begley S. Lies, damned lies and . . . *Newsweek*. http://www.newsweek.com. Published July 12, 2008. Accessed September 22, 2009.

CHAPTER 10

Evidence-Based Public Health: A Fundamental Concept for Public Health Practice

Christopher M. Maylahn, MPH, Ross C. Brownson, PhD, and Jonathan E. Fielding, MD, MPH, MA, MBA

INTRODUCTION

The 30-year gain in life expectancy in the United States that occurred in the 20th century is a notable achievement.[1] Much of this increase is due to the provision of safe water and food, sewage treatment and disposal, tobacco use prevention and cessation, injury prevention, control of infectious diseases through immunization and other means, and other population-based interventions.[2] Many public health strategies have been proven to reduce risk among individuals and entire communities, yet have not been carried out. Further gains in health and longevity, as well as a reduction in health disparities, would result from widespread adoption of interventions for which there is evidence of effectiveness.[3–7]

The design and use of public health actions that are effective in promoting health and preventing disease underlie the growing field of evidence-based public health (EBPH), which emerged in the 1990s to improve the *practice* of public health. The definition of this term has evolved over time, as discussed later in the chapter. It generally is viewed as "the development, implementation, and evaluation of effective programs and policies in public health through application of principles of scientific reasoning, including systematic uses of data and information systems, and appropriate use of behavioral science theory and program planning models."[4, p. 4] Ideally, public health practitioners always should incorporate scientific evidence in selecting and implementing programs, developing policies, and evaluating progress.[8, 9]

Society pays a high opportunity cost when interventions that yield the highest health return on an investment are not implemented.[10] In practice, decisions about how to address a health issue or problem often are based on circumstances favoring a pragmatic, politically feasible approach, lack systematic planning and fail to incorporate the latest evidence about what works. These concerns were noted two decades ago when the Institute of Medicine (IOM) determined that decision making in public health

---

This chapter is reprinted, with permission, from the *Annual Review of Public Health*, Volume 30

often is driven by "crises, hot issues, and concerns of organized interest groups."[11, p. 4] Greater use of EBPH has numerous direct and indirect benefits, including access to more and higher quality information on what works, a higher likelihood of successful programs and policies being implemented, greater workforce productivity, and more efficient use of public and private resources.[4, 12, 13] Yet, in a recent survey of 107 U.S. public health practitioners, fewer than 60 percent of programs in their agencies were "evidence-based," defined as being guided by the most current evidence from peer-reviewed research.[14] Barriers to implementing EBPH include the political environment, and deficits in relevant and timely research, information systems, resources, leadership, and the required competencies.[8, 15–17]

Several concepts are fundamental to achieving a more evidence-based approach to public health practice. First, scientific information is needed about programs and policies most likely to be effective in promoting health.[4, 12, 18, 19] Evaluation research is needed to generate sound evidence and, optimally, builds on multiple studies conducted among different populations. An array of effective interventions is now available from numerous sources, including the *Guide to Community Preventive Services*,[20, 21] the *Guide to Clinical Preventive Services*,[22] Cancer Control PLANET,[23] the National Registry of Evidence-based Programs and Practices,[24] and Public Health Partners.[25] Much of this information is now widely accessible on the Internet. Second, to translate science into actual practice, the application of evidence-based interventions from the peer-reviewed literature must consider *features* of a specific "real-world" environment [4, 26, 27] and make needed adjustments, without compromising what makes the intervention work. This entails defining the processes that lead to evidence-based decision making. A combination of scientific evidence and values, resources, and context enters into decision making. Finally, widescale dissemination of proven interventions must occur more consistently at state and local levels.[28]

The focus of this chapter is on the importance of EBPH for state and local public health departments because of their responsibilities to assess public health problems, develop appropriate programs or policies, and ensure that they are implemented effectively in states and local communities.[11, 29] Government public health agencies are viewed as the primary force in organizing and mobilizing public health practice in most communities.[30] The challenges these agencies confront in applying EBPH principles in local public health practice are daunting as resources shrink, new health issues emerge, and scientific evidence to address them is slow to materialize. Often, the study aims and populations involved in published research do not match the characteristics and circumstances of a target community in which action is needed, making it problematic to assume a similar impact will occur. When a policy or public health program is launched, the intervention "dose" must be sufficient to achieve the desired change. The chapter is organized into four sections that describe: (1) the concepts and principles underlying EBPH; (2) analytic tools to enhance the adoption of evidence-based decision making; (3) dissemination and implementation in public health practice; and (4) challenges and opportunities for more widespread use of EBPH, especially through state and local health departments. It is adapted from our article published in the 2008 *Annual Review of Public Health*.[31]

TENETS OF EVIDENCE-BASED PUBLIC HEALTH

EBPH was first defined by Janicek in 1997 as the "conscientious, explicit, and judicious use of current best evidence in making decisions about the care of communities

and populations in the domain of health protection, disease prevention, health maintenance and improvement."[32] Two years later, Brownson et al. described a six-stage process that enables practitioners to take a more evidence-based approach to decision making [4, 8] and provided practical guidance on how to choose, carry out, and evaluate evidence-based programs and policies. That same year, Glasziou and Longbottom posed a series of questions to enhance uptake of EBPH (for example, Does this intervention help alleviate this problem?) and identified 14 sources of high-quality evidence.[5] Kohatsu et al. broadened earlier definitions to include the perspectives of community members, fostering a more community-centered approach.[26]

An excellent resource for public health practitioners is a glossary of EBPH terminology published in 2004 by Rychetnik et al. [33] that draws on the published literature, experience gained over several years of analysis of the topic, and discussions with students. The glossary is useful in defining terms, some of which were just emerging at the time, so that they may be interpreted and applied in a consistent manner.

In summarizing these various aspects of EBPH, key characteristics include the following:

- Making decisions based on the best available peer-reviewed evidence (both quantitative and qualitative research);
- Using data and information systems systematically;
- Applying program planning frameworks (that often have a foundation in behavioral science theory);
- Engaging the community in assessment and decision making;
- Conducting sound evaluation; and
- Disseminating what is learned to key stakeholders and decision makers.

In little more than a decade, EBPH has become a fundamental concept for public health practice. Yet, the field and terminology continue to evolve. The term "evidence-informed" has been proposed to describe a process for critical appraisal of primary research studies and systematic reviews to inform effective public health practice.[34] It depends on the use of quality checklists and other tools to identify rigorous studies with valid conclusions for implementation and assessment. Within the general area of community research, what constitutes knowledge and evidence is hotly debated, or even if the notion of evidence in evaluating community interventions applies at all.[35]

Adherence to EBPH principles and methods requires a skill called "evidence based decision making"—that is, that decisions about the use of public health resources are based on the best available scientific evidence about what works and take into account stakeholder values, available resources, and the context of the problem to be addressed (see figure 10.1).[3, 4, 33, 36, 37] Knowledge of the evidence base, strong communication skills, common sense, and political acumen often are necessary. Furthermore, understanding the context in which many most public health policies and programs are implemented—in states and local communities—demands an examination of the opportunities and challenges arising in these settings.

Defining Evidence

In its most basic form, evidence involves "the available body of facts or information indicating whether a belief or proposition is true or valid."[38] The idea of

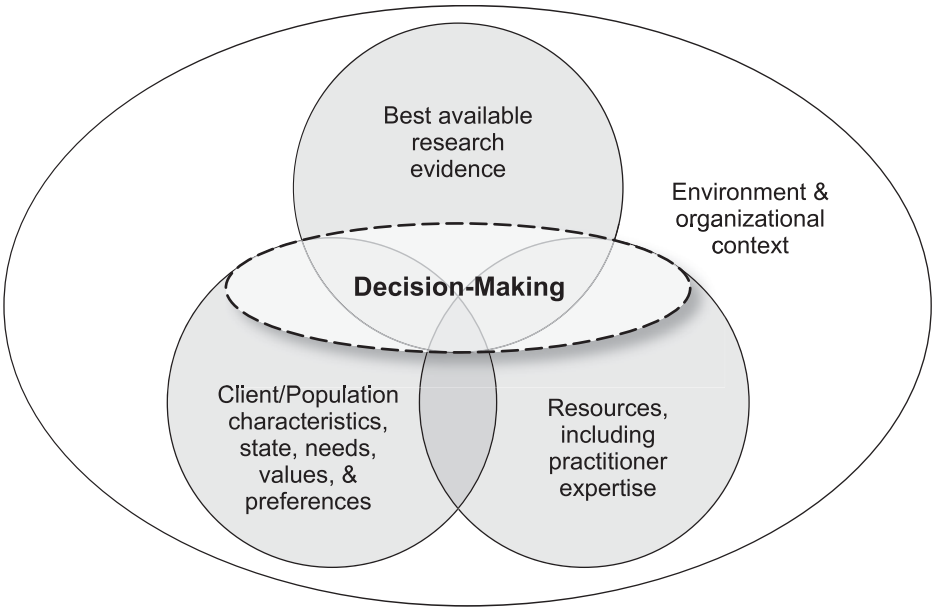

Figure 10.1 Domains that influence evidence-based decision making. (*Source:* From Spring et al.)

evidence often derives from legal settings in Western societies. Here, evidence is manifest as stories, witness accounts, police testimony, expert opinions, and forensic science.[35] The everyday use of the word evidence carries high expectations, implying that there can be no doubt if one has evidence. For a public health professional, evidence is some kind of data—including quantitative information from epidemiologic studies, results of program or policy evaluations, or qualitative data—to help make judgments or decisions. Public health evidence usually is the result of a complex cycle of observation, theory, and experiment.[39, 40] However, the value of evidence is in the eye of the beholder—its utility may vary by type of stakeholder.[41] The demand for high-quality evidence to justify a particular public health action is growing among many stakeholder groups as competition for limited resources intensifies. Yet, evidence is usually imperfect and, as noted by Muir Gray, "The absence of excellent evidence does not make evidence-based decision making impossible; what is required is the best evidence available, not the best evidence possible."[3]

There are three types of scientific evidence for public health practice.[4, 8, 33] Type 1 evidence defines the causes of diseases and the magnitude, severity, and preventability of risk factors and diseases. It suggests that "*something* should be done" about a particular disease or risk factor. Type 2 evidence describes the relative impact of specific interventions that do or do not improve health, adding "*specifically,* this should be done."[4] There are different sources of Type 2 evidence (see table 10.1). These categories emphasize the "weight of evidence" and a wider range of considerations beyond efficacy. Four categories are defined within a typology of scientific evidence for decision making: evidence-based, efficacious, promising,

and emerging interventions. Type 3 evidence shows how and under what contextual conditions interventions were implemented and how they were received, thus identifying "*how* something should be done."[33] The context of an intervention is important in understanding and applying evidence from a particular study. It is more difficult to find good type 3 studies. Most of the published studies overemphasize internal validity (for example, well-controlled efficacy trials) while giving sparse attention to external validity (for example, the translation of science to the various circumstances of practice).[42, 43] Public health practitioners need this kind of information to design community-wide programs and to adjust interventions to settings different from where the evidence was obtained.

Understanding the Context for Evidence

Numerous authors have written about the role of context in informing evidence-based practice,[12, 33, 41, 44–48] but there is little consensus on its definition. One useful definition of context specifies that information is needed to adapt and implement an evidence-based intervention in a particular setting or population.[33] The context for type 3 evidence includes five overlapping domains that are included in an ecological model of health promotion (see table 10.2).[34, 49, 50] First, there are characteristics of the target population for an intervention such as their education level and health history. Next, interpersonal variables provide important clues about context. For example, a person with a family history of cancer might be more likely to undergo cancer screening. Third, organizational variables are important. For example, whether an agency is successful in carrying out an evidence-based program will be influenced by its capacity (for example, a trained workforce, agency leadership). Fourth, social norms and culture are known to shape many health behaviors. Finally, larger political and economic forces affect context. For example, a high rate for a certain disease may influence a state's political will to address the issue in a meaningful and systematic way. These contextual issues are being examined more fully in the new "realist review," which is a systematic review process that seeks to examine not only whether an intervention works but also *how* interventions work in real-world settings.[51]

Triangulating Evidence

Triangulation involves the accumulation of evidence from a variety of sources to gain insight into a particular topic,[52] and often combines quantitative and qualitative data,[4] using multiple methods of data collection or analysis in a complementary manner to determine points of commonality or disagreement.[53, 54] Although quantitative data provide an excellent opportunity to determine how variables are related for large numbers of people, these data provide few clues to understand why these relationships exist. Qualitative data, on the other hand, can explain quantitative findings. There are many examples of the use of triangulation of qualitative and quantitative data to evaluate health programs and policies including AIDS prevention programs,[55] occupational health programs and policies,[56] and chronic disease prevention programs in community settings.[57] At the state and local level, it often is necessary to use both approaches for evaluation because either one, by itself, is not sufficient.

Table 10.1
Typology for Classifying Interventions by Level of Scientific Evidence

Category	How Established	Considerations for Level of Scientific Evidence	Data Source Examples
Evidence–based	Peer review via systematic or narrative review	Based on study design and execution	Community Guide
		External validity	Cochrane reviews
		Potential side benefits or harms	Narrative reviews based on published literature
		Costs and cost-effectiveness	
Effective	Peer review	Based on study design and execution	Articles in the scientific literature
		External validity	Research-tested intervention programs
		Potential side benefits or harms	Technical reports with peer review
		Costs and cost-effectiveness	
Promising	Written program evaluation without formal peer review	Summative evidence of effectiveness	State or federal government reports (without peer review)
		Formative evaluation data	Conference presentations
		Theory-consistent, plausible, potentially high-reach, low-cost, replicable	
Emerging	Ongoing work, practice-based summaries or evaluation works in progress	Formative evaluation data	Evaluability assessments[a]
		Theory-consistent, plausible, potentially high-reach, low-cost, replicable	Pilot studies
		Face validity	NIH CRISP data base
			Projects funded by health foundations

a. A "pre-evaluation" activity that involves an assessment is an assessment prior to commencing an evaluation to establish whether a program or policy can be evaluated and what might be the barriers to its evaluation.
Source: Brownson, RC et al. *Evidence-Based Public Health.* New York: Oxford University Press, 2003.

Table 10.2
Contextual Variables for Intervention Design, Implementation, and Adaptation

Category	Examples
Individual	Education level
	Basic human needs[a]
	Personal health history
Interpersonal	Family health history
	Support from peers
	Social capital
Organizational	Staff composition
	Staff expertise
	Physical infrastructure
	Organizational culture
Socio-cultural	Social norms
	Values
	Cultural traditions
	History
Political and economic	Political will
	Political ideology
	Lobbying and special interests
	Costs and benefits

a. Basic human needs include food, shelter, warmth, safety.[49]

Source: Adapted from Ciliska, D et al. *An Introduction to Evidence Informed Public Health and a Compendium of Critical Appraisal Tools for Public Health Practice.* Hamilton, Ontario, Canada: Collaborating Centre of Methods and Tools, McAlaster University, 2008.

Audiences for EBPH

There are four overlapping user groups for EBPH. The first includes public health practitioners with executive and managerial responsibilities who want to know the scope and quality of evidence for alternative strategies, such as different program or policy approaches. In practice, however, the range of options is limited, especially because funds from federal, state, or local sources usually are earmarked for a specific purpose. Still, the public health practitioner should carefully review the evidence for alternative ways to achieve the desired health goals. The next user group is policy makers at varying geopolitical levels who must decide how to allocate the public resources for which they are stewards. This group has the additional responsibility of making policies on controversial public issues. A third audience is the stakeholders who receive or will be affected by an intervention—that is, the public, especially those who vote, as well as interest groups formed to support or oppose specific policies, such as the legality of abortion, whether the community water supply should be fluoridated, or whether adults must be issued handgun licenses if they pass background checks. The final user group is composed of researchers who may be responsible for evaluating the impact of a specific policy or program, or using evidence to answer research questions.

Similarities and Differences between EBPH and Evidence-Based Medicine

While the concept of evidence-based practice is well established in numerous disciplines including psychology,[58] social work,[59] and nursing,[60] it is probably best established in medicine. The doctrine of evidence-based medicine (EBM) started in 1992,[61] not long before the concept of EBPH was introduced. Its origins can be traced back to the seminal work of Cochrane, who noted that many medical treatments lacked scientific effectiveness.[62] A basic tenet of EBM is to deemphasize unsystematic clinical experience and place greater emphasis on evidence from clinical research. Although the formal terminology of EBM is relatively recent, its concepts are embedded in earlier efforts, such as the Canadian Task Force for the Periodic Health Examination [63] and the *Guide to Clinical Preventive Services*.[64]

Important distinctions can be made between evidence-based approaches in medicine and public health. First, the type and volume of evidence differ. Medical studies of pharmaceuticals and procedures often rely on randomized controlled trials of individuals, the most scientifically rigorous of epidemiologic studies. In contrast, public health interventions usually rely on cross-sectional studies, quasi-experimental designs, and time-series analyses. These studies sometimes lack a comparison group and require more caveats to interpret the results. EBPH has borrowed the term "intervention" from clinical disciplines, implying a specific and discrete set of actions. In public health, however, there is seldom a single "intervention," but rather a program that involves a blending of several change strategies at once. Large community-based trials can be more expensive to conduct than randomized experiments in a clinic. Population-based studies generally require a longer time period between intervention and outcome. For example, a study on the effects of smoking cessation on lung cancer mortality would require decades of data collection and analysis. Contrast that with treatment of a medical condition (for example, an antibiotic for symptoms of pneumonia), which is likely to produce effects in days or weeks, or even a surgical trial for cancer with endpoints of mortality within a few years.

ANALYTIC TOOLS AND APPROACHES TO ENHANCE THE USE OF EBPH

Typically, public health practitioners must answer a variety of questions to improve the health of the communities they serve. As a start, they will want to know the size of a particular public health problem and whether there are effective interventions to address it. Because context is important in judging whether a given strategy will work, they will need information about the community, target population characteristics, and resources available to influence the health of people living there. They must decide whether a particular program or policy is worth doing (for example, is it better than alternatives?) and whether it will provide a satisfactory return on investment, measured in monetary terms or in health impacts. Several analytic tools and planning approaches are available to help practitioners answer such questions.

Public Health Surveillance

Public health surveillance is a critical tool for those using EBPH. It involves the ongoing systematic collection, analysis, and interpretation of specific health data,

closely integrated with the timely dissemination of these data to those responsible for preventing and controlling disease or injury.[65] Public health surveillance systems should have the capacity to collect and analyze data, disseminate data to public health programs, and regularly evaluate the effectiveness of the use of the disseminated data.[66] For example, measuring the prevalence of elevated levels of lead (a known toxicant) in blood in the U.S. population was used to justify eliminating lead from paint and then gasoline, and then to document the effects of these actions.[67] In tobacco control, agreement on a common metric for tobacco use enabled comparisons of the prevalence of smoking across the states and an early recognition of the doubling and then tripling of the rates of decrease in smoking in California after passage of its Proposition 99,[68] and then a quadrupling of the rate of decline in Massachusetts compared with the other 48 states.[69]

Systematic Reviews and Evidence-Based Guidelines

A systematic review is a synthesis of results from many studies on a particular topic. Reading a good review can be one of the most efficient ways to become familiar with a body of knowledge about a health issue and how to address it. The use of explicit methods (that is, decision rules) applied in a consistent manner limits bias and reduces chance effects, thus providing more reliable results upon which to make decisions.[70] One of the most useful sets of reviews for public health interventions is the *Guide to Community Preventive Services* (the *Community Guide*),[21, 71] which provides an overview of current scientific research literature through a well-defined, rigorous method. The *Community Guide* helps public health practitioners to (1) discover what interventions have been evaluated and what have been their effects; (2) select interventions of proven effectiveness; and (3) estimate their costs and likely impact. Several authors have provided "checklists" for assessing the quality of a systematic review article.[72–74] A good systematic review should allow the practitioner to understand the local contextual conditions that contribute to successful implementation.[75]

Economic Evaluation

Economic evaluation assesses the relative value of alternative expenditures on public health programs and policies. In cost-benefit analysis, all of the costs and consequences of the decision options are valued in monetary terms. More often, the economic investment associated with an intervention is compared with the health impacts, such as cases of disease prevented or years of life saved. This technique, cost-effectiveness analysis (CEA), shows the relative value of alternative interventions (that is, health return on dollars invested).[76] CEA has become an increasingly important tool for researchers, practitioners, and policy makers. The data needed in this type of analysis, however, may be difficult to obtain, especially for programs and policies designed to improve health.[77, 78]

Health Impact Assessment

Health impact assessment (HIA) seeks to estimate the probable impact of a policy or intervention on the health of the population that is carried out in nonhealth

sectors, such as agriculture, transportation, and economic development.[79] Recently, Dannenberg et al. [80] reviewed 27 HIAs completed in the United States from 1999 to 2007. Topics studied ranged from policies about living wages and afterschool programs to projects about power plants and public transit. An excellent illustration is the assessment of a Los Angeles living wage ordinance.[81] Researchers used estimates of the effects of health insurance and income on mortality to project and compare potential mortality reductions attributable to wage increases and changes in health insurance status among workers covered by the ordinance. Estimates demonstrated that the health insurance provisions of the ordinance would have a much larger health benefit than the wage increases, thus providing valuable information for policy makers who may consider adopting living wage ordinances in other jurisdictions or modifying existing ordinances.

Participatory Approaches

Community-based participatory research is a relatively new paradigm that integrates education and social action to improve health and reduce disparities.[82] Participatory approaches actively involve community members in research and intervention projects [83–85] to enhance the conduct and utility of EBPH.[26] Practitioners, academicians, and community members collaboratively define issues of concern, then develop intervention strategies and evaluate their impact. This approach relies on input from stakeholders who are individuals or agencies that have a vested interest in the issue at hand.[86] In the development of health policies, for example, policy makers are especially important stakeholders.[87] Stakeholders should include those who potentially would receive, use, and benefit from the program or policy being considered. Participatory evaluation strategies may involve community groups and coalitions in identifying potential outcomes or indicators of change.[82] Participatory approaches also may present challenges in adhering to EBPH principles, especially in deciding which approaches are most appropriate for addressing a particular health problem.[88]

Active Ingredients

EBPH relies on the transferability of evidence about effective interventions to new community settings. Practitioners need to identify the most important components or "active ingredients" of an intervention, which are the essential elements that produce the desired results, and how context matters. Often, constraints require some modification of the original intervention. In these situations there is an inherent tension between fidelity (maintaining the original program design) and reinvention (changes needed for replication or adoption in a new setting or for a different population).[89] Close monitoring of the intervention is called for to be sure that it incorporates the active ingredients and that the same intermediate and long-term outcomes documented in the original research are occurring.

CHALLENGES IN THE DISSEMINATION AND USE OF EBPH

The field of EBPH has steadily matured and is now recognized by most public health professionals as an important tool in public health. Yet, the dissemination

and use of effective intervention strategies, which is critically important in improving public health practice, remains a significant challenge.[90, 91] Drawing on wide experience in the clinical world, dissemination of evidence-based clinical guidelines using passive methods (for example, publication of consensus statements in professional journals, mass mailings) has been largely ineffective, resulting in only small changes in the uptake of a new practice.)[92] Similarly, single-source prevention messages generally are ineffective.[93] The dissemination and use of an evidence-based program necessitates time-efficient approaches, ongoing training, and placing a high organizational value on research-informed practice.[94] Furthermore, translation of research to practice among organizations, practitioner groups, or the general public is likely to occur in stages,[95] suggesting that the decision to adopt, accept, and utilize an innovation in EBPH is a process rather than a single act. This section addresses specific challenges standing in the way of wider dissemination of evidence-based practices and greater use of evidence-based decision making to improve the public's health.

Organizational Culture

EBPH often relies on "evidence champions" who are willing to challenge the status quo within an organization and promote scientifically proven approaches to solve health problems. New ways of making decisions and novel management strategies may be needed to achieve the desired changes. Government institutions, including public health agencies, are key users of EBPH. Yet they are not known for their organizational or budgetary flexibility, making it difficult to quickly mount a new program, alter an established way of doing business, or promote new policies in other organizations. These agencies typically are bound to rigid civil service and union-bargained requirements about how staff can be hired, remunerated, evaluated, and terminated, as well as how money can be spent. In Los Angeles County, for example, the pay scale for nutritionists and health educators is so low that it is difficult to attract even entry-level individuals. Once hired by the county, these professionals often are attracted to higher paying administrative positions that do not use their primary expertise.

An organizational climate that supports change is required for innovation,[96] but it is often not found in public health agencies. Rigid personnel systems often make it difficult to keep up with rapidly evolving technology. For example, in many health agencies, there are no suitable job classifications for a health economist or for a Web designer, making it virtually impossible to hire at competitive salaries. Within a hierarchical bureaucracy, few incentives exist to press superiors for changes in programs or approaches even when the evidence is compelling. Unlike in some private sector organizations that encourage risk taking and provide substantial monetary rewards for success, most public sector organizations have a culture that discourages "out of the box" thinking and entrepreneurship.[97] The tendency to continue doing what has been done in the past is a powerful impediment to change. In many bureaucracies, when change does occur, it is usually in small incremental steps. Continuing past practices requires less effort than working through all of the implications of a different approach based on newer evidence.

Leadership

The attitude toward EBPH among agency leadership is important because it helps to determine the organizational culture and use of finite resources. In a survey of 152 city and county health departments in the United States, one of the main predictors of strong public health system performance was the attention of organizational leadership to the science base, quality, and performance.[98] Even public health leaders who understand and embrace EBPH, however, have challenges in choosing and implementing innovative approaches. How should they choose priority opportunities for programs and policies among all those recommended based on evidence reviews? As in clinical medicine, there are more recommendations than are practical for any department to introduce. What criteria should leaders consider when selecting among options? Some worthy considerations include population-attributable disease or illness burden, preventable fraction, relative cost-effectiveness, skills of key staff, prior experience with other approaches, opportunities for leverage through partnerships with other stakeholders, and consistency with an agency's strategic plan, goals, and objectives.

Politics

Having good scientific evidence often is insufficient to convince policy makers (for example, Congress, state governors, boards of county supervisors, city councils) to initiate evidence-based changes.[99, 100] Researchers rely on experimental and observational studies to test specific hypotheses in a deliberate and systematic way [101, 102] and their influence derives from having specialized knowledge. On the other hand, policy making can happen quickly and is built on generalized knowledge and demands from stakeholders.[92, 100] Policy makers have to sell, argue, advocate, and get reelected in light of their available political capital.[78] In some cases, the evidence for a particular action does not necessarily lead to policy change.[103, 104] Public health agencies often face obstacles from other stakeholders in proposing or implementing new evidence-based practices.

Programmatic and policy changes often result in "winners" and "losers" who can be at odds in the EBPH process.[105] A contractor who financially supports an elected decision maker may have more clout than the agency, regardless of the merits. Public health agencies, because of their mission to improve the population's health, often seek to advance measures that expand the power and reach of government, raising objections from those who want less government. For example, in the debates surrounding public smoking ban proposals, public health agencies were forced to combat arguments that the smoking bans simply were a way for the government to limit personal freedoms. Overcoming this resistance often requires that public health leaders create coalitions of partners that extend well beyond the realm of public health.

The prevailing political ideology may be contrary to what science recommends, such as for water fluoridation or needle-exchange programs. The federal government recently has allowed federal funds to be used for needle-exchange programs, *two decades* after they were shown to be effective. In other cases, those without a background in scientific methods may be skeptical that a systematic review process yields a better idea of what to do about a problem than simply following advice of an individual they trust, even when the trusted advice is at variance with the best

evidence.[106] Lack of skills in putting together coalitions of partners who support a particular EBPH intervention also can reduce the likelihood of convincing policy makers. Public health action that requires legislation may face opposition from local jurisdictions that could jeopardize its passage.

Public health leaders occasionally encounter situations in which the political will to implement a particular intervention exists before there is evidence to support it. A prime example is the Drug Abuse Resistance Education (D.A.R.E.) program, which is the most widely used school-based drug use prevention program in the United States reaching over 70 percent of elementary school-age children.[107] The program costs approximately $130 per student (in 2004 dollars) to implement. Systematic reviews of methodologically sound D.A.R.E. program evaluations have shown the program to be ineffective.[108]

Funding

Another challenge to implementing EBPH is the need to adhere to the requirements of the funding agencies. Most public health funding at all levels of government is categorical and restricted as to how the money may be spent—that is, "hardening of the categories,"[109] and limits the flexible use of funds to implement new evidence-based programs. Public health leaders are beginning to recognize the benefits to program integration and have articulated principles to enhance integration efforts.[110] In addition, appropriating legislation or voter initiatives may contain explicit language about restrictions, which in turn often is influenced by key stakeholders. For example, in California, no more than 20 percent of funding coming from Proposition 99, a statute in the Tobacco Tax and Health Protection Act enacted in 1988, can be used for antitobacco education in schools and communities.[111] We are not aware of any legislation or executive branch guidance that limits expenditures to evidence-based recommendations or requires that these be used whenever available. More governmental agencies appear to be referencing the best sources of evidence-based recommendations, including the *Community Guide*,[21] as important inputs into the state and local planning processes.[112] In many states, funding of tobacco control efforts is tied to the use of evidence-based strategies for tobacco use prevention and control.

Workforce Competencies

While the formal concept of EBPH is relatively new, the underlying skills are not. For example, reviewing the scientific literature for evidence or evaluating a program intervention are skills often taught in graduate programs in public health or other academic disciplines and are building blocks of public health practice. Yet, the skills and competencies of the public health workforce to carry out EBPH functions vary tremendously. The emphasis on principles of EBPH is not uniformly taught in all the disciplines represented in the public health workforce. For example, a public health nurse is likely to have had less training in how to locate the most current evidence and interpret alternatives than an epidemiologist. A recently graduated health educator with a master's degree in public health is more likely to have gained an understanding of the importance of EBPH than an environmental health specialist holding a bachelor's degree. Fewer than half of public health workers have any formal training in a public health discipline such as epidemiology or health education.[113] An even smaller percentage

of these professionals have formal graduate training from a school of public health or other public health program. Currently, it appears that few public health departments have made continuing education about EBPH mandatory. Consequently, public health workers must rely on routine, on-the-job training to gain the skills required to apply the principles of EBPH.

FUTURE OPPORTUNITIES

The United States spends nearly $30 billion annually on health-related research,[114] but only a small portion of these expenditures is dedicated to research relevant to the practice of public health. Nonetheless, evidence for addressing a number of priority public health problems now exists. Unfortunately, the translation from research to clinical or community applications often occurs only after a delay of many years.[4, 28, 115] Several actions are required to accelerate the production of new evidence and the adoption of evidence-based interventions to protect and improve health in the United States.

Expanding the Evidence Base

The growing literature on the effectiveness of preventive interventions in clinical and community settings [21, 22] does not provide equal coverage of health problems. For example, the evidence for how to increase immunization levels is much stronger than how to prevent HIV infection or reduce alcohol abuse. A greater investment of resources to expand the evidence base is therefore essential. Even in cases in which we have interventions of proven effectiveness, the populations in which they have been tested often do not include subpopulations with the greatest disease and injury burden. Expanding the base of evidence requires a reliance on well-tested conceptual frameworks, especially those that pay close attention to dissemination and use of EBPH. For example, RE-AIM helps program planners and evaluators to pay explicit attention to *R*each, *E*fficacy/*E*ffectiveness, *A*doption, *I*mplementation, and *M*aintenance,[116, 117] concepts that are consistent with evidence-based interventions.

Overcoming Barriers to Dissemination and Use of EBPH

More knowledge is needed on effective mechanisms to translate evidence-based practice to public health settings. Several important questions deserve answers:

- Why have some types of evidence languished while others have been quickly adopted?
- What are cost-effective strategies to disseminate and encourage use of EBPH?
- How can funding agencies accelerate the replication and adaptation of evidence-based interventions in a variety of settings and populations?
- What specific processes best integrate community health assessment and improvement activities into health system planning efforts?
- How can we harness new tools, such as the Internet, to improve intervention effectiveness and dissemination?
- What changes in organizational culture that promote innovation and adoption of EBPH are feasible?
- How can we increase attention on external validity in the production and systematic reviews of evidence?

Engaging Leadership

Leadership is probably the most important determinant influencing the adoption of EBPH as a core part of public health practice.[98] This includes an expectation that decisions will be made on the basis of the best science, needs of the target population, and what will work locally. Baker et al. [118] examined the role of training in government workers in EBPH and the subsequent use of these methods in public health practice. Through qualitative studies, they found that the implementation of evidence-based processes strongly depends on leadership support. Encouraging leaders to embrace evidence-based public health as the underpinning of programs in their communities is vital to program success. Further research on the link between leadership and performance—and how to strengthen local public health agency capacity—was recommended by Scutchfield et al. [98] to improve local public health practice. In some cases, additional funding may be required, but usually having the will to change (rather than dollars) is the key to success. Use of EBPH should be incorporated as part of performance reviews for key public health leaders and as part of explicit goals and objectives for all program directors.

Strengthening the Public Health Workforce

To address EBPH competencies, training programs have been developed in the United States for public health professionals in state health agencies,[14, 118] local health departments and community-based organizations,[119, 120] and similar programs have been developed in other countries.[121–123] Some programs show evidence of effectiveness.[14, 120] The most common format uses didactic sessions, computer labs, and scenario-based exercises, taught by a faculty team with expertise in EBPH. The reach of these training programs can be increased by emphasizing a train-the-trainer approach.[121] Other formats have been used, including Internet based self-study,[119, 124] CD-ROMs,[125] distance and distributed learning networks, and targeted technical assistance. Training programs may have greater impact when delivered by "change agents" who are perceived as experts yet share common characteristics and goals with trainees.[126] A commitment from leadership and staff to lifelong learning is also an essential ingredient for success.[127] Training programs to address EBPH competencies should take into account principles of adult learning. These issues recently were articulated by Bryan et al.,[128] who highlighted the need to (1) know why the audience is learning; (2) tap into an underlying motivation to learn by the need to solve problems; (3) respect and build upon previous experience; (4) design learning approaches that match the background and diversity of recipients; and (5) actively involve the audience in the learning process.

The need for a competent public health workforce is consistent with numerous IOM reports calling for public health workers to be adept in policy making, communication, science translation and other advocacy skills.[11, 29, 129] Well-prepared public health professionals are essential for an effective public health system in the United States, yet this implies an adequate supply of skilled and competent workers. Thus, investments in strengthening capacity and competencies of the public health workforce are needed, as well as research to understand how this can best be accomplished. For example, the importance of epidemiologic leadership and support for

public health practice led to the development of Competencies for Applied Epidemiologists in Governmental Public Health Agencies, which was designed to improve the practice of epidemiology in public health agencies. Koo et al. [130] found that these competencies map to domains other than epidemiology and also to activities and experiences outside the academic environment. The field of workforce development will be enhanced by linking such competencies to curricula in a collaborative effort between schools of public health, and local and state health departments.

More practitioner-focused training is needed on the rationale for EBPH, how to select interventions, how to adapt them to particular circumstances, and how to monitor their implementation. The Task Force on Workforce Development has recommended that the essential public health services [131] be used as a framework to build the basic, cross-cutting, and technical competencies required to address public health problems. The inclusion of EBPH principles and needed competencies [4, 132] would enhance this framework. Also, because many of the health issues needing urgent attention in local communities will require the involvement of other organizations (for example, nonprofit groups, hospitals, employers), their participation in training efforts is essential.

Enhancing Accountability for Public Expenditures

Public funds should be targeted to support evidence-based strategies. Grants made by public health agencies to outside organizations should contain language explicitly requiring the use of such strategies, when they exist, to accomplish grant objectives. While the science base for many topics is still evolving, it is irresponsible not to use existing evidence in the design and implementation of proven public health interventions. Evaluations of such efforts can thus contribute to a better understanding of what works in different settings. Simultaneously, the adoption of EBPH by the public health system as a whole and its impact on the community's health should be tracked. A central criterion in the accreditation of public health departments, soon to be implemented,[133] must be the use of best evidence in every effort to improve health and health equity.

Understanding How to Better Use EBPH to Address Disparities

To what extent does the use of specific evidence-based approaches reduce disparities while improving overall current or future health? For many interventions, there is not a clear answer to this question. Despite the Healthy People 2010 goal of eliminating health disparities, recent data show large and growing differences in disease burden and health outcomes between high- and low-income groups.[134] Most of the existing intervention research has been conducted among higher income populations, and programs focusing on elimination of health disparities have often been short-lived.[135] Yet, in both developed and developing countries, poverty is strongly correlated with poor health outcomes.[136] When enough evidence exists, systematic reviews should focus specifically on interventions that show promise in eliminating health disparities.[137, 138] Policy interventions hold the potential to influence health determinants more broadly and could significantly reduce the growing disparities across a wide range of health problems.[139]

CONCLUSION

The most significant problems affecting population health result from a complex interplay of behavioral choices, environmental influences, and genetics. Modifiable behavioral risk factors, including tobacco use, poor diet and physical inactivity, and alcohol consumption are now the leading actual causes of death in the United States.[140] Changing these behaviors and their social determinants will require effective use of public health actions for which strong evidence exists, as well as new knowledge about how to reduce the continuing health disparities between subgroups. Unintentional and intentional injuries often result in many years spent with disability or lost years of life and have large social and economic consequences. Decreasing the global burden of injuries is among the main challenges facing the public health sector in the next century.

To meet these challenges, more research is needed to find evidence for what works and how to apply this evidence in local communities. While the library of effective interventions to reduce tobacco use is filled with success stories—and many of the recommended actions have been put into place across the country—there is less certainty about how to stem the rise in obesity; for example, the sheer size of the epidemic and the dire consequences if it is not stopped calls out for evidence-based solutions that can be broadly applied. Other health threats will inevitably surface that require public health action. We must remain vigilant in monitoring the health of our communities to identify such threats and be ready to apply proven interventions to reduce risk. Underlying most of the disease and injury burden are modifiable aspects of the physical and social environments. As we look forward, to maximize public health impact and minimize health disparities, EBPH should include interventions that address such factors as the built environment, poverty, and education attainment.

The successful implementation of EBPH in public health practice is both a science and an art. The science is built on epidemiologic, behavioral, and policy research showing the size and scope of a public health problem and which interventions are likely to be effective in addressing the problem. The art of decision making often involves knowing what information is important to a particular stakeholder at the right time. Unlike solving a math problem, significant decisions in public health must balance science and art, since rational, evidence-based decision making often involves choosing one alternative from among a set of rational choices. By applying the concepts of EBPH outlined in this chapter, decision making and, ultimately, public health practice can be improved.

REFERENCES

1. Centers for Disease Control and Prevention. Ten great public health achievements: United States, 1900–1999. *MMWR*. 1999;48:241–243.
2. Centers for Disease Control and Prevention. *Public Health in the New American Health System*. Atlanta, GA: Centers for Disease Control and Prevention; 1993. Discussion Paper March, 1993.
3. Muir Gray JA. *Evidence-Based Healthcare: How to Make Health Policy and Management Decisions*. New York: Churchill Livingstone; 1997.
4. Brownson RC, Baker EA, Leet TL, Gillespie KN. *Evidence-Based Public Health*. New York: Oxford University Press; 2003.

5. Glasziou P, Longbottom H. Evidence-based public health practice. *Aust N Z J Public Health*. 1999;23:436–440.
6. McMichael C, Waters E, Volmink J. Evidence-based public health: what does it offer developing countries? *J Public Health (Oxf)*. 2005;27:215–221.
7. Fielding JE, Briss PA. Promoting evidence-based public health policy: can we have better evidence and more action? *Health Aff (Millwood)*. 2006;25:969–978.
8. Brownson RC, Gurney JG, Land G. Evidence-based decision making in public health. *J Public Health Manag Pract*. 1999;5:86–97.
9. McGinnis JM. Does proof matter? why strong evidence sometimes yields weak action. *Am J Health Promot*. 2001;15:391–396.
10. Fielding JE. Where is the evidence? *Annu Rev Public Health*. 2001;22:v–vi.
11. Institute of Medicine. *Committee for the Study of the Future of Public Health: The Future of Public Health*. Washington, DC: National Academies Press; 1988.
12. Hausman AJ. Implications of evidence-based practice for community health. *Am J Community Psychol*. 2002;30:453–467.
13. Kohatsu ND, Melton RJ. A health department perspective on the Guide to Community Preventive Services. *Am J Prev Med*. 2000;18:3–4.
14. Dreisinger M, Leet TL, Baker EA, Gillespie KN, Haas B, Brownson RC. Improving the public health workforce: evaluation of a training course to enhance evidence-based decision making. *J Public Health Manag Pract*. 2008;14:138–143.
15. Anderson J. "Don't confuse me with facts . . .": evidence-based practice confronts reality. *Med J Aust*. 1999;170:465–466.
16. Baker EL, Potter MA, Jones DL, et al. The public health infrastructure and our nation's health. *Annu Rev Public Health*. 2005;26:303–318.
17. Haynes B, Haines A. Barriers and bridges to evidence based clinical practice. *BMJ*. 1998;317:273–276.
18. Black BL, Cowens-Alvarado R, Gershman S, Weir HK. Using data to motivate action: the need for high quality, an effective presentation, and an action context for decision-making. *Cancer Causes Control*. 2005;16(Suppl 1):15–25.
19. Curry S, Byers T, Hewitt M, eds. *Fulfilling the Potential of Cancer Prevention and Early Detection*. Washington, DC: National Academies Press; 2003.
20. Briss PA, Brownson RC, Fielding JE, Zaza S. Developing and using the Guide to Community Preventive Services: lessons learned about evidence-based public health. *Annu Rev Public Health*. 2004;25:281–302.
21. Zaza S, Briss PA, Harris KW, eds. *The Guide to Community Preventive Services: What Works to Promote Health?* New York: Oxford University Press; 2005.
22. *Guide to Clinical Preventive Services*. 3rd ed. Washington, DC: Agency for Healthcare Research and Quality, 2005. http://www.ahrq.gov/clinic/gcpspu.htm. Periodic updates. Accessed October 11, 2005.
23. Cancer Control PLANET. Links resources to comprehensive cancer control: The National Cancer Institute; The Centers for Disease Control and Prevention; The American Cancer Society; The Substance Abuse and Mental Health Services; The Agency for Healthcare Research and Quality. http://cancercontrolplanet.cancer.gov/index.html. Published 2008. Accessed May 26, 2008.
24. SAMHSA's National Registry of Evidence-based Programs and Practices. US Department of Health and Human Services. Substance Abuse & Mental Health Services Administration. http://www.nrepp.samhsa.gov/. Published 2008. Accessed August 16, 2008.
25. Public Health Partners. Partners in Information Access for the Public Health Workforce. http://phpartners.org

26. Kohatsu ND, Robinson JG, Torner JC. Evidence-based public health: an evolving concept. *Am J Prev Med*. 2004;27:417–421.
27. Green LW. Public health asks of systems science: to advance our evidence-based practice, can you help us get more practice-based evidence? *Am J Public Health*. 2006;96:406–409.
28. Kerner J, Rimer B, Emmons K. Introduction to the special section on dissemination: dissemination research and research dissemination: how can we close the gap? *Health Psychol*. 2005;24:443–446.
29. Institute of Medicine. *The Future of the Public's Health in the 21st Century*. Washington, DC: National Academies Press; 2003.
30. American Public Health Association. *Local Public Health Practice: Trends and Models*. Washington, DC: American Public Health Association; 2000.
31. Brownson RC, Fielding JE, Maylahn CM. Evidence-based public health: a fundamental concept for public health practice. *Annu Rev Public Health*. 2009;30:175–201.
32. Jenicek M. Epidemiology, evidence-based medicine, and evidence-based public health. *J Epidemiol Commun Health*. 1997;7:187–197.
33. Rychetnik L, Hawe P, Waters E, Barratt A, Frommer M. A glossary for evidence based public health. *J Epidemiol Community Health*. 2004;58:538–545.
34. Ciliska D, Thomas H, Buffett C. *An Introduction to Evidence Informed Public Health and a Compendium of Critical Appraisal Tools for Public Health Practice*. Hamilton, Ontario, Canada: Collaborating Centre of Methods and Tools, McMaster University; 2008.
35. McQueen DV. Strengthening the evidence base for health promotion. *Health Promot Int*. 2001;16:261–268.
36. Spring B. OBSSR award supports training in evidence-based behavioral practice. *Outlook: A Quarterly Newsletter of the Society of Behavioral Medicine*. 2007; Winter:10.
37. Spring B, Walker B, Brownson R, et al. *Definition and Competencies for Evidence-Based Behavioral Practice*. Chicago, IL: Northwestern University; 2008. Council on Evidence-Based Behavioral Practice White Paper.
38. McKean E, ed. *The New Oxford American Dictionary*. 2nd ed. New York: Oxford University Press; 2005.
39. McQueen DV, Anderson LM. What counts as evidence? issues and debates. In: Rootman I, ed. *Evaluation in Health Promotion: Principles and Perspectives*. Copenhagen, Denmark: World Health Organization; 2001:63–81.
40. Rimer BK, Glanz DK, Rasband G. Searching for evidence about health education and health behavior interventions. *Health Educ Behav*. 2001;28:231–248.
41. Kerner JF. Integrating research, practice, and policy: what we see depends on where we stand. *J Public Health Manag Pract*. 2008;14:193–198.
42. Glasgow RE, Green LW, Klesges LM, et al. External validity: we need to do more. *Ann Behav Med*. 2006;31:105–108.
43. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. *Eval Health Prof*. 2006;29:126–153.
44. Castro FG, Barrera M, Jr., Martinez CR Jr. The cultural adaptation of prevention interventions: resolving tensions between fidelity and fit. *Prev Sci*. 2004;5:41–45.
45. Kerner JF, Guirguis-Blake J, Hennessy KD, et al. Translating research into improved outcomes in comprehensive cancer control. *Cancer Causes Control*. 2005;16 (Suppl 1):27–40.
46. Rychetnik L, Frommer M, Hawe P, Shiell A. Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health*. 2002;56:119–127.

47. Glasgow RE. What types of evidence are most needed to advance behavioral medicine? *Ann Behav Med*. 2008;35:19–25.
48. Kemm J. The limitations of 'evidence-based' public health. *J Eval Clin Pract*. 2006;12:319–324.
49. Maslov A. A theory of human motivation. *Psychol Rev*. 1943;50:370–396.
50. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. *Health Educ Q*. 1988;15:351–377.
51. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review: a new method of systematic review designed for complex policy interventions. *J Health Serv Res Policy*. 2005;10(Suppl 1):21–34.
52. Tones K. Beyond the randomized controlled trial: a case for 'judicial review.' *Health Educ Res*. 1997;12:i–iv.
53. Denzin NK. *The Research Act in Sociology*. London, UK: Butterworth; 1970.
54. Steckler A, McLeroy KR, Goodman RM, Bird ST, McCormick L. Toward integrating qualitative and quantitative methods: an introduction. *Health Educ Q*. 1992;19:1–8.
55. Dorfman LE, Derish PA, Cohen JB. Hey girlfriend: an evaluation of AIDS prevention among women in the sex industry. *Health Educ Q*. 1992;19:25–40.
56. Hugentobler M, Israel BA, Schurman SJ. An action research approach to workplace health: integrating methods. *Health Educ Q*. 1992;19:55–76.
57. Goodman RM, Wheeler FC, Lee PR. Evaluation of the Heart to Heart Project: lessons from a community-based chronic disease prevention project. *Am J Health Promot*. 1995;9:443–455.
58. Presidential Task Force on Evidence-Based Practice. Evidence-based practice in psychology. *Am Psychol*. 2006;61:271–285.
59. Gambrill E. Evidence-based practice: sea change or the emperor's new clothes? *J Social Work Educ*. 2003;39:3–23.
60. Melnyk BM, Fineout-Overholt E, Stone P, Ackerman M. Evidence-based practice: the past, the present, and recommendations for the millennium. *Pediatr Nurs*. 2000;26:77–80.
61. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA*. 1992;17:2420–2425.
62. Cochrane A. *Effectiveness and Efficiency: Random Reflections on Health Services*. London, UK: Nuffield Provincial Hospital Trust; 1972.
63. Canadian Task Force on the Periodic Health Examination. The periodic health examination: Canadian Task Force on the Periodic Health Examination. *Can Med Assoc J*. 1979;121:1193–1254.
64. US Preventive Services Task Force. *Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions*. Baltimore, MD: Williams & Wilkins; 1989.
65. Thacker SB, Berkelman RL. Public health surveillance in the United States. *Epidemiol Rev*. 1988;10:164–190.
66. Thacker SB, Stroup DF. Public health surveillance. In: Brownson RC, Petitti DB, eds. *Applied Epidemiology: Theory to Practice*. 2nd ed. New York: Oxford University Press; 2006:30–67.
67. Annest JL, Pirkle JL, Makuc D, et al. Chronological trend in blood lead levels between 1976 and 1980. *N Engl J Med*. 1983;308:1373–1377.
68. Tobacco Education and Research Oversight Committee for California. Confronting a relentless adversary: a plan for success: toward a tobacco-free California, 2006–2008. In: California Department of Public Health, ed. *Health*. Sacramento, CA: California Department of Public Health; 2006.

69. Biener L, Harris JE, Hamilton W. Impact of the Massachusetts tobacco control programme: population based trend analysis. *BMJ*. 2000;321:351–354.
70. Oxman AD, Guyatt GH. The science of reviewing research. *Ann N Y Acad Sci*. 1993;703: 33–34,125–133.
71. Mullen PD, Ramirez G. The promise and pitfalls of systematic reviews. *Annu Rev Public Health*. 2006;27:81–102.
72. Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago, IL: American Medical Association Press; 2002.
73. Kelsey JL, Petitti DB, King AC. Key methodologic concepts and issues. In: Brownson RC, Petitti DB, eds. *Applied Epidemiology: Theory to Practice*. New York: Oxford University Press; 1998:35–69.
74. Oxman AD, Cook DJ, Guyatt GH, Evidence-Based Medicine Working Group. Users' guides to the medical literature. Part 6: how to use an overview. *JAMA*. 1994;272:1367–1371.
75. Waters E, Doyle J. Evidence-based public health practice: improving the quality and quantity of the evidence. *J Public Health Med.* 2002;24:227–229.
76. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.
77. Carande-Kulis VG, Maciosek MV, Briss PA, et al. Task Force on Community Preventive Services: methods for systematic reviews of economic evaluations for the Guide to Community Preventive Services. *Am J Prev Med*. 2000;18:75–91.
78. Brownson RC, Royer C, Ewing R, McBride TD. Researchers and policymakers: travelers in parallel universes. *Am J Prev Med*. 2006;30:164–172.
79. Harris P, Harris-Roxas B, Harris E, Kemp L. *Health Impact Assessment: A Practical Guide*. Sydney, Australia: Centre for Health Equity Training/Research and Evaluation,. Part of the UNSW Research Centre for Primary Health Care and Equity; 2007.
80. Dannenberg AL, Bhatia R, Cole BL, Heaton SK, Feldman JD, Rutt CD. Use of health impact assessment in the US: 27 case studies, 1999–2007. *Am J Prev Med*. 2008;34:241–256.
81. Cole B, Shimkhada R, Morgenstern H, Kominski G, Fielding J, Wu S. Projected health impact of the Los Angeles City living wage ordinance. *J Epidemiol Commun Health*. 2005;59:645–650.
82. Wallerstein NB, Duran B. Using community-based participatory research to address health disparities. *Health Promot Pract*. 2006;7:312–323.
83. Green LW, George MA, Daniel M, et al. *Review and Recommendations for the Development of Participatory Research in Health Promotion in Canada*. Vancouver, BC, Canada: The Royal Society of Canada; 1995.
84. Israel BA, Schulz AJ, Parker EA, Becker AB. Review of community-based research: assessing partnership approaches to improve public health. *Annu Rev Public Health*. 1998;19:173–202.
85. Cargo M, Mercer SL. The value and challenges of participatory research: Strengthening its practice. *Annu Rev Public Health*. 2008;29:325–350.
86. Soriano FI. *Conducting Needs Assessments: A Multdisciplinary Approach*. Thousand Oaks, CA: Sage Publications; 1995.
87. Sederburg WA. Perspectives of the legislator: allocating resources. *MMWR*. 1992;41(Suppl):37–48.
88. Hallfors D, Cho H, Livert D, Kadushin C. Fighting back against substance abuse: are community coalitions winning? *Am J Prev Med*. 2002;23:237–245.
89. Bauman LJ, Stein RE, Ireys HT. Reinventing fidelity: the transfer of social technology among settings. *Am J Community Psychol*. 1991;19:619–639.

90. Glasgow RE, Emmons KM. How can we increase translation of research into practice? types of evidence needed. *Annu Rev Public Health*. 2007;28:413–433.
91. Green L, Ottoson J, Hiatt R, Garcia C. Diffusion, dissemination and implementation of evidence-based public health. *Annu Rev Public Health*. In press.
92. Bero LA, Jadad AR. How consumers and policy makers can use systematic reviews for decision making. In: Mulrow C, Cook D, eds. *Systematic Reviews Synthesis of Best Evidence for Health Care Decisions*. Philadelphia, PA: American College of Physicians; 1998:45–54.
93. The Lewin Group I. *Final Report: Factors Influencing Effective Dissemination of Prevention Research Findings by the Department of Health and Human Services*. Washington, DC: The Lewin Group, Inc; 2001.
94. Dobbins M, Cockerill R, Barnsley J, Ciliska D. Factors of the innovation, organization, environment, and individual that predict the influence five systematic reviews had on public health decisions. *Int J Technol Assess Health Care*. 2001;17:467–478.
95. Rogers EM. *Diffusion of Innovations.* 5th ed. New York: Free Press; 2003.
96. Simpson DD. A conceptual framework for transferring research to practice. *J Subst Abuse Treat*. 2002;22:171–182.
97. Cohen S, Eimicke W. Understanding and applying innovation strategies in the public sector: Paper presented at 57th Annual National Conference of the American Society for Public Administration; June 29–July 3, 1996; Atlanta, GA.
98. Scutchfield FD, Knight EA, Kelly AV, Bhandari MW, Vasilescu IP. Local public health agency capacity and its relationship to public health system performance. *J Public Health Manag Pract*. 2004;10:204–215.
99. Choi BC. Twelve essentials of science-based policy. *Prev Chronic Dis*. 2005; 2:A16.
100. Choi BC, Pang T, Lin V, et al. Can scientists and policy makers work together? *J Epidemiol Community Health*. 2005;59:632–637.
101. Koepsell TD, Weiss NS. *Epidemiologic Methods. Studying the Occurrence of Illness*. New York: Oxford University Press; 2003.
102. Last JM, ed. *A Dictionary of Epidemiology*. 4th ed. New York: Oxford University Press; 2001.
103. Shulock N. The paradox of policy analysis: if it is not used, why do we produce so much? *J Pol Anal Manage*. 1999;18:226–244.
104. Almeida C, Bascolo E. Use of research results in policy decision-making, formulation, and implementation: a review of the literature. *Cad Saude Publica*. 2006;22(Suppl):S7–S19, S20–S33.
105. Abney G. Lobbying by the insiders: parallels of state agencies and interest groups. *Public Administr Rev*. 1988;48:911–917.
106. Lavis JN, Robertson D, Woodside JM, McLeod CB, Abelson J. How can research organizations more effectively transfer research knowledge to decision makers? *Milbank Q*. 2003;81: 171–172, 221–248.
107. Drug Policy Foundation. *Public Policy: Youth Drug Education/The D.A.R.E. Program*. Washington, DC: Drug Policy Foundation; 2003.
108. West SL, O'Neal KK. Project D.A.R.E. outcome effectiveness revisited. *Am J Public Health*. 2004;94:1027–1029.
109. Wiesner PJ. Four diseases of disarray in public health. *Ann Epidemiol*. 1993;3:196–198.
110. Slonim AB, Callaghan C, Daily L, et al. Recommendations for integration of chronic disease programs: are your programs linked? *Prev Chronic Dis*. 2007; 4:A34.

111. Breslow L, Johnson M. California's Proposition 99 on tobacco, and its impact. *Annu Rev Public Health*. 1993;14:585–604.
112. Brownson RC, Ballew P, Dieffenderfer B, et al. Evidence-based interventions to promote physical activity: what contributes to dissemination by state health departments. *Am J Prev Med*. 2007;33: S4–S8, S66–S73.
113. Turnock BJ. *Public Health: What it is and How it Works*. 3rd ed. Gaithersburg, MD: Aspen Publishers, Inc.; 2004.
114. Office of Management and Budget. *Budget: Department of Health and Human Services*. Washington, DC: The Executive Office of the President; 2008.
115. Balas EA. From appropriate care to evidence-based medicine. *Pediatr Ann*. 1998;27:581–584.
116. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. *Am J Public Health*. 1999; 89:1322–1327.
117. Jilcott S, Ammerman A, Sommers J, Glasgow RE. Applying the RE-AIM framework to assess the public health impact of policy change. *Ann Behav Med*. 2007;34:105–114.
118. Baker EA, Brownson RC, Dreisinger M, McIntosh LD, Karamehic-Muratovic A. Examining the role of training in evidence-based public health: a qualitative study. *Health Promot Pract*. 2009;10:342–348.
119. Maxwell ML, Adily A, Ward JE. Promoting evidence-based practice in population health at the local level: a case study in workforce capacity development. *Aust Health Rev*. 2007;31:422–429.
120. Maylahn C, Bohn C, Hammer M, Waltz E. Strengthening epidemiologic competencies among local health professionals in New York: teaching evidence-based public health. *Public Health Rep*. 2008;123:35–43.
121. Brownson RC, Diem G, Grabauskas V, et al. Training practitioners in evidence-based chronic disease prevention for global health. *Promot Educ*. 2007;14:159–163.
122. Oliver KB, Dalrymple P, Lehmann HP, McClellan DA, Robinson KA, Twose C. Bringing evidence to practice: a team approach to teaching skills required for an informationist role in evidence-based clinical and public health practice. *J Med Libr Assoc*. 2008;96:50–57.
123. Pappaioanou M, Malison M, Wilkins K, et al. Strengthening capacity in developing countries for evidence-based public health: the data for decision-making project. *Soc Sci Med*. 2003;57:1925–1937.
124. Linkov F, LaPorte R, Lovalekar M, Dodani S. Web quality control for lectures: supercourse and Amazon.com. *Croat Med J*. 2005;46:875–878.
125. Brownson RC, Ballew P, Brown KL, et al. The effect of disseminating evidence-based interventions that promote physical activity to health departments. *Am J Public Health*. 2007;97:1900–1907.
126. Proctor EK. Leverage points for the implementation of evidence-based practice. *Brief Treat Crisis Interv*. 2004;4:227–242.
127. Chambers LW. The new public health: do local public health agencies need a booster (or organizational "fix") to combat the diseases of disarray? *Can J Public Health*. 1992;83:326–328.
128. Bryan RL, Kreuter MW, Brownson RC. Integrating adult learning principles into training for public health practice. *Health Promot Pract*. 2009;10(4):557–563.
129. Institute of Medicine. *Who Will Keep the Public Healthy? Educating Public Health Professionals for the 21st Century*. Washington, DC: National Academies Press; 2003.

130. Koo D, Birkhead GS, Reingold AL. Competency-based epidemiologic training in public health practice. *Public Health Rep*. 2008;123(Suppl 1):1–3.
131. Centers for Disease Control and Prevention. *CDC Taskforce on Public Health Workforce Development*. Atlanta, GA: Centers for Disease Control and Prevention; 1999.
132. Brownson R, Ballew P, Kittur N, et al. Developing competencies for training practitioners in evidence-based cancer control. *J Cancer Educ*. 2009; In press.
133. Tilson HH. Public health accreditation: progress on national accountability. *Annu Rev Public Health*. 2008;29:xv–xxii.
134. Ezzati M, Friedman AB, Kulkarni SC, Murray CJ. The reversal of fortunes: trends in county mortality and cross-county mortality disparities in the United States. *PLoS Med*. 2008;5:e66.
135. Shaya FT, Gu A, Saunders E. Addressing cardiovascular disparities through community interventions. *Ethn Dis*. 2006;16:138–144.
136. Subramanian SV, Belli P, Kawachi I. The macroeconomic determinants of health. *Annu Rev Public Health*. 2002;23:287–302.
137. Masi CM, Blackman DJ, Peek ME. Interventions to enhance breast cancer screening, diagnosis, and treatment among racial and ethnic minority women. *Med Care Res Rev*. 2007;64:195S–242S.
138. Peek ME, Cargill A, Huang ES. Diabetes health disparities: a systematic review of health care interventions. *Med Care Res Rev*. 2007;64:101S–156S.
139. Brownson RC, Haire-Joshu D, Luke DA. Shaping the context of health: a review of environmental and policy approaches in the prevention of chronic diseases. *Annu Rev Public Health*. 2006;27:341–370.
140. Mokdad AH, Marks JS, Stroup DF, Gerberding JL. Actual causes of death in the United States, 2000. *JAMA*. 2004;291:1238–1245.

CHAPTER 11

Public Health and Medical Education in the United States

Rika Maeshiro, MD, MPH

> In spite of the progress of public health work the medical schools have too generally neglected or slighted the preventive side of medicine. This has had an unfortunate result. The average physician fails to see as clearly as he should that he is a vital part of the public health organization, that he is expected to discover and to report communicable diseases, to instruct his patients, to support the local authorities, to help create sound public opinion.[1]
>
> —George E. Vincent (President, Rockefeller Foundation, 1917–1929)

HISTORICAL BACKGROUND

The challenge of incorporating public health content into the standard medical curriculum is not new. Not surprisingly, the roots of this struggle are entwined with the historical events and trends that led to the separation, or "schism" [2] as some have described, between the practice of medicine and the practice of public health in the United States. While in the mid to late 19th century some medical schools included public health in the curriculum,[3] during the first half of the 20th century trends in science, the professionalization of the medical and public health fields, and competition between private physicians and providers at public health clinics contributed to the separation of medicine and public health. Historians have described these events of the early 20th century from the "medical side" and the "public health side" of this division. Specifically, Elizabeth Fee, Roy Acheson, et al. considered the evolution of public health education as it became distinct from medical education,[4] and Kerr White reviewed how medical schools came to deemphasize epidemiology, the social sciences, and quantitative methods.[2] Both accounts reveal how some leaders of that era had concerns about distancing public health from medical training, and both considered two public health infrastructure issues that continue to be of concern. One concern relates to medical training—for example, during their training, physicians are not assured an adequate foundation in public health knowledge

and skills that would facilitate collaborations with public health colleagues and solutions to societal health needs. Another relates to the inadequate numbers of physicians who choose public health as their career choice. According to White, until the late 19th century, most leaders in academic medicine embraced the care of individual patients and concern for unacceptable environmental and social conditions that endangered the public's health. In the late 1800s, physicians who had studied in Germany and Austria influenced medical education in the United States. German and Austrian medical school curricula had begun to reflect the experiences of academic, hospital-based consultants (rather than those of community-based practitioners) and the research interests of professors (rather than the health issues of the community). Theodor Billroth, a prominent Austrian surgeon of the time, documented the evolution of medical education in German-speaking countries in his 1875 publication, *The Medical Sciences in the German Universities: A Study in the History of Civilizations*, and included his critique of public health in his work:

> The physician, as one of the most important members of the community, is expected not only to help in cases of individual sickness, but in community diseases as well. He is even expected to do his part in curing the stupidity and indifference of humanity. . . . The fanatical champions of public health are fighting for a goal that is too high for my myopic vision. I can admire the struggle, but I cannot become interested in it.[5]

The changing teaching priorities in Europe and the United States were driven in part by scientific discoveries, particularly in bacteriology, leading to what White describes as the "Big bug hunt": "The unquestioned message, that for each disease there was a single cause and that for most known diseases there was probably a single microbe, changed the focus of medical education, research, and practice."[2] Some described this focus as a "reductionist" [2, 3] perspective in medicine, in contrast to the broader concerns of population health and public health.

Breakthroughs in science influenced the recommendations of Abraham Flexner, author of the seminal 1910 report, *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching.*[6] Flexner's report promoted reforms to improve medical education. He recommended that medical education be rooted in up-to-date science and be based on clinical experiences. After visiting all of the medical schools, he recommended that many close because of inadequate facilities, resources, and teaching. Because his 1910 report is associated with what evolved into the traditional medical education model (two years of basic sciences, followed by clinical experiences), and because the changes in medical education prompted by his report contributed to an inadequate number of physicians opting for public health careers, it can be perceived, perhaps mistakenly, as promoting a non-public-health orientation to medical education. In fact, the report included public health themes (for example, the training, quality, and quantity of physicians should meet the health needs of the public; physicians have societal obligations to prevent disease and promote health, and medical training should include the breadth of knowledge necessary to meet these obligations; and collaborations between academic medicine and public health communities should result in benefits to both parties).[7] As such, the report did not totally divorce public health missions from the responsibility of physicians. Flexner saw

advancements in science as a basis for moving medicine toward disease prevention and a population perspective:

> For scientific progress has greatly modified his ethical responsibility. His relation was formerly to his patient—at most to his patient's family; and it was almost altogether remedial. . . . But the physician's function is fast becoming social and preventive, rather than individual and curative. . . . To the intelligent and conscientious physician, a typhoid patient is not only a case, but a warning: his office it is equally to heal the sick and to protect the well.[6]

At the turn of the 20th century, the number of U.S. medical schools was declining primarily because of changes in state licensing requirements for physicians. Because longer training periods were mandated (during which no income could be earned), fewer students enrolled in medical school, and schools, particularly marginal ones, were going out of business even before the release of Flexner's findings.[8] Those that remained tended not to adopt a public health or population health perspective. Medical education reforms stemming from the 1910 report also led to a decline in the number of practicing physicians and an increase in their incomes,[9] a scenario that did not encourage many doctors to choose public health (and its lower income) as the focus of their careers.[10] Sir James Mackenzie, a Scottish cardiologist, visited U.S. medical schools in 1918, and upon visiting Johns Hopkins, felt that while their research was advancing insights into disease mechanisms, their faculty was not attuned to the needs of their patient population.[2]

Concerned that insufficient numbers of "sanitary workers" were produced through the existing education programs [3] (including postgraduate courses at medical schools [11]), the Rockefeller Foundation began to consider the training of public health workers during the second decade of the 20th century. The foundation asked Abraham Flexner, who had continued his work on medical education reform after his 1910 report, to look for a site to train health officers.[2] His recommendations led to the 1916 decision to establish schools of public health as distinct entities separate from schools of medicine. Johns Hopkins was selected as the endowed institution. Nevertheless, the discussions that led to these decisions reveal a lack of consensus regarding the separation of medical and public health training. Dr. William Welch, the dean of the Johns Hopkins School of Medicine (and future founding dean of the Johns Hopkins School of Public Health and Hygiene) and co-author of the Welch-Rose report that helped to establish schools of public health, originally had envisioned the teaching of hygiene and public health to occur within a department of the medical school, but he was not supported by his medical school faculty.[2, 8] Wickliffe Rose, the organizer of the Rockefeller Foundation's campaign against hookworm and the other co-author or the Welch-Rose report, had always preferred a national system of public health training that was separated from medical training. The Welch-Rose report was the seminal justification and blueprint for schools of public health in the United States. Its recommendations were the essential starting point for planning the future of public health and the training of its practitioners. The report called for collaboration with medical schools and hospitals to share resources, development of a method to encourage public health education

for medical students and physicians ("It is of the utmost importance that education in the principles of hygiene should be available for students and graduates in medicine who are to engage in the practice of their profession."), and a grounding in basic biomedical and clinical sciences for public health students.[12] Many physicians continued to be unsupportive and suspicious of public health at this time. The prevailing belief was that efforts to control infectious diseases could pit health officers against private practitioners,[9] and the establishment of public health clinics could threaten the income of private physicians.[13]

Furthermore, the experts who developed the model for the first schools of public health had debated whether public health education should be limited to those who already had completed medical training. However, the appreciation for the multidisciplinary expertise that was required in public health and the acknowledgment that physicians were not flocking to public health practice prompted the new schools of public health to accept applicants from a variety of backgrounds (for example, engineers, statisticians, biologists), and sought to develop a discipline distinct from medicine.[10] Nonetheless, close ties were encouraged with schools of medicine primarily on the assumption that public health students would utilize medical school facilities to receive required education in "disease" and that conversely, schools of medicine would be the public health school would be "imbued with the spirit of public health."[9]

In the 1920s and 1930s, those involved in public health education continued to hope that schools of public health would promulgate "the spirit of preventive medicine" in medical schools, but the technological advances in medical diagnosis and treatment, and the new federal and state support for public health programs facilitated their separated existence.[9] Preventive medicine was taught in medical schools, but these departments typically were small and comparatively weak. In the 1930s and 1940s, preventive medicine become identified with "socialized medicine" ("social medicine" was the British term for "preventive medicine" [2]), a somewhat pejorative term and perceived as a potential economic threat to the profession. Medical schools were "more willing to express vague support for the concept of preventive medicine than to provide active advocacy or strong financial commitment to the idea."[9] More medical educators began to share and publish their concerns about improving public health education for physicians.[14] The 49th Annual Meeting of the Association of American Medical Colleges (AAMC) in October 1938, for example, featured at least four speakers who addressed the need to improve preventive medicine education in medical school.[15–18] In 1939, the AAMC appointed a Committee on the Teaching of Preventive Medicine and Public Health to design an education program in preventive medicine and public health for undergraduate medical students. During the 1944 Annual Meeting, the Committee presented its final recommendations, which addressed the organizational structure of medical schools, the curriculum, and the AAMC. Each school was to establish a Department of Preventive Medicine and Public Health, led by a full-time director, and supported by 5 to 8 percent of the medical school budget. Four percent of the total curricular time (initiated no later than the second year of medical school, and continued in each year thereafter) was to be devoted to preventive medicine and public health, and integrated into clinical teaching. The report also suggested that the preventive medicine curricula include content that would be identified in

21st-century terminology as applied biostatistics, the epidemiology and natural history of disease, the social and environmental determinants of health, and disease prevention.[18] AAMC was to support the creation of fellowships in preventive medicine and public health. None of the recommendations appear to have been applied universally.[2]

In 1949, Thomas McKeown, a British professor of social medicine, addressed the First World Conference on Medical Education and commented that medical education in Great Britain and the United States was "in danger of losing sight of the continuing importance of prevention of disease" and that the instruction provided in preventive medicine "was divorced from clinical training" and was thought to make "little impression on the student,"[2] observations that would resonate into the 21st century.

For medical educators in preventive medicine and public health, the 1940s through the 1980s were marked by expansions in scope, and related opportunities and concerns. Many departments of preventive medicine (or hygiene or public health) that had not been converted to a school of public health, merged with bacteriology departments. These departments frequently became the home for new topics that were not directly the responsibility of clinicians (for example, nutrition, environmental exposures, injury, disease screening), and they could be given responsibilities for tuberculosis, sexually transmitted infections, substance abuse, and mental illness. Later, related subjects such as health services research, planning, and evaluation and medical sociology were added to the scope, resulting in what has been described as departments of "miscellaneous medicine."[2] The Conference of Professors of Preventive Medicine (renamed the Association of Teachers of Preventive Medicine in 1953, and renamed again in the early 21st century as the Association for Prevention Teaching and Research) was established in 1942, stemming from the American Public Health Association's (APHA) subcommittee on education and medical students.[19] Members of the original group were "health officers, directors of student health services, statisticians, pathologists, bacteriologists, practitioners, and deans" who were interested in teaching public health to medical students. They agreed to meet regularly at APHA meetings and shared short presentations on public health topics. This group's meetings and activities over the ensuing four decades reflected the broad concerns of preventive medicine departments, shifting attention from quantitative skills, to clinical prevention, health services research, and health policy. Despite these efforts to enhance the public health education of physicians, and increased medical interest in the quantitative aspects of public health to meet biomedical research needs during the 1970s and 1980s,[20] U.S. educators in public health and preventive medicine approached the end of the 20th century with many of the same concerns that had been raised in the late 1930s.[18]

CONTEMPORARY PUBLIC HEALTH CHALLENGES AND IMPLICATIONS FOR MEDICAL EDUCATION

The end of the 20th century and the dawn of the 21st century found the United States confronting challenges to health and security that continued to fuel calls to improve physicians' knowledge of public health. Chronic disease prevention, health

disparities, disaster preparedness and response, and health systems reform are a sample of the issues that require improved collaborations between medicine and public health.[21, 22, 23] In 1994, the American Medical Association (AMA) and the APHA embarked on the Medicine and Public Health Initiative to catalyze more effective partnerships between medicine and public health. The initiative's agenda was far-reaching, trying to address medicine and public health collaborations in communities, education, research, and health care. A 1997 publication highlighted effective examples of medicine–public health collaborations, including those in education,[24] and a second publication in 1998 [25] provided a fuller listing of case studies in medicine–public health partnerships that had been collected through the initiative. Cases listed under "Synergy 6b" focus on education and training that promote the linkages between medicine and public health.

Concerned experts in medicine and public health gathered and published a number of reports that addressed the need to better educate physicians about public health, identifying the barriers to effective education in this area and the scope of public health that should be incorporated in the medical curricula, including "The Medical School's Mission and the Population's Health" (1992),[26] "Education for More Synergistic Practice of Medicine and Public Health" (1999),[27] and "Contemporary Issues in Medicine—Medical informatics and population health: Report II of the Medical School Objectives Project" (1999).[28]

Participants at the December 1990 conference sponsored by The Royal Society of Medicine Foundation, Inc. and the Josiah Macy Jr. Foundation that resulted in "The Medical School's Mission and the Population's Health" believed that redefining the mission of medical schools would be necessary to fully embrace the public health perspective into medical education.[29] Medical schools and their faculty were directed to be "thoroughly familiar" with the distribution of health issues in populations, particularly in their institution's service area,[30] and all medical students, not just those with interest in public health, were to receive basic education in the population perspective.[31] The Josiah Macy Foundation sponsored another conference on this topic in June 1998.[27] Participants in this conference recommended flexible approaches to encouraging synergy between medical and public health schools. Specific suggestions included restructuring grand rounds to incorporate the behavioral, social, and environment determinants of the patients' problems (along with the usual assessment of the biological determinants, diagnosis, and treatment) and the development of continuing education programs on clinically relevant public health topics so that faculty would be better role models for their trainees. Epidemiology, biostatistics, health policy, and prevention were identified as the recommended public health content for medical students. The barriers to improved synergy included financial incentives that emphasized treatment and did not recognize prevention efforts, and distrust between the two disciplines.

The AAMC's "Medical School Objectives Project Report" in 1999 identified three major barriers to the improved teaching of population health: the lack of a consistent medical school department "owning" the scope of the content; lack of funding to support population health curriculum development and teaching; and the view within the academic medical community that the interest in population health was primarily a response to managed care. The experts who contributed to this report believed that public concern over the health care system and public expectations that

physicians are trained to meet the health needs of the public, combined with students' expectations to be better educated in public health, and influences from the managed care movement, would help overcome these obstacles.[28]

Although not focused solely on physician training, the Institute of Medicine's (IOM) report on the public health workforce, *Who Will Keep the Public Healthy?*[32] recommended that "all medical students receive *basic* public health training in the population-based prevention approaches to health" and that "a significant proportion of medical school graduates should be *fully* trained in the ecological approach to public health at the master's of public health (MPH) level. A second report issued by the IOM in the 1990s, *Training Physicians for Public Health Careers*,[33] focused on preparing physicians who specialize in public health, but declared that all physicians "intersect with public health in many activities of their practice" and "are part of the public health system." This report divided physicians into three groups:

"All physicians"; i.e., those who can help detect and respond to epidemics, exposures, and other threats; who rely on public health guidance on topics such as international travel, immunizations, and other preventive services; and who can serve as leaders in emergency response, health promotion, tobacco control, nutrition, and other health promotion and health protection activities.
"Physicians in practices or specialties with public health needs"; i.e., infectious disease specialists investigating health-care-associated outbreaks, or pediatricians working in school health, or emergency physicians directing Emergency Medical Services.
"Public health physicians"; i.e., those who have chosen public health as their specialty and practice.

Because all medical students will become physicians practicing in at least one of these three categories, the authors reiterated the earlier IOM recommendations regarding medical education and added additional topics to the list recommended for all medical students. In particular, medical school graduates should be able to demonstrate population health competencies in the following areas:

- Assess the health status of populations using available data (for example, public health surveillance data, vital statistics, registries, surveys, electronic health records, and health plan claims data);
- Understand the role of socioeconomic, environmental, cultural, and other population-level determinants of health on the health status and health care of individuals and populations;
- Integrate emerging information on individuals' biologic and genetic risk with population-level factors when deciding on prevention and treatment options;
- Appraise the quality of the evidence of peer reviewed medical and public health literature and its implications at patient and population levels;
- Apply primary and secondary prevention strategies that improve the health of individuals and populations;
- Identify community assets and resources to improve the health of individuals and populations;
- Explain how community-engagement strategies may be used to improve the health of communities and to contribute to the reduction of health disparities;
- Participate in population health improvement strategies such as community-based interventions;

- Discuss the functions of public health systems;
- Understand the organization and financing of the U.S. health care system and effects on access, utilization, and quality of care for individuals and populations;
- Discuss the ethical implications of health care resource allocation and emerging technologies on population health; and
- Understand quality improvement methods to improve medical care and population health.

Recognizing that medical education begins with medical school, but continues through internships and residencies (graduate medical education, GME) and continuing medical education (CME), the contributors to this report addressed the continuum of medical education. The report recommended that national organizations representing medical education, public health education, and public health and preventive medicine practitioners collaborate to develop models for integrating training in public health principles and practice into physician education at both the undergraduate and graduate levels. GME programs were directed to identify and include public health concepts and skills relevant to the practice of that specialty. Medical specialty societies were urged to provide CME and self-assessment opportunities to address the elements of public health that are included in the practice of their specialty. The report also recommended that periodic recertification examinations include these elements. To help ensure the quality of these education efforts and identify potential role models, medical schools and GME programs were advised to include faculty who were trained and experienced in public health.

The Healthy People Curriculum Task Force's Clinical Prevention and Population Health Curriculum Framework [34] presented a listing of population health and clinical prevention topics that should be included in the education of all health professionals. The framework was developed by educators in medicine, nursing, pharmacy, dentistry, and physician assistant training in response to the Healthy People 2010 Objective 1-7: "Increase the proportion of schools of medicine, schools of nursing and health professional training schools whose basic curriculum for healthcare providers includes the core competencies in health promotion and disease prevention." The curriculum framework is based on the previous "[i]nventory of knowledge and skills relating to disease prevention and health promotion."[35] Although the Healthy People Objective calls for public health competencies, the interprofessional task force agreed that competency statements would be profession-specific and thus chose to identify broad topics, leaving the depth and breadth of teaching in each area to the professions.

CONTEMPORARY MEDICAL SCHOOL CURRICULA AND PUBLIC HEALTH

Incentives for curricular change in medical schools include accreditation standards, student demand, and the content of national certifying examinations. Recent trends show that these factors may be aligning to facilitate a greater emphasis on public and population health. The Liaison Committee on Medical Education (LCME) is the national accrediting authority for medical education programs at medical doctorate–granting schools in the United States and Canada, and is responsible for establishing their accreditation standards. Accreditation is a process of quality

assurance in postsecondary education that determines whether an institution or program meets established standards for function, structure, and performance. The accreditation process also fosters institutional and program improvement. As of October 2009, the accreditation standards include general language regarding the inclusion of "preventive medicine" in the medical school curricula. Standard ED-11 states that "[the curriculum] must include the contemporary content of those disciplines that have been traditionally titled anatomy, biochemistry, genetics, physiology, microbiology and immunology, pathology, pharmacology and therapeutics, and preventive medicine."[36] In late 2009, the LCME released proposed revisions to Standards ED-11 and ED-15 that would include "the public health sciences" in the annotations for the proposed new structure of the standards. The potential changes were discussed during the February 2010 meeting of the LCME.

The Commission on Osteopathic College Accreditation is responsible for establishing accreditation standards for osteopathic doctorate–granting medical schools. Accreditation standards for osteopathic schools identify both preventive medicine and public health in their guidance for medical school curricula.[37] Standard 6.1 states:

> The education should at least include, but not be limited to . . . principles, history and practice of osteopathic medicine, human anatomy, biochemistry, pharmacology, genetics, physiology, pathology, microbiology, physical and differential diagnosis, medical ethics and legal aspects of medicine; internal medicine, family medicine, pediatrics, geriatrics, obstetrics and gynecology, preventive medicine and public health, psychiatry, surgery, radiology.

In addition, public health is included in a discussion of Core Competencies, with Standard 6.5.1 stating a demonstration of "knowledge of professional, ethical, legal, practice management, and public health issues applicable to medical practice."[37]

Because medical schools do not share one standard data system by which curricular content can be tracked across all schools, assessing the inclusion of public health topics in current medical school curricula is imperfect. Some public health topics are included in the curriculum questions of the LCME Annual Medical School Survey, Part II.[38] Although the quality of the curriculum data is limited by its reliance on self-reporting, the survey enjoys a 100 percent response rate, and thus provides a general view of trends. Beginning in 1979, community preventive medicine, environmental health hazards, health care delivery, medical jurisprudence, nutrition, patient education, and population dynamic were among the topics included in a new question that inquired about the inclusion of subjects within the medical curriculum. The number of public health topics and the wording of the curriculum questions have changed over the years but reflect a continued interest in these issues by the LCME. In the 2008 survey, some 41 topics associated with public health were included. Over the years, great strides have been made in the inclusion of some of these topics (for example, prevention and health maintenance), but others, such as occupational medicine and health policy, remain relatively underrepresented.

Another barometer of the inclusion of public health in the medical school curriculum is reflected in the AAMC's Graduation Questionnaire, an annual survey of graduating fourth-year medical students.[39] A portion of this survey asks students about their perception of the time devoted to particular topics during medical

school. The question asks "Do you believe that your instruction in the following areas was inadequate, appropriate, or excessive?" Between 2006 and 2008, more than 40 percent of respondents felt that their instruction time in occupational medicine was inadequate, and more than 45 percent felt that their instruction time in health policy was inadequate, mirroring the results of the LCME survey.

The United States Medical Licensing Examination™ (USMLE™) program is the licensure examination for graduates of LCME-accredited, doctorate-granting medical schools. The examination includes three "steps," and successful completion is intended to provide state licensing authorities evidence that the graduate has the minimum knowledge and skills required for initial licensure. Advocates of improving public health content in medical school education have been concerned that the USMLE does not include sufficient or appropriate public health content. A comprehensive review of the USMLE began in 2004.[40] The proposed changes to the examinations include efforts to assess skill in accessing relevant information, evaluating its quality, and applying it appropriately in a clinical scenario,[41] a potential opportunity to improve the quantitative public health content (biostatistics, epidemiology) of the examinations. Medical school faculty experts in public health have been invited to participate in the redesign process, providing additional opportunities to increase and improve the USMLE's public health content.

Barriers to the Effective Integration of Public Health into the Medical Curricula

Many of the obstacles to the inclusion of public health content into the medical curricula have been persistent over the years:

- Confusion about what is included public health: Within the medical community and the general public, the scope of public health can be misconstrued as medical care for the underserved or limited to services provided by government public health units. Medical school faculty and administrators who have an inaccurate or incomplete understanding of the breadth of public health may have difficulty implementing and assessing the public health content in their curricula.
- Varying terminology: Beginning in the 19th century, "public health" referred to the health of the public in broad terms, without a necessary connection to medical practice. "Sanitary reform," "sanitary engineering," and "sanitary science" are terms that are not in wide use in the 21st century, but in the 19th century, they were associated with efforts to clean up industrial environments, water, food, and city streets. "Hygiene" became associated with the German emphasis on research at the same time that "public health" reflected the British emphasis on the administrative, or practice, perspective of public health.[9] "Public health medicine," "preventive medicine," "social medicine," and "community medicine" were terms that were created to acknowledge the medical contributions to public health.[10] "Preventive medicine" is a centuries-old term that was often synonymous with public health until the 20th century when it became more closely identified with clinical preventive medicine. "Community medicine" is a 20th-century term that in the United States implies a special focus on the community aspects of health and health care delivery. In the late 20th century, "population health" became a popular term, particularly in the academic community.[42]
- Student disinterest: A 1920 survey by E. O. Jordan [43] revealed attitudes of medical students that still resonate: Insufficient knowledge of the field, lack of patient contact,

politics, lower salary, and dislike of working within a bureaucracy were reasons medical students were not interested in working in public health. Some students felt that public health was a legitimate choice for physicians ("Public health is at present a branch science of medicine."), whereas others reflected a disregard for public health ("The medical profession does not itself take public health and preventive medicine seriously.").

- Lack of role models and faculty: Medical students may not interact with public health physicians while in medical school,[44] and public health content may be taught primarily by nonphysicians, who may not effectively present the relationships between public health and clinical care.[2] Although the number of full-time faculty with MPH and other public health–related degrees grew between 1990 and 1998, a significant number of schools had few faculty with formal public health training—57 of the 125 schools, at the time of the study, had fewer than 10 faculty members with MPH degrees.[45]
- Lack of curricular time: The 1944 Final Report on the Teaching of Preventive Medicine and Public Health found that "[f]acing an already overcrowded and not entirely elastic curriculum, proponents of preventive medicine and public health have been forced to insinuate these subjects into the teaching schedule in an opportunistic rather than a systematic manner."[18] This statement continues to be true and encourages creative approaches to incorporating public health content into existing learning opportunities, rather than the creation of new courses.

Innovations in Public Health Education at Medical Schools

To respond to the need for the effective and engaging integration of public health content into the medical curricula, educators at schools across the country have developed creative curricular innovations that emphasize the clinical relevance of public health topics, engage experiential learning techniques, and involve public health partners. The innovations include introducing public health concepts during the first sessions in medical school,[46] population health case studies that incorporate local health data,[47] population health projects,[48] hospital policy projects,[49] required public health or community health clerkships,[50, 51] partnerships with academic health departments,[52, 53, 54] mock health policy hearings,[55] pandemic exercises,[56] population health "ward rounds" and grand rounds, and "community windshield tours" that have students tour the regions from which their patients come to better understand social and environmental determinants of health and to appreciate the community-based services that are available.[7]

National programs also have addressed this need through competitive grants programs to schools. The Health of the Public: Academic Challenge program was launched in 1986 to challenge academic health centers to broaden their missions to address health needs of their surrounding communities.[57] Thirty-four academic health centers were involved over 11 years with funding provide by the Pew, Rockefeller, and Robert Wood Johnson Foundations. Seven program objectives guided the projects, addressing health professions education, research, and the academic health center's interactions with their community:

- Provide basic competencies in population-based subjects, including epidemiology and preventive medicine, to all health professions students
- Provide enhanced population-based education for selected students
- Include clinical prevention knowledge and skill-building activities at all levels of health professions education

- Conduct substantive scholarly studies in subjects related to population-based medicine
- Assume institutional responsibility for maximizing the health of a defined population within available resources
- Involve the academic health center in decision making about the development and deployment of health resources
- Involve the academic health center in the social-political process as an advocate of the health of the public

Some courses and programs that were developed through this program still exist, including the MD-MPH program at Tufts and the public health course at the University of Kansas.

"Undergraduate Medical Education for the 21st Century: A Demonstration of Curriculum Innovations to Keep Pace with a Changing Health Care Environment" (UME-21) was sponsored by Health Resources Services Administration from 1997 to 2002 to support curriculum development at 18 schools to work with partners to improve medical student education in light of "new health systems," including the introduction of managed care.[58, 59] Curricular innovations were to be focused on nine content areas (that is, health systems finance, economics, organization, and delivery; evidence-based medicine; communication skills; ethics; informatics; leadership; quality measurement and improvement; systems-based care; and wellness and prevention). This initiative was unique because of the participation of more than 50 external partners, including managed care organizations, health plans, community health centers, and local health departments.

The cooperative agreement between the Centers for Disease Control and Prevention (CDC) and AAMC resulted in a national effort, the Regional Medicine-Public Health Education Centers,[60] to improve public and population health education for medical students, and later for residents. Sixteen medical schools have received funding through the cooperative agreement to enhance their public health curriculum for all of their medical students, and to achieve these improvements through partnerships with public health colleagues. At a minimum, these public health partners were to partner with their local or state health agencies. In addition to implementing the kind of curricular innovations described previously, the grantees worked with representatives from AAMC and CDC to develop "Population Health Competencies" for medical students.[7] Prior recommendations regarding relevant public health topics were transformed into competency (or learning objective) statements that facilitate curricular needs assessments and student evaluations.

An alternate approach to integrating public health content into medical education is MD-MPH programs, in which medical students can pursue their MPH degree while in medical school.[61–65] In response to student interest, the number of medical schools that offer MD-MPH opportunities has increased through the end of the end of the 20th century and into the early 21st century.[66] The MPH is offered through either a school of public health or a graduate program in public health. Although a few programs manage to schedule the dual training into four years, most programs require five years for completion.

GRADUATE MEDICAL EDUCATION AND PUBLIC HEALTH

Whereas medical school is an opportunity to introduce all medical students to the foundations of public health, residency curricula provide an important opportunity to demonstrate how public health can be integrated into specialty-specific practice, as noted in the 2007 IOM report.[33] GME, internship, and residency training that lasts three to eight years, is required after medical school to complete specialty training. Because medical school graduates no longer enter practice directly from medical school, GME has been identified as the phase of medical education that will affect how physicians will practice medicine.[67] The Accreditation Council for Graduate Medical Education (ACGME) accredits training programs that follow the MD degree. The ACGME has residency review committees (RRCs) for each of 26 specialties. RRCs are responsible for identifying program requirements for each specialty, and evaluating programs against those requirements.

In 2007, the ACGME identified six common program requirements that are to be applied across all specialty programs: patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice. The common program requirements and the specialty-specific program requirements affect five specialties (emergency medicine, family medicine, obstetrics, pediatrics, and psychiatry) that reflect public health content. Across these specialties, the program-specific public health requirements fell into the patient care, medical knowledge, and systems-based practice categories.[68] Family medicine and pediatric requirements reflect a comprehensive approach to identifying and integrating clinically relevant public health content. To address the need to improve public health education in medical education across all specialties, the Regional Medicine-Public Health Education Centers Initiative, supported by AAMC and CDC, was expanded to include GME in 2007. A competitive process resulted in 13 funded sites that represented a full spectrum of specialties, including emergency, family, and internal medicine, obstetrics, pediatrics, psychiatry, and surgery.[69]

THE SPECIALTY OF PREVENTIVE MEDICINE

Preventive medicine is the medical specialty with expertise in public health. Efforts to establish medical specialty board certification in preventive medicine and public health began in the mid-1940s. Representatives from the APHA, the AMA, the Canadian Public Health Association, the Southern Medical Association, and the Association of Schools of Public Health helped plan the establishment of the American Board of Preventive Medicine and Public Health, which was incorporated in 1948. The Health Officers section of the APHA felt that specialty board certification would better prepare health officers for their responsibilities.[70]

In 2007, 79 ACGME-accredited preventive medicine programs were available in the United States. Preventive medicine residencies are based in medical schools, schools of public health, hospitals, and health departments. The U.S. Army, Navy, Air Force, and CDC also offer preventive medicine residencies.[33] Current requirements

for residency training include a minimum of one year of clinical training in an ACGME-accredited residency, followed by an academic year leading to an MPH or its equivalent, and a practicum requirement year. Most preventive medicine residencies do not offer the first year of clinical training, and approximately one-third of entering preventive medicine residents have completed a previous residency program. Three subspecialties include Preventive Medicine: Public Health/General Preventive Medicine (PH/GPM), Occupational Medicine, and Aerospace Medicine. PH/GPM is the specialty that is most closely associated with public health. The practicum year requirements vary by subspecialty, and PH/GPM requires a practicum experience at a public health agency for at least one month. The medical knowledge content areas for PH/GPM include: health services administration, public health practice, and managerial medicine; environmental health; biostatistics; epidemiology; and clinical preventive medicine.[71]

Preventive medicine residencies do not typically receive Medicare GME assistance, unlike most residency programs, because much of the training occurs in non-hospital settings. Consequently, many programs struggle to arrange for funding to support their residents, combining support from state and local health agencies, community health organizations, the Department of Veterans Affairs, and limited Title VII funds from the Health Resources and Services Administration.[33] As a result of these funding difficulties, some preventive medicine residency programs have been closed. The IOM report *Training Physicians for Public Health Careers* recommended strengthening the capacity of preventive medicine residency programs. The IOM suggested that federal graduate medical education funds that are not linked to clinical care support the expansion of existing PH/GPM residency programs and the creation of new programs to graduate a minimum of 400 graduates per year.

CONCLUSION

Unfortunately, the theme to improve physicians' understanding of public health is not new in medical education. Multiple generations of physicians, medical educators, and public health practitioners have voiced their concerns about how the promises of medicine, no matter how technologically sophisticated, would not be achieved without better integrated clinical care and public health systems. Foundations, medical specialty societies, and government agencies have supported a series of studies and initiatives to foster an improved appreciation for public health in all physicians (a theme that was ironically included in two historic reports—one that provided the framework for contemporary medical education and one that established public health schools separate from schools of medicine). The escalating investment in medical care, and the occasionally dubious health status indicators that are achieved fuel these concerns, along with predictions of decreased life expectancy for future generations, continued health disparities, and the threat of natural and man-made public health disasters.

The early 21st century does give cause for some optimism. The framework in which medical education exists (for example, accreditation criteria for both medical school and residency training, national examination content) appears to be acknowledging the need for physicians to have a population perspective. As reflected in the

demand for MD-MPH opportunities and in student surveys, medical students are seeking public health education. The biomedical research enterprise also appears to be appreciating the need for "implementational" sciences and "translational" research, as illustrated by the National Institutes of Health's Clinical and Translational Science Awards. Much like the 1970s and 1980s, research needs may help the medical community embrace the public health sciences. In addition to the calls from special interests within the medical and public health communities, communities that are concerned about their health and health care may accelerate the decades-old efforts to heal this schism. In conclusion, "Whoever wishes to investigate medicine properly [must consider] . . . the effects of the winds, . . . waters, . . . city, . . . ground, . . . [and] the mode in which [people] live.[72]

REFERENCES

1. Vincent GE. Rockefeller Foundation President's Review: Rockefeller Foundation annual report for 1924: 27–28. http://www.archive.org/details/reportro1924rockuoft. Accessed January 30, 2010.
2. White KL. *Healing the Schism: Epidemiology, Medicine, and the Public's Health.* New York: Springer-Verlag; 1991.
3. Viseltear AJ. The emergence of pioneering public health education programs in the United States. In: Fee E, Acheson RM, eds. *A History of Education in Public Health: Health that Mocks the Doctors' Rules*. Oxford, UK: Oxford University Press; 1991:145.
4. Fee E, Acheson RM, eds. *A History of Education in Public Health: Health that Mocks the Doctors' Rules*. Oxford, UK: Oxford University Press; 1991.
5. Billroth T. *The Medical Sciences in the German Universities: A Study in the History of Civilization*. Welch WH, trans. New York: Macmillan; 1924.
6. Flexner A. *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching*. Boston, MA: Updyke; 1910. Carnegie Foundation Bulletin No.4.
7. Maeshiro R, Johnson I, Koo D, et al. Medical education for a healthier population: reflections on the Flexner Report from a public health perspective. *Acad Med*. 2010;85(2):211–219.
8. Starr P. *The Social Transformation of American Medicine: The Rise of a Sovereign Profession and the Making of a Vast Industry*. New York: Basic Books; 1982.
9. Fee E. Designing schools of public health for the United States. In: Fee E, Acheson RM, eds. *A History of Education in Public Health: Health that Mocks the Doctors' Rules*. Oxford, UK: Oxford University Press; 1991.
10. Fee E, Porter D. Public health, preventive medicine, and professionalization: Britain and the United States in the nineteenth century. In: Fee E, Acheson RM, eds. *A History of Education in Public Health: Health that Mocks the Doctors' Rules*. Oxford, UK: Oxford University Press; 1991.
11. Ruis AR, Golden RN. The schism between medical and public health education: a historical perspective. *Acad Med*. 2008;83(12):1153–1157.
12. Welch WH, Rose W. Institute of hygiene: being a report by Dr. William H. Welch and Wickliffe Rose to the General Education Board Rockefeller Foundation, 1915. In: Delta Omega Honorary Public Health Society, ed. *The Welch-Rose Report: A Public Health Classic*. http://www.deltaomega.org/WelchRose.pdf. Updated 1991. Accessed February 6, 2010.

13. Brandt A, Gadrner M. Antagonism and accommodation: interpreting the relationship between public health and medicine in the United States during the 20th century. *Am J Public Health*. 2000;90(5):707–715.
14. FitzGerald JG. The place of preventive medicine in the medical curriculum. *J Assoc Am Med Coll*. 1939;14:16–20.
15. Leathers WS. The integration of the teaching of preventive and clinical medicine. *J Assoc Am Med Coll*. 1939;14:21–25.
16. Russell FF. The place of preventive medicine in the medical curriculum. *J Assoc Am Med Coll*. 1939;14:26–30.
17. Mustard HS. Utilizing of health districts in the teaching of preventive medicine to medical students. *J Assoc Am Med Coll*. 1939;14:31–36.
18. Final report of the Committee on the Teaching of Preventive Medicine and Public Health. *J Assoc Am Med Coll*. 1945;20:152.
19. Barker WH, Jonas S. The teaching of preventive medicine in American medical schools, 1940–1980. *Prev Med*. 1981;10:674–688.
20. Brandt A. Collaboration and competition: tracing the historical relationship of medicine and public health in the 20th century. In: Hager M., ed. *Education for More Synergistic Practice of Medicine and Public Health*. New York: Josiah Macy, Jr. Foundation; 1999.
21. Beitsch LM, Brooks RG, Glasser JH, et al. The Medicine and Public Health Initiative: ten years later. *Am J Prev Med*. 2005;29:149–153.
22. Davis RM. Marriage counseling for medicine and public health: strengthening the bond between these two health sectors. *Am J Prev Med*. 2005;29(2):154–157.
23. McGinnis JM. Can public health and medicine partner in the public interest? *Health Aff*. 2006;25(4):1044–1052.
24. Lasker R. *Medicine and Public Health: The Power of Collaboration*. New York: New York Academy of Medicine; 1997.
25. Lasker RD, Abramson DM, Freedman GR. *Pocket Guide to Cases of Medicine and Public Health Collaboration*. New York: New York Academy of Medicine, 1998.
26. White KL, Connelly JE, eds. *The Medical School's Mission and the Population's Health*. New York: Springer-Verlag; 1992.
27. Hager M, ed. *Education for More Synergistic Practice of Medicine and Public Health*. New York: Josiah Macy, Jr., Foundation; 1999.
28. Contemporary issues in medicine: medical informatics and population health: report II of the Medical School Objectives Project. *Acad Med*. 1999;74:130–141.
29. White KL, Connelly JE. Redefining the Mission of Medical School. In: White KL, Connelly JE, eds. *The Medical School's Mission and the Population's Health*. New York: Springer-Verlag; 1992.
30. Marmot MG, Zwi AB. Measuring the burden of illness in general populations. In: White KL, Connelly JE, eds. *The Medical School's Mission and the Population's Health*. New York: Springer-Verlag; 1992.
31. Black D. Discussion. In: White KL, Connelly JE, eds. *The Medical School's Mission and the Population's Health*. New York: Springer-Verlag; 1992.
32. Gebbie K, Rosenstock L, Hernandez LM, eds. *Who Will Keep the Public Healthy? Educating Public Health Professionals for the 21st Century*. Washington, DC: National Academies Press; 2003.
33. Hernandez LM, Munthali AW, eds. *Training Physicians for Public Health Careers*. Washington, DC: National Academies Press; 2007.
34. Allan J, Barwick TA, Cashman S, et al. Clinical prevention and population health: curriculum framework for health professions. *Am J Prev Med*. 2004;27:471–476.

35. Wallace RB, Wiese WH, Lawrence RS, et al. Inventory of knowledge and skills relating to disease prevention and health promotion. *Am J Prev Med*. 1990; 6:51–56.
36. Liaison Committee on Medical Education. Functions and structure of a medical school. http://www.lcme.org/functions2008jun.pdf. Published 2008. Accessed October 19, 2009.
37. Commission on Osteopathic College Accreditation. Accreditation of colleges of osteopathic medicine: COM Accreditation Standards and Procedures (effective July 1, 2009). http://www.osteopathic.org/pdf/SB03-Standards%20of%20Accreditation%20July%202009.pdf. Accessed October 26, 2009.
38. Liaison Committee on Medical Education. Part II: annual medical school questionnaire. http://www.lcme.org/qunaires.htm. Accessed October 23, 2009.
39. Association of American Medical Colleges. 2008 medical school graduation questionnaire: final all schools summary report. http://www.aamc.org/data/gq/allschoolsreports/2008_pe.pdf. Accessed September 13, 2009.
40. US Medical Licensing Examination. Comprehensive Review of USMLE:CRU. http://www.usmle.org/General_Information/CRU/index.html. Accessed February 1, 2010.
41. Committee to Evaluate the USMLE Program. Comprehensive Review of USMLE. http://www.usmle.org/General_Information/CEUP-Summary-Report-June2008.PDF. Accessed October 19, 2009.
42. Tricco AC, Runnels, V, Sampson M, et. al. Shifts in the use of population health, health promotion, and public health: a bibliometric analysis. *Can J Public Health*. 2008;99:466–471.
43. US Public Health Service. *The Future of Public Health in the United States and the Education of Sanitarians*. Washington, DC: Government Printing Office; 1922. Public Health Bulletin. No.124:19–21.
44. Coker RE Jr, Kosa J, Back KW. Medical students' attitudes toward public health. *Milbank Mem Fund Q*. 1996;44(2):155–180.
45. Dismuke SE, Sherman L. Identifying population health faculty in US medical schools. *Am J Prev Med*. 2001;20(2):113–117.
46. Ornt DB, Aron DC, King NB, et al. Population medicine in a curricular revision at case western reserve. *Acad Med*. 2008;83(4):327–331.
47. Kekering KW, Novick LN. An enhancement strategy for integration of population health into medical school curriculum: employing the framework developed by the Healthy People Curriculum Task Force. *Acad Med*. 2008;83(4):345–351.
48. Chamberlain LJ, Wang E, Ho ET, et al. Integrating collaborative population health projects into a medical student curriculum at Stanford. *Acad Med*. 2008;83(4): 338–344.
49. Jacobsohn V, DeArman M, Moran P. Changing hospital policy from the wards: an introduction to health policy education. *Acad Med*. 2008;83(4):352–356.
50. Finkel ML, Fein O. Teaching about the changing US health care system: an innovative clerkship. *Acad Med*. 2004;79(2):179–182.
51. McIntosh S, Block RC, Kapsak G, Pearson TA. Training medical students in community health: a novel required fourth-year clerkship at the University of Rochester. *Acad Med*. 2008;83(4):357–364.
52. Keck CW. Health department-medical school collaboration. In: Hager M., ed. *Education for More Synergistic Practice of Medicine and Public Health*. New York: Josiah Macy, Jr., Foundation; 1999.
53. Novick LF, Greene C, Vogt RL. Teaching medical students epidemiology: utilizing a state health department. *Public Health Rep*. 1985;100(4):401–405.

54. Melville SK, Coghlin J, Chen DW, Sampson N. Population-based medical education: linkages between schools of medicine and public health agencies. *Acad Med*. 1996;71(12):1350–1352.
55. Blumenthal DS. Improving community health through 'hands on' approach. *Natl AHEC Bull*. 2003;20(1):22–23.
56. Finkelstein JA, McMahon GT, Peters A, et. al. Teaching population health as a basic science at Harvard Medical School. *Acad Med*. 2008;83(4):332–337.
57. Edelman N, Guttman N. Evaluation of health of the public: an academic challenge. http://www.rwjf.org/pr/product.jsp?id=17974. Accessed October 19, 2009.
58. Pascoe JM, Babbott D, Pye KL, et al. The UME-21 Project: connecting medical education and medical practice. *Fam Med*. 2004;36(Suppl):S12–S14.
59. Mahoney JE, Cox M, Gwyther RE, et al. Evidence-based and population-based medicine: national implementation under the UME-21 Project. *Fam Med*. 2004; 36(Suppl):S31–S35.
60. Maeshiro R. Public health practice and academic medicine promising partnerships: regional medicine public health education centers—two cycles. *J Public Health Manage Pract*. 2006;12(5):493–495.
61. Chauvin SW, Roedenhauser P, Bowdish BE, Shenoi S. Double duty: students' perception of Tulane's MD-MPH Dual Degree Program. *Teach Learn Med*. 2000; 12(4):221–230.
62. Boyer MH. A decade's experience at Tufts with a four-year combined curriculum in medicine and public health. *Acad Med*. 1997;72(4):269–275.
63. Stellman JM, Cohen S, Rosenfield A. Evaluation of a one-year masters of public health program for medical students between their third and fourth years. *Acad Med*. 2008;83(4):365–370.
64. Harris R, Kinsinger LS, Tolleson-Rinehart S, Viera AJ, Dent G. The MD-MPH Program at the University of North Carolina at Chapel Hill. *Acad Med*. 2008; 83(4):371–377.
65. Smith SR. Perceptions of the MD-MPH Option at Brown. *Acad Med*. 1996;71(10): 1024–1025.
66. Maeshiro R. MD-MPH opportunities at US medical schools: Poster presented at Association of American Medical Colleges Annual Meeting; April 30, 2004; Boston, MA.
67. Whitcomb ME. Flexner redux 2010: *graduate* medical education in the United States. *Acad Med*. 2009;84:1476–1478.
68. Accreditation Council for Graduate Medical Education. Common program requirements: general competencies. http://www.acgme.org/outcome/comp/GeneralCompetencies Standards21307.pdf. Accessed January 30, 2010.
69. Association of American Medical Colleges. Regional medicine-public health education centers: GME. http://www.aamc.org/members/cdc/rmphec/rmphecgme_detail.htm. Accessed February 7, 2010.
70. Ring AR. History of the American Board of Preventive Medicine. *Am J Prev Med*. 2002;22(4):296–319.
71. Accreditation Council for Graduate Medical Education. ACGME program requirements for graduate medical education in preventive medicine, 2007. http://www.acgme.org/acWebsite/downloads/RRC_progReq/380pr07012007.pdf. Accessed on February 1, 2010.
72. Hippocrates. *On Airs, Waters, and Places*. Adams F, trans. The Internet Classics Archive by Daniel C. Stevenson, Web Atomics. http://classics.mit.edu/Hippocrates/airwatpl.mb.txt. Accessed January 31, 2010.

SECTION 5

ASSURING THE HEALTH OF THE PUBLIC: PUBLIC HEALTH CHALLENGES IN THE 21ST CENTURY

This page intentionally left blank

CHAPTER 12

The Stem Cell Controversy: Navigating a Sea of Ethics, Politics, and Science

Ryan Cauley, MD

Sitting in the far corner of Dr. Shahin Raffi's lab at Weill-Cornell Medical College's Institute of Regenerative Medicine is a beating heart. Not a whole heart, but a piece of living heart tissue that has been produced from human embryonic stem cells. Dr. Raffi created the cells by introducing them to a series of growth factors typically present when the fetal heart develops in the womb. The heart tissue lies in a Petri dish and contracts between 60 and 70 beats per minute—normal for human cardiac tissue.

Roughly two blocks away, in New York Presbyterian Hospital's Cardiac Intensive Care Unit lies Mr. Smith, a recent recipient of a quadruple coronary bypass surgery. Severe heart disease runs in Mr. Smith's family, and despite trying to maintain a healthy diet, he has had three heart attacks in the past three years. Each heart attack has caused irreversible damage to his cardiac tissue, drastically increasing his risk for subsequent attacks and eventual death from heart disease. He now becomes so fatigued from simply walking across the room, that he tends to sit most of the time. Having exhausted all other reasonable options, a heart transplant is his only hope of living a normal life. At any given time, more than 2,500 patients are on the national heart transplant list. Of those, slightly over half will receive a new heart.[1] More than 450 of these patients will die waiting.

Few areas of biomedical science have aroused as much controversy as embryonic stem cell research. Since the derivation of the first human embryonic stem cells in 1998, the issue has been at the forefront of scientific, ethical, and political debates. Stem cells appear to offer unprecedented opportunities for developing new medical therapies for many debilitating diseases and a new way to explore fundamental questions of biology. In many ways, stem cell therapies may be the only foreseeable hope to many individuals who suffer from a variety of diseases for which there is currently no cure or effective treatment. Despite the great

---

This chapter was originally printed as Cauley, R. "The Stem Cell Controversy: Navigating a Sea of Ethics, Politics, and Science" in Finkel, M.L. *Truth, Lies, and Public Health*. Westport, CT: Praeger Publishers, 2007. Reprinted with permission.

aspirations of stem cell proponents, Dr. Raffi would be the first to admit that his heart cells are not yet ready for implantation in humans. After all, there are inherent risks in transplanting cardiac tissue derived from stem cells in someone's heart; but with more research, the future may be very bright.

With the global economy now faltering, the financial industry, politicians, physicians, and scientists envision tremendous economic benefits of a burgeoning stem cell industry. Individuals with incurable diseases envision the medical miracles that possibly could cure a multitude of diseases. Nevertheless, opponents of stem cell research still speak of the immorality of utilizing human cells, even for potential medical cures. While it would be ideal to have a rational and coherent national dialogue about this complex and controversial issue, radical viewpoints on both sides of have made it difficult to reach a compromise regarding a way forward for stem cell research.

This chapter focuses on the stem cell debate and addresses the issue from a medical, ethical, and political perspective. How close are we to curing diseases using stem cells? What are the ethical and moral issues involved in researching these cures? What political issues have arisen over the funding of stem research and how has this affected its progress? But firstly, what are stem cells anyway?

WHAT ARE STEM CELLS?

Most cells in humans are committed to becoming a single type of cell with a specific function within the body—that is, muscle cells, blood cells. In contrast, pluripotent stem cells are "uncommitted"—that is, able to become a number of different cell types, providing a number of different functions. Because pluripotent stem cells give rise to almost all of the cell types of the body, they hold great promise for both research and medical care. Pluripotent stem cells may serve as a source of generated cells and tissues for transplantation, potentially treating many diseases and conditions.

Each time these cells differentiate, they potentially can become fewer different types of cells. Stem cells that no longer can become every type of cell in the human body are called "multipotent" cells, loosely translating to "potentially many" (usually of a single cell class, like blood cells). For example, stem cells derived from adult bone marrow, where blood is made, only become blood cells, and not liver, heart or nerve cells.[2] Eventually each of the stem cells differentiates into "committed" cells. These committed cells sometimes can divide and produce copies of themselves, but they can never become any other type of cell.[3]

Although stem cells are present in the greatest quantities during human embryonic development, some stem cells still are maintained in the adult human body, albeit in very small quantities. It is thought that in most tissues these "adult stem cells" provide a source of new cells to replace those lost due to organ damage or natural cell death. These adult stem cells produce copies of themselves throughout the lifetime of the organism, providing a permanent source of cellular repair.[4] As such, adult stem cells may have limited potential compared to pluripotent stem cells derived from embryos or fetal tissue. While adult stem cells can be useful in certain therapies, it previously was thought that they may not be as potentially powerful as a truly "pluripotent" stem cell, which can become any other type of cell in the human body. Recently, however, scientists have discovered ways to "de-differentiate" or reprogram these adult cells, in an attempt to turn them back into pluripotent stem

cells. By creating pluripotent cells from adult tissues, it may be possible to circumvent some of the ethical concerns of using stem cells strictly from embryonic sources. Researchers currently are studying whether or not these induced pluripotent stem cells truly are equivalent to the embryonic stem cell.

HISTORY OF STEM CELL RESEARCH

In 1953, almost by accident, research on stem cells began. While investigating the effects of cigarette papers and tobacco on laboratory mice, a young scientist named Leroy Stevens noticed a tumor in one of his lab mice. Strangely, this tumor seemed to be completely unrelated to the effects of the smoking trials. Located in the testicles of one of his adult male mice, the tumor was found to be a teratoma, or a mass of wrongly differentiated cells, containing bone, teeth, and hair. Dr. Stevens found that by injecting stem cells derived from the inner cell mass of embryonic mouse blastocysts into the testes of other mice, he could induce the formation of a teratoma. In a series of experiments, he proved that stem cells could both be derived from the embryo, and forced to differentiate when placed in a live organism.[5] A year later, in 1954, Dr. John Enders of Harvard University began to use stem cells derived from a fetal kidney to produce poliovirus. For this work, Dr. Enders later would be awarded the Nobel Prize in Medicine.

It was not until the late 1960s that the first medical therapies based on the use of stem cells became available. In 1968, several children with severe immune deficiency disorder (known commonly as the "bubble boy disorder," where no functional white blood cells are made) were successfully given bone marrow transplants. After the transplants, the children were found to be making new white blood cells, proving that the transplanted marrow both contained blood stem cells, and that these cells were capable of surviving and dividing in a new organism.[6]

In the 1970s and 1980s, embryonic stem cells derived from blastocysts were demonstrated to spontaneously give rise to a number of different cell types while allowing them to divide and replicate in a Petri dish. One of the most exciting discoveries related to stem cell research occurred in 1996 when scientists at the Roslin Institute in Scotland announced the birth of Dolly the Sheep, the first animal cloned from adult cells. To clone Dolly from her "mother," the scientists had taken skin cells from an adult sheep, extracted the genetic information and placed it into a fertilized sheep egg (with its genetic information already removed). This fertilized egg, now with Dolly's mother's genes, was then implanted in the womb of a surrogate sheep to be allowed to come to term. Several months later Dolly was born, and history was made.[7]

Researchers in the United States also were working in this nascent field. In 1998, two separate research teams led by Drs. James Thompson of the University of Wisconsin and John Gearhart of Johns Hopkins University developed the first embryonic stem cell lines. In both cases, the research was funded privately (no federal funds were used). Stem cell lines are stem cells that have been placed in a Petri dish and induced to replicate, producing a permanent source of identical stem cells. Dr. Thompson and his colleagues derived their cell line from cells taken from surplus embryos donated voluntarily by couples undergoing fertility treatment at an in vitro fertilization (IVF) clinic.[8, 9] Dr. Gearhart's cell line, from early, nonliving

fetuses obtained from first trimester abortions, produced cells that could be replicated indefinitely and were shown to have the potential to grow into any tissue or organ in the body, thus holding great promise for treatment and cures. Before this time, animal embryos were the only source of embryonic stem cells.

In response to these ground-breaking studies, in 1999, the journal *Science*, in a special cover article and editorial, declared pluripotent stem cell research to be the scientific 'breakthrough' of the year.[10]

HOW ARE STEM CELL LINES MADE?

Stem cells can come from several different sources in the human body, specifically from adult organs and tissues, embryonic tissues, and most recently from umbilical cord blood, which possesses a high concentration of stem cells. Adult stem cells are taken from adult tissues, such as skin, the liver, and bone marrow rather than from embryos. As stem cells are present in greater quantities in adult bone marrow than in most other adult tissues, it is not surprising that marrow was one of the first places that adult stem cells were successfully harvested and used therapeutically.

Adult stem cells have been used therapeutically since the 1960s, when the first successful bone marrow transplants were performed. Yet stem cell lines are far more difficult to create and maintain when starting with adult stem cells. Adult stem cells can be made to divide and replicate in culture; however, scientists have found that they often cannot induce the cells to divide indefinitely. Adult cells lack a gene for "telomerase," an enzyme that allows for a cell to continue dividing. Without it, after a certain number of divisions, the cell lines simply will die.[11] Until quite recently, it had been thought that stem cells originating from adult organs could only become cells found in the organ from which the stem cell was taken. In other words, it was thought that stem cells in liver could only make liver cells, and stem cells found in the nervous system could only make nerve cells. Several experiments conducted over the past few years, however, have found that stem cells originating from one organ possibly can become cells of another organ type if encouraged in "the right way." For example, some researchers have shown that adult liver cells could be transformed relatively easily into insulin-producing pancreas cells.[12]

Embryonic tissues historically have been the most reliable source of stem cell lines. Pluripotent stem cells are found in great quantities in the human embryo. Embryonic stem cell lines can be derived from early embryos before they implant in the uterus. The greatest advantage of using embryonic stem cells is their "pluripotency," or ability to become any type of cell. A single stem cell line from an embryo therefore potentially could cure a larger range of diseases than a single adult stem cell line from an adult organ. With the recent advent of induced pluripotent stem cells, this may change. Currently, inducing adult tissues is so inefficient that it may not be ready for bedside therapy for quite some time. Although stem cell lines are difficult to create from induced adult stem cells, large numbers of embryonic stem cells can be grown relatively easily in culture. By placing the cells in Petri dishes with feeder cells (which help support the stem cells) and several chemical agents, embryonic stem cells will divide and flourish indefinitely. In fact, the first embryonic stem cell lines created during the late 1990s are healthy and continue to divide to this day.[4] With thousands of surplus embryos, the

byproducts of IVF therapy, embryonic stem cells theoretically are readily available for research purposes.

Umbilical cord blood is a new and potentially exciting source of adult stem cells. The blood, which is now often collected from the umbilical cord after birth, is typically rich in several different types of adult stem cells (although the majority are blood stem cells). As the collection procedure is painless and quick, it is possible that this could be a major source of stem cells in the future. In fact, in 2003, Congress passed the Cord Blood Stem Cell Act to establish a national network to prepare, store, and distribute the cells. Just after this act was passed, nongovernmental organizations such as the National Bone Marrow Donation Center and the Red Cross also began to set up national cord blood banking programs to encourage the donation of cord blood and to take advantage of this rich source of cells.[5]

Inducing adult cells to become pluripotent stem cells always had been the goal of stem cell researchers hoping to avoid the ethical arguments over the use of embryonic stem cells. In a landmark study Takahashi and Yamanaka [13] found that by introducing several key genes into adult mouse cells they could create "induced pluripotent stem cells" (or iPS cells). As these iPS cells were created from adult cells, they immediately were hailed as a means of eliminating the use of embryos in stem cell therapy and research. With this possibility in mind, iPS cells quickly became a favorite of the conservative movement and opponents of embryonic stem cell research.[14] While iPS cells potentially could reduce the reliance on embryonic stem cells, the most important aspect of iPS cells is actually their possible ability to create cells for use in personalized or customized medicine. Until this time, the only means of creating cells that were identical to the patients they were meant to treat was through therapeutic cloning. By using "personalized stem cells" to cure disease, physicians could avoid the use of potentially toxic antirejection drugs.[15] For stem cell researchers, the development of personalized stem cell lines always has been the holy grail of stem cell treatment.

STEM CELL RESEARCH AND CLONING

One goal of stem cell research is to provide cells that could be implanted in humans to repair damaged organs and tissues. The range of diseases that could be helped by this type of therapy is tremendous. Many considerations need to be taken into account, however, when placing foreign material in any human being. First and foremost, there is the consideration of the possibility of rejection. The human body has an excellent immune system that is built to recognize foreign material. When a foreign organ is transplanted into an individual, the individual's immune system will work to attack and destroy the organ. For this reason, organs must be "matched" to their recipient to minimize the chance of rejection. Using a series of complex tests, doctors can tell the likelihood of a certain individual rejecting a given organ. Of course, except in the case of identical twins, no donor is going to completely genetically match a recipient. Therefore, doctors have discovered that by using a cocktail of medications they can suppress or "turn off" the immune system so that the transplanted organ can survive. As these types of immune suppression therapies can wreak havoc on an individual, leaving them more at risk of infection and some cancers, they are used only when absolutely necessary.

In stem cell therapies, a foreign body, albeit a much smaller one, is being transplanted into an individual. If the stem cell line is not a complete genetic match for the recipient, there will be an immune response that will reject the foreign stem cells. It is therefore of utmost importance to either (1) have stem cells that will be a complete genetic match for the recipient, or (2) have a sufficient number of unique stem cell lines available that a near perfect match can be made. In this vein, scientists have begun to conceive of ways to produce stem cell lines that satisfy these criteria. One possibility for creating genetically identical stem cells is therapeutic cloning.

Cloning is a time-intensive and expensive process. Theoretically, only one human egg is required to create each new stem cell line. However, therapeutic cloning is actually quite a bit more difficult than this implies. Using current techniques, only 1 percent of eggs that have been injected with new genetic information go on to become stem cell lines. This means that for each stem cell line that is created, more than 100 eggs will be needed. Each egg will need to be donated by women willing to undergo the painful procedure of egg harvesting. Although doctors are now making great strides in increasing the yield of stem cell lines from cloned eggs in mouse studies, much work still is being done to continue to improve the process.

Through the use of therapeutic cloning, it is possible to produce embryonic stem cell lines that are perfect genetic matches for patients. Reaching this goal would mean being one step closer to realizing the tremendous therapeutic potential that embryonic stem cells appear to offer for the future. The stakes are high and the pressure to be the first to produce stem cell lines by cloning cells can lead some to take irresponsive action. The biggest scandal to date occurred in South Korea.

Scientist Dr. Hwang Woo Suk and his colleagues of Seoul National University published a paper in the acclaimed journal *Science* in 2004 claiming to have produced stem cell lines by cloning cells derived from adult patients.[16] The researchers alleged that these cell lines were genetically identical to the patients from whom they were cloned, and therefore perfect for future stem cell therapy. Scientists around the globe became ecstatic, as their goal of using genetically identical stem cells for "personalized medicine" seemed closer than ever to being realized. Dr. Hwang's apparent accomplishments were received by the medical establishment as a harbinger of future success in embryonic stem cell research. His experiments were deemed to be proof of the success of stem cell research and were used to justify increases in state and private spending. But, in December 2005, Dr. Hwang admitted to falsifying his experimental records and abruptly resigned from his university post.[17] A panel of investigators found that Dr. Hwang's laboratory had no record of ever having successfully created a genetically identical stem cell line through the use of cloning.

Despite Dr. Hwang's unfortunate falsification of his records, many other stem cell researchers have pressed on in the quest to create human stem cells through the use of therapeutic cloning. The process of therapeutic cloning to create stem cell lines is still its infancy; thus far, no study has reported the successful creation of a stem cell line from an embryo derived through somatic cell nuclear transfer (SCNT) or therapeutic cloning.[18] However, a number of scientists have been able to use therapeutic cloning to create early human embryos.[19] Although none of these human embryos have been implanted in a human uterus to create a child, they appear able to progress though early stages of development. If these embryos could

be developed further, they theoretically could become a source for embryonic stem cells that are identical to the donor of its genetic information.

INDUCED PLURIPOTENT STEM CELLS: THE FUTURE OF STEM CELL RESEARCH?

Since the discovery of pluripotent stem cells, the direction of stem cell research often has been dictated by political and ethical controversy. It is thought that with therapeutic cloning, stem cells could be created that are identical to the individuals they will be used to treat. When utilizing stem cells that are identical to the patient, it likely would not be necessary to give potentially dangerous drugs to suppress the immune system and avoid rejection. With many religious leaders opposed to the use of therapeutic cloning in the production of new stem cell lines, new ways of creating pluripotent stem cells for use in personalized medicine have been investigated.

In a landmark study in 2006, Takahashi and Yamanaka discovered a way of reprogramming adult cells to become iPS cells, effectively avoiding the arguments over the use of cloning or embryos to produce stem cells. To create these pluripotent stem cells, Takahashi and Yamanaka used four genes known to be critical in allowing pluripotent stem cells to divide indefinitely. By inserting these genes into the DNA of a type of adult mouse cell called a "fibroblast," these adult cells began to exhibit characteristics of pluripotent stem cells. These iPS cells were found to have the crucial ability to differentiate into other cell types, a feature unique to stem cells.[13] The cells expressed genes typically found only in pluripotent stem cells and even could contribute to development when they were injected into mouse embryos (as stem cells are typically able to do). For all intensive purposes, it appeared that these cells were pluripotent stem cells. Since these initial mouse studies, Takahashi, Yamanaka, and others have created induced pluripotent stem cells using adult cells from a human.[20]

In the ensuing years, a number of researchers began to find significant differences between these iPS cells and embryonic stem cells. One of the major sources of these differences is thought to be something known as "epigenetics." Epigenetics is a recent discovery in the field of cell biology. It always had been thought that all genetic information was determined by the sequence of genes in the DNA. However, researchers began to notice that some genes were turned "on" or "off" not only by other genes (such as gene sequences in the DNA known as "promoters" or "repressors"), but also by factors that occur after the DNA sequence was completed. It was found that DNA could be changed after it was made, allowing a cell to pass on genetic information that was not contained purely in the sequence of DNA (the precise order of bases that make up a gene). In fact, the term "epigenetics" literally means "outside of genetics." Cells that may appear to have the same gene sequence, and thus the same genetic information, may in fact have different sets of genes epigenetically turned on or off, effectively resulting in different cells.

For stem cells to replicate, they need to have certain specific genes turned "on." In the process of creating iPS cells, however, it is not known how the epigenetics of a cell are affected. Since the process of reprogramming described by Takahashi and Yamanaka involves only the introduction of new genes, it changes only the cells genetic sequence. Epigenetically, no changes are made. As genes can be

turned off due to epigenetics, not reprogramming the cells epigenetically makes little sense. Without reprogramming all aspects of the cells genetic material, we never will be able to create cells that are truly identical to embryonic stem cells.

The successful formation of stem cell lines from iPS cells is startlingly low: 0.0006–3 percent,[21] versus 50–69 percent from early embryos.[22] This low rate of return makes it costly and difficult to reliably create stem cell lines from iPS cells. Currently, embryos remain a much more reliable and easier source of pluripotent stem cells. It has been found that epigenetics may be one reason that the reprogramming process is so inefficient. Clearly, embryonic stem cells are in a different epigenetic state than iPS cells, which largely possess the epigenetic state of an adult cell.[23] For this reason, many of the induced stem cells likely have many embryonic genes epigenetically "turned off." For these cells to successfully transform into stem cells, it is thought that random events causing epigenetic changes may be required to "turn on" the necessary genes. As these random events occur at a low rate, it becomes a long and arduous process to successfully create iPS cells.

Genes typically found in stem cells often are found in tumor cells. Some scientists have been concerned that the injection of tumor genes into normal cells may increase the likelihood that they could become cancerous. Recent studies have concentrated on reducing this risk by decreasing the number of cancer-related genes used to produce iPS cells. Researchers are continuing to find ways to reduce the risks of iPS cells so that they may be used safely in future therapies. Despite these worries, iPS cells remain one of the most likely sources of stem cells for use in personalized medicine, and an extremely important discovery in the field of stem cell research.[24]

THE POTENTIAL OF STEM CELL RESEARCH

Millions of people suffer from a wide range of diseases, many of which are either difficult to treat or are incurable with current medical therapies. The potential use of stem cell therapy to affect a cure, or ameliorate symptoms, is fueling research. It is hoped that with more focused studies, stem cell treatments will be found for a great number of diseases. One reason for hope is that a number of diseases already are being treated successfully with stem cell–based therapies. One of the best examples of this cell-based treatment is leukemia. A sometimes-devastating blood cancer, Leukemia often is treated through the use of radiation or chemotherapy, followed by bone marrow transplants. Research has shown that stem cells from cord blood are a viable alternative to bone marrow as a source of new blood stem cells. In fact, it recently was suggested that cord blood stem cells might not have to be as closely genetically matched to a recipient to avoid detection and rejection by the patient's immune system.[25] As cancer patients often do not have relatives that would be suitable genetic matches for bone marrow donation, cord blood from unrelated donors potentially could be a life-saving alternative. Since this discovery, transplants of stem cells from cord blood have begun to replace bone marrow transplants in the treatment for leukemia, often with great success.[26]

Recent animal trials have shown partially restored eyesight in mammals with macular degeneration, the most common cause of blindness in human beings (essentially due to older age).[27] Research is focusing on isolating adult stem cells from

a blind individual and reimplanting these cells in a patient's retina. If successful, this technique could revolutionize the way physicians treat blindness. Much work still needs to be done, of course, to ensure that this kind of therapy will be feasible.

Spinal cord trauma always has been one of the most difficult injuries for physicians to treat. Unlike the cells of our skin, which can divide and replace themselves when the skin is cut or damaged, nerve cells normally are not capable of regeneration. For many years it was thought that spinal cord cells never would be capable of repair, leaving little hope for people with spinal cord injuries.

In 2009, the Geron Corporation became the first group to receive clearance from the Federal Drug Administration (FDA) for a human trial of embryonic stem (ES) cell therapy. The trial is a phase I multiphase study to assess the safety of ES cells in treating patients with new spinal cord injuries. At first, the study will use patients with spinal cord injuries to the thoracic spine that are between 7 and 14 days old. In the future they also will attempt to treat older injuries or those outside of the thoracic spine.

The repair of heart tissue often has been seen as a holy grail for cardiac researchers. Nearly 700,000 Americans died of heart disease in 2002, now the leading cause of death in the United States.[28] At present, it is impossible to completely reverse the heart damage that inevitably occurs during a myocardial infarction, the major cause of cardiac-related deaths. If heart tissue could be replaced or repaired, it would be possible to greatly reduce the catastrophic nature of this illness. As is the case with spinal cord damage, stem cells offer one of the most promising therapies for the repair of cardiac tissue. The current therapy for people with pacemaker damage is the implantation of an electronic pacemaker. As with any heart surgery, the implantation of an artificial pacemaker poses a considerable amount of risk to patients, and it can lead to many more cardiac complications than a natural pacemaker. If the heart's pacemaker could be repaired using ES cells, it is possible that the result would be far more stable and reliable than the current electronic therapies.

In another exciting study, researchers at Weill-Cornell Medical College and Memorial Sloane-Kettering Cancer Institute found that congenital heart defects could be repaired in utero using ES cells.[29, 30] Congenital heart defects can be highly lethal for newborns and often are difficult to surgically repair at birth. Amazingly, these researchers found that congenital heart defects could be partially repaired simply by injecting ES cells into the afflicted fetal mouse's mother. After being injected, the ES cells were found to secrete certain growth factors and chemical signals that caused the offspring's own heart cells to regenerate themselves.

Among those suffering from Parkinson's disease (PD), a neurodegenerative disorder primarily characterized by the loss of nerve cells within the brain that secrete the neurotransmitter dopamine, stem cell research is viewed as a mean of alleviating a host of symptoms, including tremors, muscle rigidity, and a general slowing of physical movement. This debilitating neurodegenerative disease affects more than 5.5 million Americans. ES cells have been viewed as being useful in treating those with Parkinson's disease. Currently, there are two main types of therapy: dopamine-boosting medication and deep brain stimulation. Medication, which had been the primary mode of therapy, mainly focuses on increasing the amount of dopamine in the brain.

The main dilemma in PD is the loss of dopamine secreting neurons; therefore, to treat PD, scientists must first be able to create these neurons from ES cells. Takagi

et al., of Kyoto University in Japan, were the first to do so at the end of 2004. When these newly created dopamine-secreting neurons were implanted in monkeys with symptoms of Parkinson's disease, tremors were significantly reduced.[31]

Diabetes currently affects more than 16 million Americans. As obesity rates skyrocket in the United States, adult onset diabetes mellitus is increasingly prevalent.[32] Diabetes can lead to a host of problems throughout a person's life, including blindness, loss of limb function, heart disease, and kidney failure. Current diabetes therapies are based on the replacement of insulin through the use of pills or an injection, depending on how much insulin is needed. To match insulin dosage with blood glucose levels (which fluctuate throughout the day), diabetics often are required to test their glucose levels several times each day. The goal is to maintain this delicate balance of insulin and glucose to keep glucose levels as close to normal as possible. Stem cells are being considered as a means to help diabetics better regulate their insulin. As diabetes is the result of losing a single specific type of cell, stem cell therapy to replace this cell population is an optimal treatment. A number of different sources for insulin-producing cells exist.[33] Adult pancreatic stem cells were one of the first sources of cells capable of insulin replacement. However, these cells are in very low numbers within the pancreas, and they do not replicate once removed, making them a difficult source for therapy. Bone marrow stem cells sometimes can produce insulin, and some even will differentiate into fully fledged beta cells.[34, 35] Although they currently differentiate into beta cells very inefficiently, they may be a good source of cells if the right growth factors are identified. More research will need to be done to investigate their potential. In October 2005, three scientists at the University of Wisconsin announced that they had produced synthetic beta islet cells using embryonic stem cells in rats.[36] If human embryonic stem cells could be made to efficiently transform into beta cells capable of producing insulin, and responding to the level of sugar in the blood like real pancreatic beta cells, diabetes and all of its complications would be things of the past.

President Ronald Reagan's death from Alzheimer's disease triggered an outpouring of support for ES cell research. But, in contrast to Parkinson's disease, diabetes, and spinal injuries, Alzheimer's disease involves the loss of huge numbers and varieties of nerve cells in the brain. The complexity of the brain makes stem cells an unlikely therapy for this disease.

Despite the stunning advances in stem cell research, much more needs to be understood before individuals can maximally benefit from all of these potential stem cell therapies. Early research results are extremely promising, but ethical, political, and legal issues have clouded the debate. Detractors of ES cell research tend to play down the scientific merits of stem cell use and focus on the more difficult ethical and moral issues. The crux of the matter is that the extraction of human stem cells to create a stem cell line currently requires the destruction of a harvested embryo. The stem cell debate is now focused on the status of the embryo. Is it a living "human being"? Should embryos be destroyed for the sake of future advances in medical science?

CHALLENGES IN STEM CELL RESEARCH

Since the landmark study by Takahashi and Yamanaka in 2006, a second source of pluripotent stem cells has seemed within reach. The induction of pluripotent stem

cells from harvested adult cells is an exciting alternative source of personalized stem cells to therapeutic cloning, a source that is fraught with potential ethical concerns. However, the insertion of genes into the adult cells to "reprogram" them back to pluripotency has caused some scientists to worry about an increased risk of cancer formation. The genes used to reprogram these cells are essentially "cancer genes," giving the cells the ability to copy themselves indefinitely. Without any form of regulation, this ability to replicate may cause the induced pluripotent stem cells to be even higher risk of causing teratoma or tumor formation than traditional pluripotent stem cells derived from embryos.[37] Researchers have now focused on eliminating the need to use cancer genes used in the reprogramming process. If a method can be found to induce pluripotency in adult cells without inserting potentially dangerous genes it would go a long way to ease fears of tumor formation.

Aside from the possible dangers of inducing disease through stem cell transplantation, there are a number of legal and regulatory factors that could prove difficult to overcome. As stem cells can be harvested from individuals, there is the very intriguing legal question of ownership. Who owns a stem cell? The scientist? The donor? The recipient? These are questions that have been dealt with for some time by sperm banks, organ donation, and egg surrogacy; however, they will arise again with the advent of human stem cell treatments.[37] Along with the question of ownership is the issue of privacy. Like other organ or tissue donations, a stem cell always can be traced back to an individual through the genetic code. For donors of embryos, bone marrow, or cord blood, it will be especially important to have protections in place to ensure that privacy is upheld. The federal government and the NIH currently are examining these issues.

ETHICAL ISSUES

The human embryonic stem cell debate is often framed as part of a larger discussion on the definition of human life and the role of medical science in maintaining it. The extraction of human embryonic stem cells to create a stem cell line currently requires the destruction of the harvested embryo. Although much research is being done on inducing adult cells to become pluripotent stem cells, the challenging nature of this process likely will ensure that embryos will be a major source of stem cell lines in the near future. As a result, the question of the embryo's moral status often is considered the most controversial question in the stem cell research debate. At the center of the dialogue is the question of an embryo's "personhood." Do embryonic stem cells represent a life? That is, are the pluripotent stem cells human and do they have the same rights as born humans? Are pluripotent embryonic stem cells morally protected entities or are they more like other disposable tissues gleaned from the human body?[38] In essence the debate focuses on when life begins, a question for which there is no easy answer.

A current method of avoiding this controversy has been to find ways to extract stem cells without harming or destroying a human embryo. It has been thought that by discovering benign harvesting techniques, stem cell research could be unhampered by the ethical and religious debates surrounding the question of the embryos personhood and human right to life.

Several new methods for producing embryonic stem cell lines have shown great promise. The use of Pre-implantation Genetic Diagnosis (PGD) for the harvesting

of embryonic stem cells is a benign procedure that has been used by IVF clinics for many years to determine the genetic health of embryos before their implantation in the mother's uterus. By using this technique, IVF clinics can avoid using embryos that are predisposed to developing lethal genetic diseases such as Tay-Sachs, Huntington's, Muscular Dystrophy, and Cystic Fibrosis. Since two days after the meeting of the sperm and the egg an embryo consists of only eight cells, by using special techniques, it is possible to remove a single cell while allowing the remaining cells to develop into a human being.[22]

Alexander Meissner and Rudolf Jaenisch of the Whitehead Institute for Biomedical Research recently suggested another alternative for creating stem cell lines without causing the destruction of an embryo: Alternative Nuclear Transfer (ANT), which is designed to create a modified cell that is incapable of fully developing into a human individual.[39] Meissner and Jaenisch believe that if the cell cannot become a human, it cannot be considered to possess personhood.

The response to both of these alternative techniques has been highly varied. After the announcement of the new methods in the journal *Nature*, a spokesman from the U.S. Conference of Catholic Bishops stated that while the two reported techniques still raise some ethical questions, they do represent "a step in the right direction."[40] Some social conservative leaders, such as Rep. Roscoe G. Bartlett of Maryland, a self-described pro-life advocate, believe that "except for the small minority in the pro-life community that doesn't even support IVF therapy, [these techniques] circumvent all of the ethical arguments against stem cell research."[41] Other leaders, like Dr. John Shea, medical advisor to the Campaign for Life Coalition, came out against these techniques, saying that the PGD technique does not benefit the child and thus cannot be used without the child (embryo's) consent.[42] Similarly, Tony Perkins of the Family Research Council, wrote that "it is not clear what effect [PGD] would have on the children born after having had one of their cells removed."[43]

The debate will continue until the answers to a number of important questions can be found. What if stem cells could be produced without embryo loss? Would this then make a difference? As it happens, a small biotech company says that it has found a way to produce human embryonic stem cells without destroying an embryo.[44] Researchers at Advanced Cell Technology grew a colony of stem cells, leaving the embryo unharmed, from a single cell removed from an embryo that had only 8 to 10 cells. Presently, physicians routinely remove a cell from an eight-cell embryo to screen for chromosomal abnormalities before implantation. Hence, logic has it that deriving stem cells from this method adds no additional risk since a diagnostic screening procedure already relies on this technique.

Many questions need to be answered: Would this new technique satisfy those who believe that it is unethical to remove a cell purely for stem cell research? Would this technique satisfy those who believe that life is being destroyed? For those who believe that a single cell removed from an early embryo may have the potential to produce life, this new technique probably will not change their mind. For those who are proponents of stem cell research, what this newest development shows is that stem cells can be produced without destroying an embryo and does not destroy the potential for life.

THE POLITICS OF STEM CELL RESEARCH

With the inherent ethical and moral issues of stem cell research so difficult to resolve, the political debate over stem cell research has raged from its inception. At the end of the 1992 Presidential Campaign, Bill Clinton announced his intention to overturn the de facto prohibition of research on human embryos that had been put in place by President George H. W. Bush. On June 10, 1993, the newly elected President Clinton signed legislation authorizing the NIH to begin to conduct and fund human embryo research. But, worrying that federal funds could be abused for research on human cloning, he issued an executive order in 1994 to prohibit the creation of human embryos for research purposes. To many ethicists, scientists, and politicians, this executive order was deemed insufficient to make sure that the considerable funds of the NIH would not be misused. Therefore, in the summer of 1995, members of Congress decided to attach a rider to the Health and Human Services Appropriations Act that was used to fund the NIH each year. The "Dickey-Wicker amendment," as it became known, prohibited the NIH from using appropriated funds for the creation of human embryos for research purposes. The amendment defined a human embryo as being an organism capable of becoming a human being when implanted in a uterus.[45] Using this broad language, it initially was thought that the act prevented the use of federal funds for almost any research related to human embryonic stem cells.

In 1998, after the initial successes of the research groups from the University of Wisconsin and Johns Hopkins, the Clinton administration decided to reevaluate its position on the support of embryonic stem cell research. The NIH requested a legal opinion from the Department of Health and Human Services (DHHS) on whether federal funds could be made available to researchers working with the human ES cells produced by the groups of Wisconsin and Johns Hopkins. In January 1999, Harriet Rabb, the general counsel of DHHS, found that the Dickey-Wicker amendment could not apply to human embryonic stem cells. The Dickey-Wicker amendment officially defines a human embryo as being an *organism* capable of becoming a human being when implanted in a uterus. Because an ES cell cannot develop into a human being even when implanted in a uterus, Rabb determined that it could not be considered a human embryo. According to this logic, the DHHS maintained that despite the amendment, it could fund any research related to human ES cells, as long the cells *initially* were created with private funding.[3] That is, after careful consideration, DHHS concluded that because human pluripotent cells are not embryos, current federal law does not prohibit DHHS funds from being used for research utilizing these cells.

In April of 1999, NIH director Harold Varmus appointed an oversight committee to begin drafting guidelines and oversight for the federal funding of ES cell research. The working group included scientists, clinicians, ethicists, lawyers, patients, and patient advocates. By February 2000, more than 50,000 comments had been received by experts in fields as far ranging as medicine, philosophy, ethics, biology, and neuroscience. Finally, in summer 2000, NIH published the final set of guidelines, NIH Guidelines for Research Using Human Pluripotent Stem Cells, in the *Federal Register*, which became effective on August 25, 2000.[46] The purpose

of the NIH Guidelines was to set forth procedures to help ensure that NIH-funded stem cell research was conducted in an ethical and legal manner. Among other stipulations, the guidelines prescribed that for studies using human pluripotent stem cells derived from human embryos, NIH funds may be used only if the cells were derived from frozen embryos that were created for the purpose of fertility treatment, were in excess of clinical need, and were obtained after the consent of the donating couple.

The Clinton administration's guidelines for stem cell research were actually relatively conservative in comparison to the policies of other developed countries. In accordance with the Dickey-Wicker amendment, which had been renewed on every DHHS appropriations bills since 1997, the guidelines only allowed federal funding for studies using stem cells derived from embryos created for the purposes of *in vitro* fertilization, and only if they were in excess of the clinical need for such embryos. In addition, it was decided that the NIH could not fund any research that actually involved the derivation or creation of ES cells, as this was explicitly barred by the Dickey-Wicker amendment.[47]

Furthermore, the Clinton administration decided to outlaw the use of NIH funds for research involving ES cells derived using therapeutic cloning (or SCNT), even if the actual derivation of the cells was performed with private funds.[48] SCNT is the most well-researched technique that potentially could create embryonic stem cells that are genetically identical to an individual. That is, the cloned cell is used to create a stem cell line (not to create a new human being) that would be a perfect genetic match for a patient. Stem cells that are created by this method would likely avoid immune rejection, the primary concern of tissue transplantation. Without the ability to use therapeutic cloning, scientists utilizing federal funding would not be able to participate in research related to the "personalized medicine" that had become the ultimate goal for many stem cell researchers.

With the new guidelines in place, the NIH began to accept grant applications from research projects using human ES cells. The first review of these grants was supposed to occur by April 2001, several months after the Clinton administration left office. In mid-April, however, the DHHS decided to postpone the meeting until the incoming Bush administration could review the department's policies. After several months of review, on August 9, 2001, President George W. Bush announced the first federal funding of human embryonic stem cell research. Funding, however, would be available only to researchers using the 78 human ES cell lines that had been created before that date. President Bush believed that the government could explore the promise and potential of stem cell research without crossing a fundamental moral line. Of the 78 cell lines that were originally eligible for federal funding, only 15 are currently available. The remainder of the eligible stem cell lines was either unavailable or unsuitable for research. With so few lines actually available, relatively few federal dollars actually have been spent on human stem cell research.

During the U.S. presidential election of 2008, stem cell research again rose to the forefront of national political debate. Both major party candidates for president, then Democratic senator Barack Obama and Republican senator John McCain, announced their strong support for increasing federal funding of embryonic stem cell research. Mr. Obama even declared that soon after assuming office he would

repeal the Bush administration's 2001 executive order preventing the use of federal funds for research involving embryonic stem cell lines created after August 9, 2001.[49] Although neither candidate made explicit the degree to which they would increase funding, proponents of stem cell research were ecstatic to receive the support of both major party candidates.

In January 2009, upon the inauguration of Mr. Obama as president, supporters of stem cell research received a significant boost. On March 9, 2009, just weeks after he was sworn into office, he officially repealed the 2001 Bush executive order, laying the groundwork for federal funding of new embryonic stem cell lines. In the executive order, the president stated,

> for the past eight years, the authority of the Department of Health and Human Services, including the National Institutes of Health (NIH), to fund and conduct human embryonic stem cell research has been limited by Presidential actions. The purpose of this order is to remove these limitations on scientific inquiry, to expand NIH support for the exploration of human stem cell research, and in so doing to enhance the contribution of America's scientists to important new discoveries and new therapies for the benefit of humankind.[50]

As of the writing of this chapter, Congress was just beginning to reexamine the issue of stem cell research. The Dickey-Wicker amendment, which bans the use of federal funds to create embryos for research purposes, is under examination. Rep. Diana DeGette (D-Col.), a staunch supporter of stem cell research, is planning on reintroducing the Castle-DeGette bill by the summer of 2009 to officially sanction the use of federal funds for embryonic stem cell research on stem cell lines created after 2001.[51] The broad language used by President Obama in his executive order has given the NIH the power to redefine the specific federal funding guidelines for stem cell research, effectively returning science to the scientists.

Since the election of Mr. Obama in November 2008, the scientific community has hoped for a major increase in the federal funding of stem cell research. The NIH budget for the fiscal years 2008 and 2009 was just over $29 billion. In 2008, stem cell research accounted for $938 million, but only $88 million was given to researchers working with human embryonic stem cell lines.[52] It is not yet clear how much the federal funding of stem research would change over the course of the year.

The Obama administration used broad language in its executive order calling for the use of federal funds for stem cell research, effectively allowing the NIH to determine the details of the new federal funding policy. The NIH was due to release new guidelines outlining these new funding policies in the summer of 2009. As of yet the U.S. Congress has not yet passed a bill to complement the strongly worded executive order released by the Obama administration in March 2009. The Dickey-Wicker amendment, which still outlaws the use of federal funds for embryos created solely for research purposes, may need to be repealed or amended in order for the NIH to legally fund some of the most promising stem cell research. Therapeutic cloning (SCNT), for example, in which the genetic material of an adult patient is injected into a human egg to create a personalized stem cell line, would not be legal under the Dickey-Wicker amendment.

Rep. Diana DeGette, one of the original sponsors of the pro stem cell research Castle-DeGette bill twice vetoed by former President Bush, stated that "in consultation

with experts," the sponsors of a new bill "are reviewing past legislative efforts to assess what needs to be done [legislatively] going forward."[49] The congressional climate has changed dramatically since the original Castle-DeGette bill, with a great majority of congressional leaders now supporting the research. Currently, stem cell research proponents have high hopes that Congress will move swiftly to allow for significant funding increases in the year 2009. Whether or not the federal government will be able to support the use of therapeutic cloning, or even the creation of embryonic stem cell lines from embryos created specifically for research, remains a question. All 78 embryonic stem cell lines that currently are eligible for federal funding were created from excess embryos donated from in vitro fertilization clinics. Even with the new executive order, the only ES cell lines eligible for federal funding will continue to be those created from donated surplus IVF embryos. Without congressional action and a more permanent change in federal law, this would continue to be a significant limitation for researchers interested in creating personalized stem cell lines for therapeutic interventions free of immune rejection.

WHERE DO WE GO FROM HERE?

Few areas of biomedical science have aroused as much controversy as ES cell research. Since the derivation of the first human ES cells in 1998, the issue has been at the forefront of scientific, political, and ethical debate. Proponents tend to emphasize the considerable therapeutic potential of stem cell research while opponents speak of the immorality of using human cells for this purpose. Yet, to those individuals suffering from debilitating diseases for which stem cells may offer a cure, such as Parkinson's disease, diabetes, and spinal cord injuries, they view the use of ES cells as the best means to treat or even cure their illness.

Stem cell research involves such unprecedented opportunities to improve medical science that it is hard not to be overwhelmed by its sheer potential. The major legal, ethical, religious, and political hurdles continue to fuel the debate. Both proponents and opponents make cogent arguments for and against ES cell research. What is needed is a scientific resolution to the moral dilemmas, with input from both science and medical ethics. Yet, given the scope of the issue, it is unlikely that the issue will be resolved quickly, and the broader application of ES cell research to those who potentially could benefit is still a hope and a dream. Hopefully researchers will continue to discover new ways of creating and using stem cells, making their future even brighter.

REFERENCES

1. United Network of Organ Sharing. UNOS Scientific Registry. www.UNOS.org. Accessed May 29, 2009.
2. Prosper F, Verfaillie CM. Human pluripotent stem cells from bone marrow. In: Chiu A, Rao MS. *Human Embryonic Stem Cells*. Totowa, NJ: Humana Press; 2003: 89–112.
3. Fischbach GD, Fischbach RL. Stem cells: science, policy, and ethics. *J Clin Invest*. 2004;114:1364–1370.
4. Bongso A, Lee EH. Stem cells: their definition, classification, and sources. In: Bongso A, Lee EH, eds. *Stem Cells: From Bench to Bedside*. NJ: World Scientific; 2005: 1–13.

5. Parson A. Proteus effect: stem cells and their promise in medicine. Washington, DC: Joseph Henry Press; 2004.
6. Van Bekkum DW. Bone marrow transplantation. *Transplant Proc*. 1977;9:147–154.
7. Usdin S. Ethical issues associated with pluripotent stem cells. In: Chiu A, Rao MS, eds. *Human Embryonic Stem Cells*. Totowa, NJ: Humana Press; 2003:3–26.
8. Thompson JA, et al. Embryonic stem cell lines derived from human blastocysts. *Science*. 1998;282:1145–1147.
9. Shamblott MJ, et al. Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc Natl Acad Sci U S A*. 1998;95(23):13726–13731.
10. Breakthrough of the year. *Science*. 1999;286(5448):2221–2416.
11. Wilmut I, Paterson LA. Stem cells and cloning. In: Sell S, ed. *Stem Cells Handbook*. Totowa, NJ: Humana Press; 2004.
12. Horb ME, et al. Experimental conversion of liver to pancreas. *Curr Biol*. 2003;13:105–115.
13. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006;126:663–676.
14. Kastenberg, ZJ. and Odorico, JS. Alternative sources of pluripotency: science, ethics and stem cells. *Transplant Rev*. 2008;22:215–222.
15. Wilmut I. The first direct reprogramming of adult human fibroblasts. *Cell Stem Cell*. 2007;1:593–594.
16. Hwang SW, et al. Patient-specific embryonic stem cells derived from human SCNT blastocysts. *Science*. 2005;308(5729):1777–1783.
17. Cyranoski D. South Korean scandal rocks stem cell community. *Natl Med*. 2006;12:4.
18. Cervera RP, Stojkovic M. Commentary: somatic cell nuclear transfer: progress and promise. *Stem Cells*. 2008;26:494–495.
19. French AJ, et al. Development of human cloned blastocysts following somatic cell nuclear transfer with adult fibroblasts. *Stem Cells*. 2008;26:485–493.
20. Yu J, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007;318:1917–1920.
21. Silva J, et al. Nanog promotes transfer of pluripotency after cell fusion. *Nature*. 2006;441:997–1001.
22. Chung Y, et al. Embryonic and extraembryonic stem cell lines derived from single mouse blastomeres. *Nature*. 2006;439:216–219.
23. Hochedlinger K, Plath K. Epigenetic reprogramming and induced pluripotency. *Development*. 2009;136:509–523.
24. Condic ML, Rao M. Regulatory issues for personalized pluripotent cells. *Stem Cells*. 2008;26:2753–2758.
25. O'Brien TA, et al. No longer a biological waste product: umbilical cord blood. *Med J Aust*. 2006;184:407–410.
26. Advani AS, Laughlin MJ. Umbilical cord blood transplantation for acute myeloid leukemia. *Curr Opin Hematol*. 2009;16:124–128.
27. Young MJ. Stem cells in the mammalian eye: a tool for retinal repair. *APMIS*. 2005;113:845–857.
28. Centers for Disease Control and Prevention. CDC Statistics. www.cdc.gov/nchs/fastats/lcod.htm. Published June 20, 2006.
29. Fraidenreich D, et al. Rescue of cardiac defects in id knockout embryos by injection of embryonic stem cells. *Science*. 2004;306(5694):247–252.
30. Fraidenreich D, Benezra R. Embryonic stem cells prevent developmental cardiac defects in mice. *Nature Clin Practice Cardiovasc Med*. 2006;3:S14–S17.
31. Takagi Y, et al. Dopaminergic neurons generated from monkey embryonic stem cells function in a Parkinson primate model. *J Clin Invest*. 2005;115:102–109.

32. Boyle JP, et al. Projection of diabetes burden through 2050. *Diabetes Care*. 2001; 24:1936–1940.
33. Jones PM, et al. Cell based treatments for diabetes. *Drug Discov Today*. 2008;13:888–893.
34. Tang DQ, et al. In vivo and in vitro characterization of insulin-producing cells obtained from murine bone marrow. *Diabetes*. 2004;53:1721–1732.
35. Karielli O, et al. Generation of insulin-producing cells from bone marrow mesenchymal stem cells by genetic manipulation. *Stem Cells*. 2007;25:2837–2844.
36. Wisconsin Technology Web site. http://www.wisctechnology.com/article.php?id-2340. Accessed July 6, 2006.
37. Ahrlund-Richter L, et al. Isolation and production of cells suitable for human therapy: challenges ahead. *Cell Stem Cell*. 2009;4:20–26.
38. Green RM. *The Human Embryo Research Debates: Bioethics in the Vortex of Controversy*. New York: Oxford University Press; 2001.
39. Meissner A, Jaenisch R. Generation of nuclear transfer-derived pluripotent ES cells from cloned cdx2-deficient blastocysts. *Nature*. 2006;439(7073):212–215.
40. Studies may calm stem cell qualms. Associated Press. October 16, 2005.
41. Wade N. Stem cells with ethics. *New York Times*. October 16, 2005.
42. Bush's stem cell policy received with mixed emotions. http://www.lifesite.net.
43. Perkins T. Embryonic stem cell studies raise questions, not cures. *Washington Update*. October 16, 2005.
44. Stem cell without embryo loss. *New York Times*. August 26, 2006: A35.
45. Bonnicksen AL. *Crafting a Cloning Policy: From Dolly to Stem Cells*. Washington DC: Georgetown University Press; 2002: 77–79.
46. National Institutes of Health. The need for guidelines to govern research using pluripotent stem cells. http://www.nih.gov/news/stemcell/index.htm.
47. Wertz DC. Embryo and stem cell research in the United States: history and politics. *Gene Therapy*. 2002;9:674–678.
48. Casell JH. Lengthening the stem: allowing federally funded researchers to derive human pluripotent stem cells from embryos. *Univ Mich J Law Reform*. 2001;34:547–572.
49. Holden C. A fresh start for embryonic stem cells. *Science*. 2008;322:1619.
50. Obama B. Presidential Executive Order: Removing barriers to responsible scientific research involving human stem cells. White House Press Release. March 9, 2009.
51. Holden C. For congress and NIH, Headaches ahead on stem cells. *Science*. 2009;323:1552–1553.
52. National Institutes of Health Research Portfolio Online Reporting Tool. Estimates of funding for various research, condition, and disease categories. http://www.HHS.gov. Published January 15, 2009.

CHAPTER 13

Application of Novel Analytical Tools in Global Disease Monitoring: Remote Sensing in Public Health Research and Practice

Jesse C. McEntee, MA, Denise Castronovo, MS, Jyotsna S. Jagai, MS, MPH, PhD, Kenneth K. H. Chui, MS, MPH, PhD, and Elena N. Naumova, PhD

INTRODUCTION

Epidemiological research typically aims to characterize disease occurrence in terms of geographic space, temporal pattern, and human behavior. Community-based or hospital-based biomonitoring that tracks these three dimensions of disease incidence require the power of sophisticated analytical tools to extract useful information from multisourced databases. Such tools may include statistical and mathematical modeling, simulation techniques, Geographic Information Systems (GIS), and dynamic mapping. They enhance traditional methods of epidemiologic investigations, allow effective use of emerging or underutilized data sources, and facilitate comprehensive approaches to data visualization across space and time. They help to better understand the causes of disease dynamics; combine traditional surveillance data, hospital records, and vital statistics with novel sources of data that describe environment, social infrastructure, and cultural background; and eventually track progress in alleviating disease risks and effects. During the past decade, public health professionals have accrued skill in informatics, modeling, mobile communication, and GIS analysis that can facilitate linking disease risk and environmental variables as well as enhance active disease control and prevention.

This chapter provides an overview of a number of advanced computational and analytical techniques that open new opportunities to examine the role of environmental drivers and forecast disease transmission and manifestation. We review applications of various remote-sensing (RS) techniques and present the relatively nascent epidemiological applications of this technology. Although our review provides an in-depth examination of a number of remote sensing techniques for monitoring disease, we acknowledge that it is not a comprehensive review of *all* techniques in existence in the 21st century. The particular focus of this article is on remote sensing applications relating to satellite-retrieved data in order to answer research questions in the field of public health. We will first illustrate the utility of novel technologies, practical applications, and examples. Then we will delineate

key principles in data acquisitions and analyses. Finally, we will highlight future directions for public health research.

WHAT IS REMOTE SENSING?

The American Society for Photogrammetry and Remote Sensing states that remote sensing techniques are used to gather and process information about an object without direct physical contact.[1] RS emerged in the 1960s and 1970s as a result of intensified wartime investment in aerial photo interpretation.[2] Before 1972, all remotely sensed images were either obtained from ground sensors or aerial cameras on planes. On July 23, 1972, the *Landsat 1* land-surface observation satellite system began recording Earth resource data, which changed everything.[1] Traditionally used to observe land cover information about objects and geographic areas on the Earth's surface by measuring electromagnetic energy emitted from these areas, RS increasingly has been used by geographers and epidemiologists to measure environmental variables that have direct pertinence to the spread of infectious disease. Using RS to predict epidemics of infectious disease has been shown to be cost-effective and helpful.[3, 4] The framework for utilizing remotely sensed data in epidemiology involves linking measures of radiation, typically obtained by satellite, to measures of the geographic distribution of a disease and its vector.[5]

On February 16, 2005, a global initiative was implemented in which 61 countries agreed to implement the Global Earth Observation System of Systems (GEOSS), which aims to bring together countries' global Earth Observation System (EOS) hardware and software to streamline format, ensure compatibility, and incorporate shared international resources, such as ocean buoys, satellites, and weather monitoring stations.[6] EOS data are used to measure characteristics of air, water, and land through the use of visible and nonvisible radiation. The GEOSS effort represents a heightened awareness of the potential value of remotely sensed imagery in a number of fields, such as the biological sciences, urban planning, and public health.

Remotely sensed imagery is obtained through measurement of reflectance. Depending on the wavelength that is measured, a number of precise data values can be obtained and inputted into computer algorithms that subsequently are used to measure land use values.[7] Spatial epidemiology aims to analyze the spatial distribution of disease data to identify risk populations and possible causal factors.[8] RS technology, for example, could serve to provide a more accurate identification of disease-carrying vector habitat.

Key Principles of Remote Sensing Data

Public health research and practice utilize a wide variety of health outcomes that originate from large data systems, including vital statistics records, hospitalization claims, local, regional, and national registry of diseases, diagnostic data repositories, and surveillance systems. Typically, each source has its own population coverage, timeliness, and ability to properly reflect disease presence (see figure 13.1). These important characteristics have to be taken into account when analyzed to detect systematic patterns and particular aberrations. Technological advancements in compiling and exchanging large volumes of data reveal new potential for

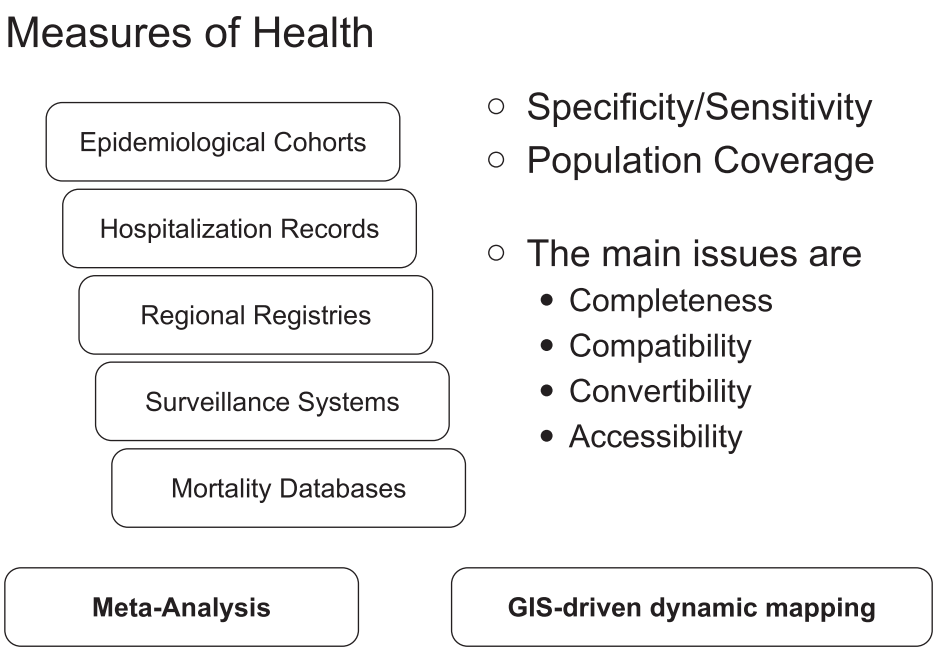

Figure 13.1 Measures of Health.

knowledge mining and retrieval.[9] This process also reveals the need for better understanding of limitations and requirements for a new generation of disease-tracking systems.

Worldwide disease tracking systems historically are supported by a network of hospitals, outpatient clinics, and diagnostic facilities and are operated in cooperation with local and regional public health institutions. Overall, the established infrastructure facilitates consistent improvement of data quality, essential for reliable disease monitoring. The major data sources that form the basis for many national surveillance systems strive to ensure a complete and comprehensive coverage of a population.

The most recent lessons of remote-sensing data (RSD) utilization in public health practice and research can be summarized as following. Initially, *conceptual support* is needed for a study, followed by *compatibility* of health outcome measures and exposure measures derived from RS. Ideally, there should be complete uninterrupted temporal overlap in health and exposure data. Such temporal *completeness* expands the array of applicable analytical techniques; many statistical methods for time-series data have serious limitation if there are gaps in the time-series. Poor overlap in time periods of available data for health outcomes and exposure measure may lead to a substantial reduction in statistical power to detect an effect because the sample size available for the analysis will be smaller in a joint time-series. Typically, time-referenced health outcome data are recorded as daily, weekly, monthly, quarterly, or annual counts of health-related events. The temporal period of RSD depends on orbits of satellites; therefore, the measurements are collected in corresponded temporal scales, for example, every 10 or 16 days. Therefore, to properly link RSD with health outcome it is important to ensure *convertibility of time units*

and perform a basic time unit alignment. Finally, the *spatial alignment* has to be carefully addressed; health outcome data need to be georeferenced to ensure proper selection of the target or catchments areas for abstracting RSD.[10]

Public Health Applications of Remote Sensing

Pavlovsky [11] was the first to explicitly state that diseases are found only where the environmental habitat requirements of the parasite, host, and vector are satisfied. As obvious as it may seem, this logical inference began to form what is known in the 21st century as spatial epidemiology. Efforts to illustrate this assertion continue today. To effectively use RS to monitor or predict infectious disease, the disease's habitat must be known. Conversely, a disease's habitat may be revealed through the analysis of remote sensing data given known patterns of outbreak. Some researchers have cited the inadequacy of current infectious disease surveillance and response systems and the corresponding benefits of improving these systems, such as advanced surveillance and modeling that can predict the temporal and spatial risks of epidemics using environmental data.[3, 12, 13, 14] For example, remotely sensed data can be converted into vegetation, land surface temperature (LST), atmospheric moisture (also referred to as cold cloud duration [CCD]), soil moisture, rainfall, and sea surface temperature (SST) indexes, which in turn can be used to track infectious diseases for public health purposes.[15]

MAJOR TYPES OF RS INDEXES

Variations in vegetation cover on the Earth's surface are indicative of different biological, meteorological, and human-induced conditions. Orthophotos of the Earth's surface can be incredibly revealing in illustrating what types of vegetation lie below. In 1971, for example, an early NASA flight yielded strong empirical evidence of a relationship between *Aedes sollicitans* (salt marsh mosquitoes) and water hyssop (wetland plant species) using color, color-infrared and multiband sensor, and film combinations.[13] In the 21st century, the normalized difference vegetation index (NDVI) is the most commonly used vegetation index in epidemiological studies because of its ability to be an indicator of moisture regime (NDVI is most commonly used to measure chlorophyll or "greenness" [16]) and its ability to be derived from satellite sensors that produce a regular and frequent time-series (see figure 13.2).[10, 17]

NDVI values are directly affected by both the time of year due to plant senescence as well as the types of vegetation in the study. If study areas that contain less than 20 percent ground cover are unable to yield accurate NDVI values, this could be a major limiting factor in using NDVI to study urban areas. Although debate exists over appropriate application of NDVI data, it remains not only the most commonly used vegetation index, but also the most widely used RS variable in epidemiological studies. This wide use could be attributed to the fact that vegetation often is referenced as representing the combined impact of rainfall, temperature, humidity, topography, soil, water availability, and human activities. Vegetation readings are independent of height aboveground, which makes it a desirable proxy because this can vary between the top surface of biomass and the enclosed climate.[18]

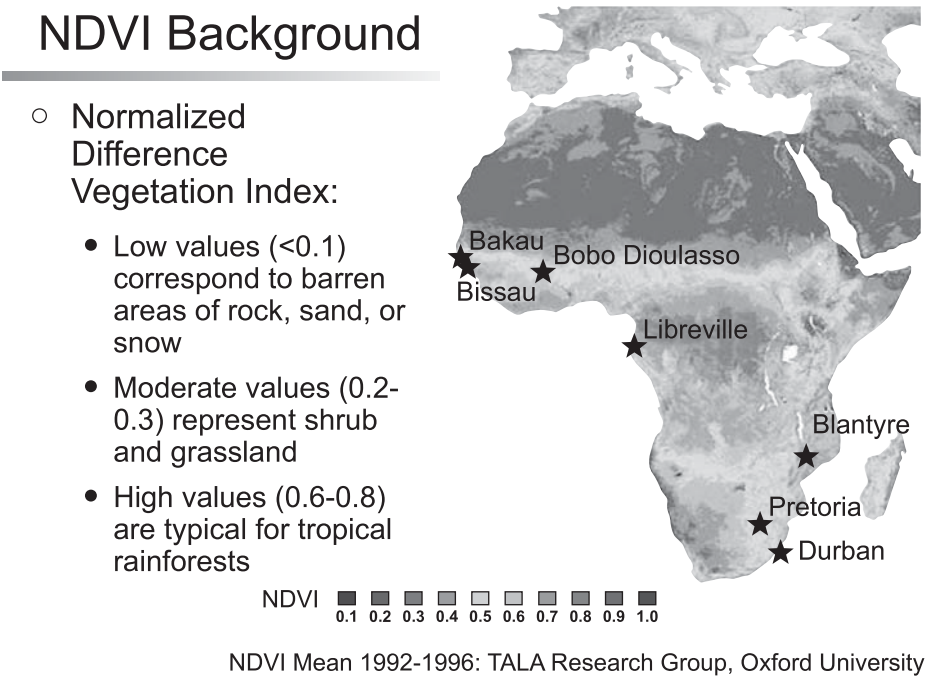

Figure 13.2 NDVI Background.
Source: Hay et al., (2006) *Advances in Parasitology*, 62, 37–77.

Other tools include Spectral Vegetation Indexes (SVIs) and the Global Vegetation Moisture Index (GMVI), which also are used to provide information on vegetation water content at the canopy level. NDVI should not be used for this purpose because it should not be assumed that in all species plant chlorophyll content is related to water content.

An example of the use of these techniques is illustrated by the Bavia et al. study designed to identify the landscape epidemiology of American *Visceral Leishmaniasis* (VL) in Bahia, Brazil.[19] Using NDVI and climate data on rainfall and temperature as well, their GIS consisted of political maps of the study area, georeferenced maps of the 33 municipal study areas, ground-collected vegetation information, NDVI values for municipalities, and disease prevalence information. Using geospatial methods, statistical analysis yielded correlations that could be used to estimate the spatial distributions of VL.

Kitron and Kazmierczak [16] used similar techniques to map Lyme disease distribution in Wisconsin. Specifically, they collected county-level data on tick distribution, human population density, Lyme disease case distribution, and proportion of wooded areas to assess measurement techniques used in the explanation of tick distribution in Wisconsin. The researchers used NDVI data to measure "greenness" in the spring and fall to find associations between county-level NDVI values, tick distribution, human population density, number of cases by county of exposure and residence, and incidence rates. Significant correlations between NDVI values and

human exposure and tick distribution were found in northern regions of the state. This correlation decreased in southern sections of the state, becoming negatively correlated in highly populated deforested areas.

These two studies represent typical applications of NDVI in attempting to use RS data for epidemiological purposes. Although other methods exist, correlating NDVI data to other variables involved in disease placement (typically disease vectors and hosts) and their habitat is quite common.

Land Surface Temperature and Diurnal Temperature Difference

Different land, soil, and geologic compositions emit radiance differently across thermal spectrums, which therefore indicates a range of temperatures that may be suitable for certain diseases and their vectors. A number of studies have successfully correlated either LST or temperature difference to environmental variables that could be used to measure disease distribution. Malone et al., for example, found that temperature difference maps could be used to determine water table levels. They concluded that surface water may be an environmental determinant of *S. mansoni* infection risk in the Nile Delta.[20]

Sea Surface Temperature

Although multiple studies exist that use SST for epidemiological applications, only a few use remotely sensed SST data. One study that has made a notable contribution is by Lobitz et al.[21] Because SST has been shown to be related to phytoplankton concentration and sea surface height (SSH), researchers investigated whether a temporal trend exists between remotely sensed image availability and timing of the disease being studied—cholera in this case. By superimposing cholera incidence plots on SST and SSH data maps, researchers found a consistent annual bimodal cycle trend with certain years of outbreaks occurring on a level of statistical significance. The real-world application of this finding is vital to the future of cholera monitoring techniques, especially in vulnerable areas such as Bangladesh (the location of the Lobitz et al. study). A temporal lag was found between the time when SST increased, SSH rose, and the cholera outbreak occurred; this remotely sensed data could be used to predict future outbreaks.

Strong and McClain [22] found that using remotely sensed data was a reliable source of SST data, which they verified by examining data from stationary as well as drifting buoys. Remotely sensed SST data has proven useful in monitoring the environmental phenomenon El Niño, which causes considerable economic damage to South American countries when the global trade winds relax in the central and western Pacific as a result of changes in the marine food chain.[1]

Rainfall Indexes

Hay and Lennon [23] established that rainfall was more effectively predicted by remotely sensed sources than interpolation methods. Others have used this as a rational basis for using CCD as a proxy for rainfall because the CCD pixels in a remotely sensed image represent the time that that particular location was covered by rain clouds.[15] Some researchers simply use CCD as another remotely sensed variable to find relationships, such as Thompson et al.'s study of meningitis in Africa.[24] Most studies use CCD as a proxy for rainfall and therefore as measure of suitable habitat for disease vectors.

The Use of Novel Analytical Approaches to Understand Seasonality of Waterborne Infection

Waterborne infections are caused by pathogens ingested with contaminated drinking or recreational water. Cholera, typhoid, and emerging infectious diseases caused by waterborne pathogens (*Cryptosporidium*, *Campylobacter*, *Giardia*, rotavirus, norovirus, *E. coli*, and potentially *Salmonella* and *Shigella*) are examples of waterborne diseases. These pathogens (protozoa, bacteria, viruses) of human and animal waste and feces naturally are present in water bodies; however, their concentration, pathogenicity, and the effect on human health might change dramatically because of natural and man-made changes in the environment. For example, cryptosporidiosis is a diarrheal illness caused by protozoan, which are spread though water and food contaminated with human or animal waste and feces. Although cryptosporidiosis is self-limiting in immunocompetent people, it can be life threatening to immunocompromised individuals, such as those with AIDS and the malnourished.

Considering the nature and etiology of waterborne diarrheal infections, it is likely that seasonal patterns of diseases, such as cryptosporidiosis can be predicted with RSD on a global scale. Cryptopsporidiosis typically manifests through a low endemic level and well-pronounced seasonal outbursts, indicating a strong effect of meteorological and environmental factors. Studies conducted in tropical climates, for example, have found an increase in cryptosporidiosis incidence during the rainy season.[25–27] In the temperate climate of Massachusetts, cryptosporidiosis peaks about six weeks after ambient temperature reached its annual maxima.[28] The seasonal patterns in cryptosporidiosis incidence can be affected substantially by seasonal variations in exposure level associated with water quality, access to clean water, as well as wildlife and agricultural activities.

Jagai et al. [10] conducted a meta-analysis to examine how an increase in cryptosporidiosis relates to precipitation and ambient temperature and to investigate the potential use of NDVI as a proxy for exposure to *Cryptosporidium parvum*. Included in the analysis were more than 60 locations worldwide, representing a wide diversity of climates, that reported monthly cryptosporidiosis incidence over one year or longer. The study showed that an increase in temperature and precipitation predicted an increase in cryptosporidiosis; the strength of relationship varied by climate subcategory. In moist tropical locations, for example, precipitation is a strong seasonal driver for cryptosporidiosis whereas in mid-latitude and temperate climates, temperature is the driver. Cryptosporidiosis infection rates increased after heavy rainfall because of an elevated presence of recreational and drinking water.

This example provides a quantitative link between the incidence of cryptosporidiosis and meteorological parameters on a global scale and illustrates a strong potential for NDVI as a suitable proxy for exposure to *Cryptosporidium parvum*, especially in the humid mid-latitude climate zones. Of course, extreme meteorological events, such as heavy rainfall, droughts, and heat waves, may substantially alter a seasonal pattern in disease incidence. In warm and wet locations, expected precipitation can serve as a reliable predictor for incidence of cryptosporidiosis with one month lead time. Weather forecasting on a local and global scale can be useful for disease forecasting, so public health measures for disease prevention can be better targeted and focused. Because the mean NDVI is strongly associated with temperature in specific geographic regions, it is very likely that RSD and indexes reflecting vegetation water content in particular can be very useful for predicting the incidence

of waterborne infections on a large geographic scale and RSD can be applied to a variety of waterborne diseases and infections caused by thermosensitive pathogens.

CHALLENGES OF RSD APPLICATIONS

The major challenge in using RSD data in public health application is our ability to achieve proper temporal and spatial alignment. In the examples presented in this chapter, each study location had specific challenges in gathering data: large urban areas and coastal regions needed to be properly treated in the data extraction process, NDVI values obtained from different sources have different time and image resolution, and access to complete data can be a challenge as well. In using RSD, it is crucial to develop methodologies for uniform study area identification and relevant data abstraction. This is one of the many challenges facing emerging fields of RS analysis.

That being said, perhaps the greatest challenge facing the remote sensing community is the potential discontinuation of the *Landsat* series of satellites. For more than 30 years *Landsat* images have informed decisions on relationships between land use, water quality, agricultural production, vegetation health, and epidemiology. In the *Strategic Plan for the U.S. Integrated Earth Observation System*, the authors acknowledge the benefits of *Landsat* as well as the funding maladies, but they offer no pragmatic solution other than a vague need for global observation systems.[29][1] A functional EOS relies on complete data sets, including not only remotely sensed data, but also data on disease monitoring, population characteristics, food production, and natural disasters.

Although *Landsat* faces a set of funding issues that most likely will be remedied by a government contract with a commercial entity, some experts have encouraged international cooperation to account for data gaps.[30] The 1992 Land Remote Sensing Policy Act assessed the potential of four different avenues for continuing *Landsat* and national RS operations. Included were an assessment of private sector funding for RS systems, establishment of an international consortium for RS systems, a solely U.S.-funded and -managed RS system, and a cooperative effort between the U.S. government and commercial entities for future RS systems. The 1992 Act's goals became more important, however, when *Landsat 7* suffered a severe malfunction in 2003, thus disrupting data continuity until 2011 when the *Landsat Data Continuity Project Mission* is scheduled to launch.

FULL INTEGRATION IN PUBLIC HEALTH RESEARCH AND PRACTICE

Integration of remote sensing-related resources, including personnel, technology, and data, has yet to occur. For instance, though multiple strategies exist that examine RS data and disease location, the general public as well as organizations that could benefit from this research, have yet to be made aware of its existence and

1. According to this document, "the development of the Strategic Plan addresses the first goal of the Interagency Working Group on Earth Observations and serves as the initial step towards the development and implementation of the US Integrated Earth Observation System" (Interagency Working Group on Earth Observations 2005, 10).

successes.[31] In addition, new RS technologies have yet to be applied widely in many fields that could benefit, simply because of the difficulty of learning how to use the technology. Communication, interpretation, and cooperation are elements that must be improved for RS and GIS systems to be truly integrated across disciplines. Improved environmental monitoring is needed to realize not only geologic and natural system application, but also health-related data recording, such as cancer registries. Cromley [32] recommended that the distribution of GIS data meet the needs of the larger research community and the general public. Greater communication between epidemiologists and software programmers, as suggested by Graham et al. [33] certainly would help. Most important, not only is expert-oriented support for spatial analysis and RS techniques needed, but societywide changes must be made that pervade not only these respective fields, but also the institutions that influence them. The *Strategic Plan for the U.S. Integrated Earth Observation System* set forth a four-point plan for this type of integration:

1. Policy and Planning Integration: To maximize synergies brought about by unforeseen applications of earth observations research (research and operational oriented).
2. Issue and Problem-focused Integration: Align multiple societal benefit areas (e.g. climate disasters, agriculture) in order to pool resources as well as successfully communicate results across interest areas in common, consistent, and understandable terms.
3. Scientific Integration: Integrate information about Earth process modeling across scientific sub-disciplines, which involves comparison of data analysis processes as well as data that is collected through satellites and *in situ* measurements.
4. Technical Systems Integration: Coordinate observation system technology and data management systems, which result in research and operation applications.[29]

These recommendations address the major overarching issues standing in the way of a truly integrated EOS. Although the report from which these issues originated does a fairly adequate job of describing these points, no clear policy guidelines are provided. This omission, in combination with a lack of government funding in RS technologies, indicates an overall lack of priority on RS-related technologies, which takes the form of inadequate funding in both execution of and research into developing new technologies. Public health fields, including epidemiology are perhaps most severely affected by potential gaps in data availability, accessibility, and data quality because creation and utilization of risk-based knowledge requires careful validation between in situ and satellite measurements.

FUTURE DIRECTIONS FOR PUBLIC HEALTH APPLICATIONS OF RS

Currently, there are two primary purposes of applying RS in the field of public health. The first set of applications deals with disease-vector habitat identification. These applications involve identifying suitable habitat for disease vectors; these types of studies usually take place in stable environments in which conditions are static and/or are highly predictable and account for the majority of the studies reviewed in this paper. The second type of application involves natural phenomena and aberrations for which the conditions being studied are either for identification of vector habitats or direct environmental impacts of these events on human health (for example, effects of volcanic ash on human cardiovascular systems). The events

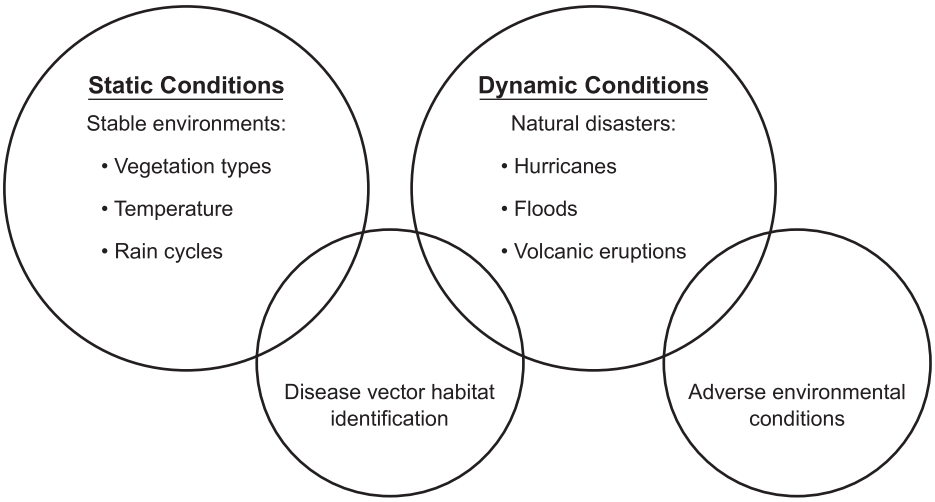

Figure 13.3 Interrelationship of RS and Public Health.

are typically dynamic and unpredictable. As illustrated in figure 13.3, both types of applications can involve disease-vector habitat identification.

In addition to studying the direct relationships along the transmission pathway and potential interaction between environment and human activities, RS would be valuable for other secondary purposes. For instance, RS can provide information on susceptibility or vulnerability of a geographic entity to certain infectious disease.[34] Examples include assessments of the risk of wild fire,[35] earthquake, flood, and landslide.[36] Another valuable use of RS would be locating and assessing physical destruction brought about by natural disasters, such as hurricanes,[37] floods,[38] landslides,[39] and forest fires,[40] so that emergency aids and long-term redevelopment efforts can be directed to the correct location.

CONCLUSION

Applications like Google Earth, Google Maps, Bing Maps, and NASA World-Wind have allowed people with little technical knowledge of GIS or RS to use these technologies for everyday purposes like finding the quickest route to the grocery store or discovering a bird's-eye view of cities halfway around the world. For RS to be securely financed, public awareness of this field and the corresponding environmental and public health applications must be enhanced. User-friendly graphic interfaces that allow people to measure and see the average temperature of the Earth's surface or watch clouds that moved over Europe a few hours ago from the perspective of a geostationary weather satellite could educate people about the benefits of this technology and therefore generate political support. The RS community must make greater efforts to show how and why this technology is important.

Public health applications of RS data are no longer new; spatial epidemiology is equally important as the strictly environmental applications for which RS was originally intended. This is not surprising, because environmental studies and epidemiology are inextricably linked. Each provides information on human health

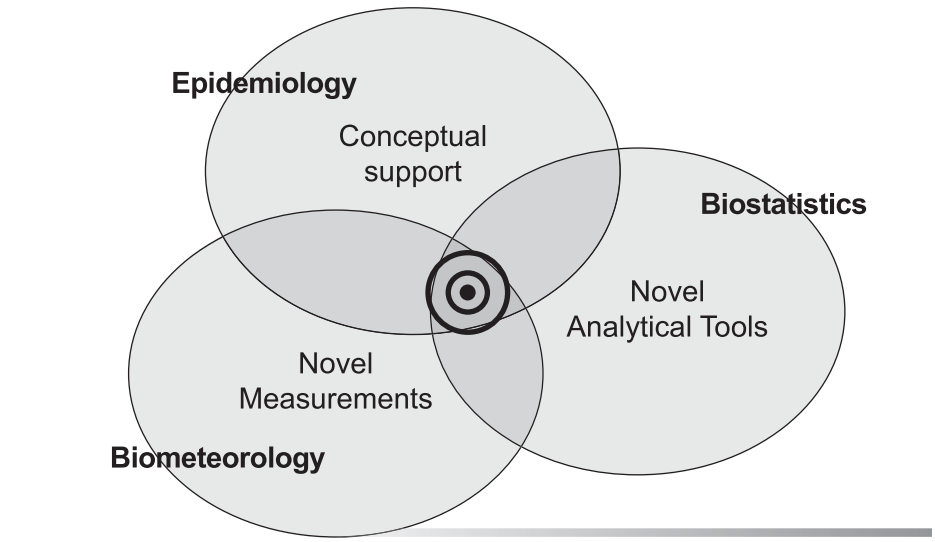

Figure 13.4 Toward Interdisciplinary Understanding.

conditions and the corresponding management of environmental resources. Climate and land use change and variability can be measured remotely and corresponding effects of alterations in natural and built environments can be predicted. Despite the uncertainty involved in forming sound habitat-parasite-host-vector connections, we expect that the promise offered by RS eventually will outweigh current fiscal concerns expressed, for example, by politicians. A globally shared and managed EOS would provide monitoring and predictive capabilities to help people around the world deal with disease, climate variability, and natural disasters. Although RS has many technical challenges, perhaps the most urgent problem is not one of scientific rigor or data availability, but rather one of social and political will to make an international effort toward the greater good.

Examples presented in this chapter indicate that future developments of novel approaches in public health are driven by a strong movement toward interdisciplinary research. New challenges to human, animal, and ecosystem health are demanding novel solutions: new diseases are emerging from new configurations of humans, domestic animals, and wildlife; new pressures on once-robust and resilient ecosystems are compromising their integrity; synthetic compounds and engineered organisms, new to the natural world, are spreading unpredictably around the globe. Globalization also is providing opportunities for infectious organisms to gain access to new hosts, changing in distribution and virulence. Many are calling for interdisciplinary health research to better understand the complex nature of the problems facing the world today (see figure 13.4). The most important requirements of interdisciplinary research include the desire, ability, and capacity to form, test, and utilize a common language that supports the development of novel hypotheses.

A necessary part of interdisciplinary research is asking questions in progressive steps during the process of insight, communication, persuasion, agreement, and decision, while considering the strengths and weaknesses of individual disciplines in terms of what each discipline can contribute to the solution. Novel sources of information and novel ways of knowledge synthesis should be embraced as a means to enhance the health of the population, globally or locally.

REFERENCES

1. Jensen JR. *Remote Sensing of the Environment: An Earth Resource Perspective*. Upper Saddle River, NJ: Prentice Hall, 2002.
2. American Society for Photogrammetry and Remote Sensing. http://www.asprs.org/society/about.html. Published 2007. Accessed April 9, 2007.
3. Myers MF, et al. Forecasting disease risk for increased epidemic preparedness in public health. In: Hay SI, Randolph SE, Rogers DJ, eds. *Remote Sensing and Geographical Information Systems in Epidemiology*. 2nd ed. London, UK: Elsevier Science Ltd; 2000:309–330.
4. Susser M, Susser E. Choosing a future for epidemiology Part 1: eras and paradigms. *Am J Public Health.* 2006;86(5):668–673.
5. Curran PJ, Atkinson PM, Foody GM, Milton EJ. Linking remote sensing, land cover and disease. In: Hay SI, Randolph SE, Rogers DJ, eds. *Remote Sensing and Geographical Information Systems in Epidemiology*. 2nd ed. London, UK: Elsevier Science Ltd; 2000: 37–80.
6. US Environmental Protection Agency. Global earth observation system of systems. http://epa.gov/geoss/. Published 2007. Accessed April 9, 2007.
7. Schmidt CW. Terra cognita: using earth observing systems to understand our world. *Environ Health Perspect.* 2005;113(2):A98–104.
8. Tran A, Gardon J, Weber S, Polidori L. Mapping disease incidence in suburban areas using remotely sensed data. *Am J Epidemiol.* 2002;156(7):662–668.
9. Ford TE, et al. Using satellite images of environmental changes to predict infectious disease outbreaks. *Emerg Infect Dis*. 2009;15(9):1341–1346.
10. Jagai JS, Castronovo DA, Monchak J, Naumova EN. Seasonality of cryptosporidiosis: a meta-analysis approach. *Environ Res.* 2008;109:465–478.
11. Pavlovskii EN. *Natural Nidality of Transmissible Diseases with Special Reference to the Landscape Epidemiology of Zooanthroponoses*. Urbana, IL: University Illinois Press; 1966.
12. Hay SI, Packer MJ, Rogers DJ. The impact of remote sensing on the study and control of invertebrate intermediate hosts and vectors for disease. *Intern J Remote Sensing.* 1997;18(14):2899–2930.
13. Hugh-Jones M. Applications of remote sensing to the identification of the habitats of parasites and disease vectors. *Parasitol Today.* 1989;5(8):244–251.
14. Epstein PR, Chikwenhere GP. Environmental factors in disease surveillance. *Lancet.* 1994;343:1440–1441.
15. Hay SI. An overview of remote sensing and geodesy for epidemiology and public health application. In: Hay SI, Randolph SE, Rogers DJ, eds. *Remote Sensing and Geographical Information Systems in Epidemiology*. 1st ed. London, UK: Elsevier Science Ltd; 2000:1–35.
16. Kitron U, Kazmierczak JJ. Spatial analysis of the distribution of Lyme disease in Wisconsin. *Am J Epidemiol.* 1997;145(6):558–566.

17. Tucker CJ, Gatlin JA, Schneider SR, Kuchinos MA. Monitoring vegetation in the Nile Delta with NOAA-6 and NOAA-7 AVHRR imagery. In: *Remote Sensing of Arid and Semi-Arid Lands*. 1982:973–987.
18. Daniel M, Kolar J, Zeman P. GIS tools for tick and tick-borne disease occurrence. *Parasitol.* 2004;129:S329–S352.
19. Bavia ME, et al. Remote sensing and geographic information systems and risk of American visceral leishmaniasis in Bahia, Brazil. *Parassitol.* 2005; 47(1):165–169.
20. Malone JB, et al. Geographic information systems and the distribution of Schistosoma mansoni in the Nile delta. *Parasitol Today.* 1997;13(3):112–119.
21. Lobitz B, et al. Climate and infectious disease: use of remote sensing for detection of vibrio cholerae by indirect measurement. *Proc Natl Acad Sci U S A.* 2000;97(4):1438–1443.
22. Strong AE, McClain EP. Improved ocean surface temperatures from space: comparisons with drifting buoys. *Bull Am Meteorol Soc.* 1984;65(2):138–142.
23. Hay SI, Lennon JJ. Deriving meteorological variables across Africa for the study and control of vector-borne disease: a comparison of remote sensing and spatial interpolation of climate. *Trop Med Intern Health.* 1999;4(1):58–71.
24. Thomson MC, et al. Potential of environmental models to predict meningitis epidemics in Africa. *Trop Med Intern Health.* 2006;11(6):781–788.
25. Adegbola RA, Demba E, De Veer G, Todd J. Cryptosporidium infection in Gambian children less than 5 years of age. *J Trop Med Hyg.* 1994;97:103–107.
26. Newman RD, et al. Longitudinal study of Cryptosporidium infection in children in northeastern Brazil. *J Infect Dis.* 1999;180:167–175.
27. Perch M, et al. Seven years' experience with Cryptosporidium parvum in Guinea-Bissau, West Africa. *Ann Trop Pediatr.* 2001;21(4):313–318.
28. Naumova EN, et al. Seasonality in six enterically transmitted diseases and ambient temperature. *Epidemiol Infect.* 2007;135(2):281–292.
29. Interagency Working Group on Earth Observations and the National Science and Technology Council Committee on Environmental and Natural Resources. *Strategic Plan for the US Integrated Earth Observation System*. Washington, DC: Executive Office of the President; 2005.
30. Williamson RA, Baker JC. Current US remote sensing policies: opportunities and challenges. *Space Policy.* 2004;20(2):109–116.
31. Kuhn K, et al. *Using Climate to Predict Infectious Disease Epidemics*. Geneva, Switzerland: World Health Organization; 2005.
32. Cromley EK. GIS and disease. *Annu Rev Public Health.* 2003;24:7–24.
33. Graham AJ, Atkinson PM, Danson FM. Spatial analysis for epidemiology. *Acta Trop.* 2004;91(3):219–225.
34. Chuvieco E, et al. Development of a framework for fire risk assessment using remote sensing and geographic information system technologies. *Ecolog Model.* 2009; In Press. doi:10.1016/j.ecolmodel.2008.11.017
35. Chandra S. Application of remote sensing and GIS technology in forest fire risk modeling and management of forest fires: a case study in the Garhwal Himalayan Region. *Geo-inf Disaster Manage.* 2005;1239–1254.
36. Tralli DM, et al. Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *ISPRS J Photogramm Remote Sensing.* 2005;59: 185–198.
37. Frickel S, Campanella R, Vincent MB. Mapping knowledge investments in the aftermath of Hurricane Katrina: a new approach for assessing regulatory agency responses to environmental disaster. *Environ Sci Policy.* 2009;12(2):119–133.

38. Fritz HM, et al. Hurricane Katrina storm surge distribution and field observations on the Mississippi Barrier Islands. *Estuar Coast Shelf Sci.* 2007;74(1–2):12–20.
39. Metternicht G, Hurni L, Gogu R. Remote sensing of landslides: an analysis of the potential contribution to geo-spatial systems for hazard assessment in mountainous environments. *Remote Sensing of Environment.* 2005;98(2/3):284–303.
40. Dlamini WM. Characterization of the July 2007 Swaziland fire disaster using satellite remote sensing and GIS. *Appl Geogr.* 2009;29(3):299–307.

CHAPTER 14

Thinking Creatively about Public Health for the 21st Century

Barry H. Smith, MD, PhD

THE DEVELOPMENT OF PUBLIC HEALTH AS A FIELD

Writing in 1920, one of the statesmen and founders of the field of public health in the United States, C-E. A. Winslow, defined public health as "the science and art of preventing disease, prolonging life and promoting health through the organized community efforts . . . and the development of social machinery which will ensure to every individual in the community a standard of living adequate for the maintenance of health."[1] This definition has proven to be a useful one over the past 90 years, helping to bring together a modern and professional discipline, but public health activities have a history that dates back thousands of years with roots in every Eastern and Western civilization. Whenever and wherever people have come together, there have been issues of preventing disease and maintaining or improving the health of communities, both large and small.

From the beginnings of human civilization, peoples of different cultures and geographies have recognized that polluted water and improper waste disposal were associated with disease. The relation of human behavior to health was also recognized long ago, with the eating of certain foods, the use or abuse of alcohol and other substances, and sexual relations being associated with disease under particular circumstances and often at specific times. Various early religions and cultural traditions, including those of the Han Chinese (Huangdi, the Yellow Emperor, *Huangdi Neijing*); Indians (Buddhism and Hinduism, Sushruta *Samhita*); Hebrews (*Torah*); Greeks (*Corpus Hippocraticum*); Muslims; Christians; and Incas provided and enforced rules designed to protect the health of individuals in these and other areas and likely were among the first organized "public health" efforts. The Romans understood the importance of proper human waste disposal; the Chinese experimented with variolation around 1,000 BCE to try to stop a smallpox epidemic; and Medieval Europe developed burning techniques and quarantine rules to stop the spread of the Black Plague and other diseases based on the then-prevalent miasma theory.

Cholera in pandemic form ravaged Europe from 1829 to 1851, but it was an outbreak in London in 1854 that led Dr. John Snow to develop epidemiology, one of the cornerstones of public health in the 21st century. The long-overlooked observation of microorganisms by Leeuwenhoek (1680) was rightly given preeminence by Pasteur's elaboration of the germ theory of disease in the 1880s. These advances, along with other public health measures, including the building of latrines and sewers, the regular collection of garbage and its incineration or placement in landfills, the provision of clean water, and the removal of standing water to prevent mosquitoes from breeding for diseases such as dengue fever and yellow fever, brought tremendous benefits to the more developed world, with a marked decrease in infectious diseases and an increase in life span. This was much less true for the poorer (developing) areas of the world, a subject to which we will return shortly.

New York City's statistics provide dramatic evidence of just how significant this increase has been in developed, especially urban, areas. Figure 14.1, from the Bureau of Vital Statistics of the New York City Department of Health and Mental Hygiene's Summary of Vital Statistics 2007 shows the deaths per 1,000 population from 1800 through 2007 for the City of New York. Striking is the decrease in the death rate from 1890 to 1920. The conquest of infectious diseases such as yellow fever, cholera, typhoid, and diphtheria, among others, accounts for a large measure of this success.[2] A new decrease in death rates from 1990 through 2007 is apparent, if not quite as striking as the earlier decrease. During this period, life expectancy for a New Yorker born in 2000 is 77.6 (for females, 80.2 years and males, 74.5 years). This overall figure is 5.2 years longer that of a decade earlier and 0.6 years longer than the national average.[3]

During this second period of decline in the death rates, an infectious disease of a new and different sort, HIV/AIDS, was a leading cause of death, but in the 10 years from 1991 to 2001 New York City deaths associated with AIDS dropped by 66 percent. A 70 percent decline in homicide deaths and a 52 percent decrease in deaths of infants less than one year of age were the greatest contributing factors to the decline in the death rate. Notably, two of the three major reasons for the death rate decline, and even the third as well, relate directly to human behavioral patterns, clearly indicative of the changing nature of public health concerns.

A rather different, but nonetheless encouraging, story of increase in life expectancy is that of India—a vast country with a population approaching 1.2 billion. India's overall life expectancy is projected to increase from the current 64.7 years to 75.6 years by 2050, not very different from the current U.S. figure of 77.5 years.[4]. In the state of Kerala, it is already 73 years overall. The Indian infant mortality rate (IMR) has fallen from 59 per 1,000 live births in 2003 to 32 per 1,000 live births in 2008, a dramatic improvement, but still high compared with that of New York City in 2007 (5.4). Improved control of infectious diseases, better nutrition, and rising literacy among women have all been factors, among others, in these encouraging statistics.

The striking progress against many of the then-existing communicable diseases was a great triumph for public health. With that progress came new foci for public health, especially in the developed world, as well as countries such as India and Brazil where development has taken place rapidly. One clear new focus was on chronic diseases, such as heart disease, diabetes, hypertension, stroke, and cancer. Beyond this, however, was the much broader issue of health maintenance and promotion. For example, Dr. Sara Josephine Baker in New York City worked with poor

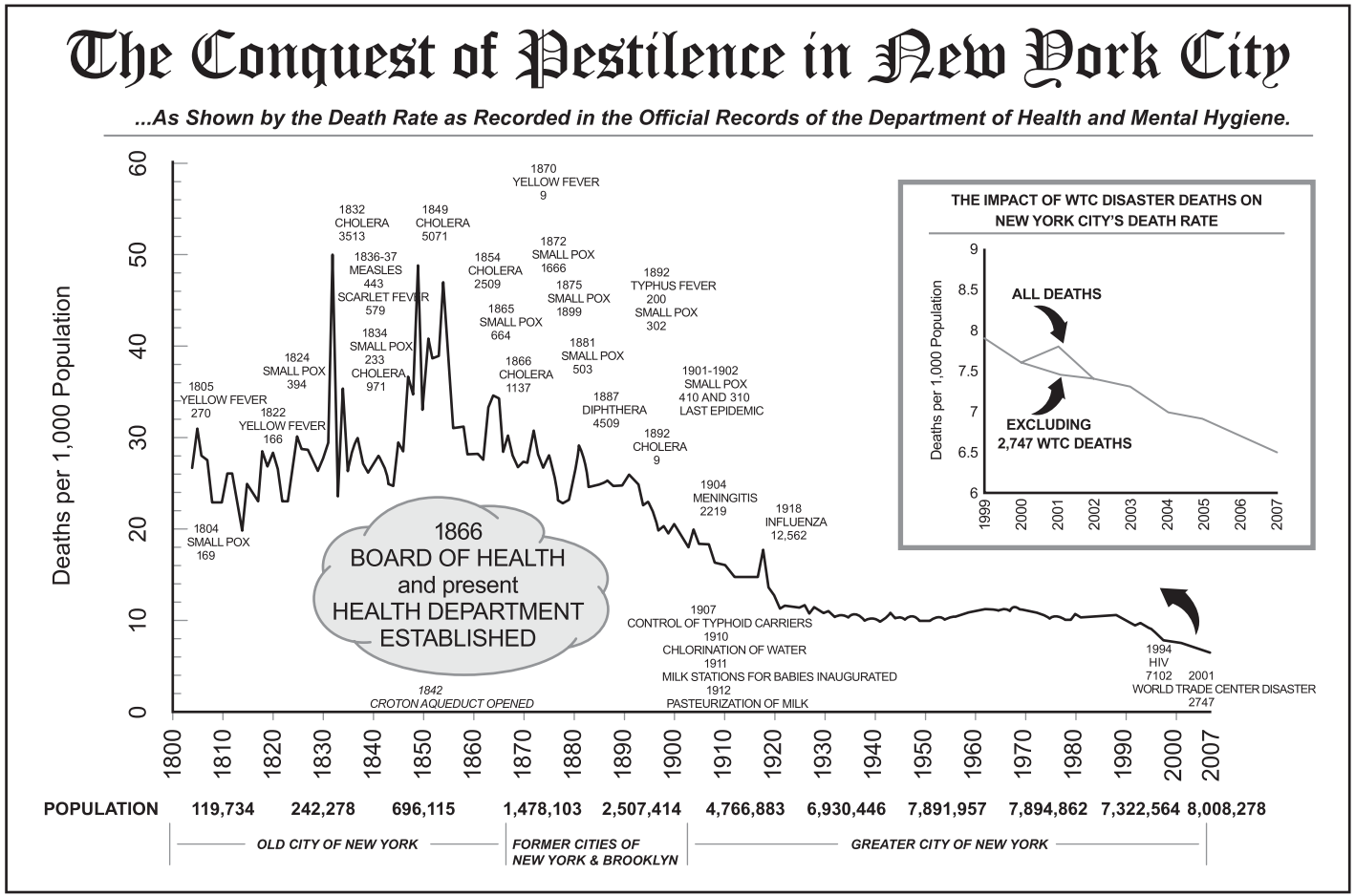

Figure 14.1 The Conquest of Pestilence in New York City.

people and their babies to keep them healthy and to reduce infant mortality rates, and the benefits of her pioneering work are continuing to make a contribution to infant health and mortality. Over the course of the 20th century the control of infectious diseases; vaccination programs for diseases such as polio, smallpox, typhoid, diphtheria; mumps, measles, and rubella; the use of seat belts in cars; regulations and education concerning the importance of use of bicycle helmets; occupational safety regulations and health screening; family planning; fluoridation of drinking water; antismoking campaigns; exercise and healthy diet promotion; and increased screening for chronic disease risks and signs led to a dramatic increase in life span and a much healthier population. With these successes, public health was beginning to mature into a field with a much broader range of interests and questions than it had in the past. The field increasingly focused on population health and its determinants rather than individual health, risk factors, and behavioral change.

Although the positive impact of public health in the 20th century was unquestionable, the end of the century and the beginning of the 21st century saw a host of new challenges. The developed world was not immune to these challenges, while the developing world struggled with many of the old infectious disease and malnutrition problems as well as new problems. Multi-drug-resistant tuberculosis, avian flu, H1N1 swine flu, and the increase of chronic diseases, such as heart disease and diabetes, and are found in both the developed and the developing world. The developing world has to contend with the rapidly rising prevalence of the chronic diseases and the new infectious disease challenges, while still fighting diseases such as polio, cholera, and malaria that have been eradicated in the developed world. All countries also continue to struggle with inequalities within their populations such that, for example, IMR was 10 per 1,000 live births in 2007 in Mississippi,[5] compared with 5.8 per 1,000 live births in New York City.[2] In India in 2006, the under-five mortality rate was 29.7 per 1,000 deaths for mothers with the highest level of education and 94.7 per 1,000 for those with the lowest. The highest and lowest wealth quintiles showed a similar, approximately threefold, difference.[6]

With the advent of the 21st century, it was now clear that health is a global issue with the health of one region dependent on the health of all the others. Because many emerging diseases as well as noncommunicable diseases do not honor nation-state boundaries, the interdependence of all humanity is evident. Consistent with and underscoring this, the new challenges have been increasingly global in nature, whether directly or indirectly. Considering the challenges of emerging infectious diseases such as HIV/AIDS, severe acute respiratory syndrome (SARS), and the H1N1 flu; epidemic type 2 diabetes; childhood and adult obesity; respiratory disease, including asthma secondary to air pollution; adolescent pregnancy; violence; terrorism and bioterrorism; medical care and drug costs; lack of access to health care; natural disasters; and global climate change, the message is clear. The continued growth of the world's population, with some of the least developed areas experiencing the greatest growth rates, makes the problems more difficult because of the enormous numbers of people needing help.

HEALTH: A BROADENING PERSPECTIVE AND NEW CHALLENGES

The public health field has changed along with the changes in global health challenges. A critical change, for example, has been the broadening definition of

health. As the Constitution of the World Health Organization (WHO) puts it, "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity."[7] One important way in which the field has changed its emphasis is to shift from a concern with individual behaviors such as the use of tobacco, to risk factors in populations. The subfield of population health took off in the 1980s and has concerned itself with issues of poverty, inequality, and education. These were spelled out in what have come to be known as the "social determinants of health." According to WHO, "the social determinants of health are the circumstances in which people are born, grow up, live, work and age, and the systems put in place to deal with illness. These circumstances are in turn shaped by a wider set of forces: economics, social policies, and politics."[8] Specific factors that are determinants of health include the following:

1. Income and social status
2. Social support networks
3. Education and literacy
4. Employment working conditions
5. Social environments
6. Physical environments
7. Life skills
8. Personal health practices and coping skills
9. Health childhood development
10. Biology and genetic endowment
11. Health services
12. Gender
13. Culture

A more strictly "social" list drawn up in Canada [9] includes the following:

1. Aboriginal status
2. Early life
3. Education
4. Employment and working conditions
5. Food security
6. Gender
7. Health care services
8. Housing
9. Income and its distribution
10. Social safety net
11. Social exclusion
12. Unemployment and employment security

From the nature of the above lists, it is easy to see how these determinants shape the health inequities that exist, whether within a given country or between countries. Recognizing this, the WHO established the Commission on Social Determinants of Health (CSDH) in 2005 to propose ways to reduce health inequities by working through the social determinants. Although medical care has changed vastly since 1900, McKinlay and McKinlay in 1987 [10] estimated that only 10 to 15 percent of the increase in longevity since that time is due to improved health care, at

least in the developed nations. Similarly, it was not improvements in individual behavior that produced these results. It was much more the improvements in daily life that were responsible. Key elements, consistent with the above lists, included education, early childhood, food processing and availability, health and social services, employment security and conditions—all social determinants of health.

Changes in the public health field have not come easily or without cost. In its 1988 report, "The Future of Public Health," the U.S. Institute of Medicine (IOM) stated, "In recent years there has been a growing sense that public health as a profession, as a governmental activity, and as a commitment of society is neither clearly defined, adequately supported, nor fully understood . . . current capabilities for effective public health action are inadequate."[11] The report defined the core functions of public health as (1) assessment: surveillance of disease and injury (trends, causes, needs); (2) policy development: evidence-based decision making, strategic approaches, and comprehensive public health policies—with broad community involvement; and (3) assurance: implementing legislative mandates and fulfilling statutory responsibilities, providing essential services and access to those services, with subsidies for those needing them.

The challenges facing public health in the 21st century are thus enormous, including, as they do, the original challenges of fighting infectious diseases, but also much more than that, that is, striving to achieve for the people of the world that "complete state of physical, mental, and social well-being and not merely the absence of disease or infirmity" proposed by WHO as the "health" part of public health.[7] Paraphrasing the IOM report referenced above, public health's mission is to create the conditions within which people can be healthy. Alternatively, in Henry Taylor's words, "Public Health shapes the context within which people and communities can be safe and healthy."[12]

TWENTY-FIRST-CENTURY CHALLENGES

The challenges for the field of public health are enormous and difficult to specify and communicate to a public that is accustomed to issues and results that are both defined, and easy to grasp, at least in United States. The issues must be seen as affecting one's personal life and should not require sacrifice on the part of that person or group of people. A new treatment or drug for a particular disease is easy to understand. Prevention of a potential problem is much less obvious. Surveillance is not glamorous and the effects of new health policies may be years, if not decades, away. Unfortunately, the support of the public for the public health field is critical and currently that support is not as strong as it should be. This is a problem that must be addressed.

At the same time, as stated clearly and explicitly in a recent IOM report entitled, *The U.S. Commitment to Global Health: Recommendations for the Public and Private Sectors*, released in May 2009, the United States and the world have a unique and critical opportunity to improve the health of all people in all nations.[13] The knowledge, innovative technologies, and tools are there, but, as the report points out, a wide gap exists between what is possible and what actually is being done in disadvantaged communities and nations. The report calls for action in four areas: (1) generating and sharing knowledge regarding problems endemic to poor

countries; (2) investing in people, institutions, and capacity building in resource-poor settings; (3) increasing U.S. financial investments in global health; and (4) ensuring that the United States is a respectful partner and leader in this effort. The report recognizes that the global health enterprise in the United States involves both many governmental and nongovernmental sectors and disciplines.

Not only must the U.S. government be committed, but also foundations, universities, nonprofit entities, and commercial units must be committed. Of course, much of what is called for is already happening, but the integration of the many disparate efforts to produce the greatest possible impact remains to be achieved. Implicit in the report is the need to educate and engage the public so that the U.S. commitment is deeply rooted and supported by a large segment of the population. The seriousness of the lack of U.S. public understanding of, and support for, public health is emphasized by the IOM report. Without the understanding and support of the public, the full potential of the U.S. global health effort will not be realized. The same is likely to be true in every country around the world.

Another significant issue is the divergence of public health from its molecular, biologically based cousin, mainstream medicine. If medicine has become more and more molecular in its approach and outlook, public health has become more "social" and more holistic in its approach and guiding principles. With its emphasis on populations, it is clearly operating on a broader and quite different stage than that of individual patients with health problems that require interventions from one or more medical specialists. Public health was originally a department within a medical college and still is in many schools in the 21st century. In the United States, approximately 50 independent schools of public health are members of the American Association of Schools of Public Health. Outside the United States, and, in countries such as India, Departments of Community Health have been established. This divergence of the conceptual base of public health from that of contemporary medicine is understandable, but its long-term effects are unlikely to be in the best interests of the health of the people or the further development of the fields of public health or medicine. Ultimately, the two disciplines are interdependent and complementary. Bringing them back together is an important task. Public health should take the initiative in this regard.

Relevant to this point is the fact that the Bill and Melinda Gates Foundation, for example, has made an enormous and broad-based commitment to global health.[14] The Foundation's Global Health program emphasizes the importance of scientific innovation and discovery. It lists three priority areas: (1) the discovery of new insights to fight serious diseases and other problems affecting developing countries; (2) development of effective and affordable vaccines, medicines, and other health tools; and (3) delivery of proven health solutions to those who need them most. Undergirding these efforts, of course, is a strong foundation of technology. The field of public health must shape the ways in which these new solutions are delivered to the people who need them and integrated meaningfully into their lives, providing education that is culturally sensitive and finding ways to deliver such solutions effectively over "the last mile." The people need to take ownership of the solutions, if this effectiveness is to be achieved.

A third obvious problem for public health is the sheer magnitude of the human population and its needs. United Nations projections [15] indicate that the world

population will have grown from 6.1 billion in 2000 to 8.9 billion in 2050—a 47 percent increase or the addition of 57 million people a year on average. This means adding a population the current size of Italy every year between now and 2050. The report goes on to note that the 50-year increase will be more than twice the current population of China and more than double that of the combined current population of all the developed regions of the world. Adding to the challenges is the fact that most of the demographic increase will take place in the less developed regions of the world. These regions will increase their population by 58 percent, while the developed regions increase by only 2 percent, according to the projections.[15] Africa is projected to add 1 billion to its population, with the continent's share of the global population rising from 13 percent to 20 percent. In other words, 99 percent of the population increase will occur in regions where public health systems are the weakest or nonexistent. Beyond that, there is the dramatic shift of people from rural areas to densely packed urban centers with an associated shift in disease patterns and the opportunities for emerging diseases to spread rapidly.

The population problem has another important aspect: the changing global demographics, with aging being the most prominent of these. The growth rate of the over 65 population around the world is expected to be 2.4 percent, with a higher figure (3 percent) seen in developing regions. Put another way, the percentage growth of the population over 65 will increase by 84.8 percent from 2000 to 2050, but by 344 percent in developing regions over this same period. The percentage of the total population over 65 will have risen from 5.2 percent in 1950 to 15.9 percent by 2050.[15] The point is not to belabor the projected numbers, but rather to emphasize the size of the problem. The chronic illnesses of the aging population are far more complex and expensive to manage than communicable diseases. Noncommunicable diseases are projected to account for more than 75 percent of all deaths by 2030 [16] and an increasing percentage of these problems will be seen in the over 65 population. Given the struggles the world already has meeting the needs of the present population burden, it is clear that the difficulties are only going to grow. The resources needed to meet the challenges are unlikely to grow at the pace required to even keep up with the problems. It would appear that many aspects of the disease burden are far outpacing the systems designed to control them.

RESPONDING TO THE CHALLENGES

If the above analysis is correct, then what does the public health field need to do to meet the challenges facing it? How should public health be shaped for the 21st century, both for its own sake as a critical field for the world's well-being and for the sake of the local and global public it serves? The answers to these questions are multilevel and multifaceted. What I will attempt to do here is suggest some of the directions and actions public health as a field must consider if it is to fulfill its potential to contribute to the public good and further develop itself as the exciting, vital field it should be.

First of all, recognizing the numbers presented above, it is clear that there will never be enough public health professionals, medical personnel, and financial resources to meet the challenges. The sheer size and continuing growth of the global human population that is adding 57 million people to the world every year, seems to

be a highly discouraging fact. How can we possibly cope with this? We need to radically change our thinking about such numbers. Yes, they represent a burden to the system, but they also bring us a solution. The people are not the problem. They are a major part of the solution. The "people" whom public health is designed to serve must be integrated into public health as actors and not simply passive recipients or objects of the field's efforts.

How is this integration to be achieved? First of all, traditional thinking about the "people" as passive recipients of the knowledge held by public health experts must be changed. I emphasize this point because biases from within the field are deep and subtle, and ultimately they are limiting and harmful. To emphasize what I mean, I recount a conversation I had in an Asian country with a Ministry of Health official. When asked for her evaluation of the status of public health programs in her country, she responded that the country had excellent health programs. After I replied that that was good to hear, the official replied, "But they don't work." Puzzled, I asked, "How can that be? The programs are excellent, but they don't work?" Her reply was quick. "That's easy. The people don't follow our instructions. They are lazy. They are stupid. You can't trust them." Perhaps that is an extreme example, but it is also an honest one, even if wrong—and it is a widespread view. We need to change our way of thinking.

Amartya Sen in his book, *Development as Freedom*, makes the point that poverty is not really the problem for poor people. The problem is the denial of opportunities for them to exercise their innate abilities to solve problems using local knowledge and skills bred of experience in a particular environment and set of challenges.[17] This does not mean that the "people" have all the knowledge, answers, or tools required to promote their own health or prevent disease, but it does mean that they can be the ones to turn the knowledge into action, belief, and behavioral change, and practice within their communities. The "people" are the ones to integrate the external ("expert") knowledge, priorities, beliefs, and practices into local priorities and realities.

How do we engage the "people" in public health? Crucial to achieving this is to get them to take ownership of the problems and make a commitment to be actively involved in the solution(s). They must see a given issue as a priority for them and not for someone else. One approach is the one utilized by the Dreyfus Health Foundation (DHF). It is Problem Solving for Better Health® (PSBH®).[18,19, 20] It involves asking people to think about the two or three most important (from their point of view) problems in their community or region. These should be problems that they believe they can do something about as individuals or in small groups. Implementing a solution must also fit within what is possible for them to do and it must be a solution that they have created. It may be informed by knowledge that has come from an external source, but it must ultimately belong to the individual who has made a commitment to solving the problem and has detailed the implementation of the solution.

PSBH® is a tool for the individual to use and consists of the scientific method adapted to community-level needs and issues. It includes a series of steps, which begin with (1) defining the problem precisely (nature, size, cause(s), contributing factors); (2) prioritizing it from the point of view of the community and asking if it can, in fact, be addressed by an individual (at least initially); (3) identifying and

sorting out possible solutions, choosing one and asking a "good question";[1] (4) developing an action plan (including background and rationale, good question, hypothesis, methods, and evaluation); and (5) taking action.[19] As a tool, PSBH must be met with a strong commitment to making a positive change, a blend of the intellect and passion. It is about stimulating community organization to achieve change and working to achieve a strong sense that it is realistic to expect it.[18, 20]

PSBH® is a means for helping to organize a community.[21] It has its own intrinsic strength and applicability, and it is intended to sow the seeds of a transformative process in that community. It is about creating change in a self-sustaining, dynamic, and forward-looking approach to continuous community improvement, with a strong emphasis on improving individual and family quality of life. It is not magic and certainly not the only approach to community organization and short- and long-term transformation, but it is a time-tested and proven means of achieving such goals. It serves to bring out what 21st century public health must do if it is to bring the "public" into public health and make the much larger impact it is going to have to make if health is to improve from where it is now and not remain as it is or deteriorate. Put another way, the public must take active ownership of the solutions, thereby ensuring their implementation, effectiveness, and sustainability.

Effective organization of communities at the grassroots level is not simple. It requires hard work and a consistent forward-looking plan that has at its heart local individuals and teams within communities. They must incorporate the transformative process into the very fabric of their community. Public health shares many things in common with political processes and should incorporate lessons from this discipline into its own planning. Examples of effective "community" organization can be found in post-1949 China, with both good and less good results, as well as Barack Obama's recent presidential campaign in the United States and Dr. Thomas Frieden's campaigns for tobacco control and trans fat reductions in New York City. With regard to his presidential campaign, Barack Obama, with a skillful use of the Internet, built a 13-million-strong grassroots network. He indicated that this network would play a critical role as the "Organizing for America" group to achieve the goals for his agenda once in the White House. Public health as a field must build similar grassroots-led networks and find ways to help them become self-sustaining and autoreplicating. The Web site www.whatispublichealth.org and sticker campaign "This Is Public Health," put together by the Association of Schools of Public Health, is one example of what can be done to help build such networks. Just as studies of populations have become important parts of public health, the engagement of populations as active agents for health promotion and prevention is a vital piece that must be added for the 21st century.

Public health professionals also should recognize that they might be part of the problem. Although it is not particularly palatable and some of the terminology bears the stamp of the 1970s, Paulo Freire's *Pedagogy of the Oppressed* makes the point clearly.[22] Freire points out that the oppressed (the "poor") live an existential duality that means that they are at once themselves and at the same time living with an image of themselves that has been provided by their "oppressors." The internalized image

1. The "good question" is as follows: Will doing **What With Whom** (and **Where**) for **How Long** achieve the **Desired Objective** (quantifiable and measurable goal)?

drawn from the oppressors is one of ignorance, lack of capability, and good for nothingness. This leads to fatalistic and self-deprecatory attitudes among poor individuals about their situation. Freire quotes poor individuals who say:

> The peasant begins to get courage to overcome his dependence when he realizes that he is dependent. Until then, he goes along with the boss and says, What can I do? I am only a peasant.[22, p. 61]
>
> The peasant feels inferior to the boss because the boss seems to be the only one who knows things and is able to run things.[22, p. 63]

Freire also adds an admonishment to those who seek to help the oppressed:

> Those who authentically commit themselves to the people must re-examine themselves constantly. The conversion [to a true humanist] is so radical as to not allow of ambiguous behavior. To affirm this commitment but to consider oneself the proprietor of revolutionary wisdom—which must then be given to (or imposed on) the people—is to retain the old ways. The man or woman who proclaims devotion to the cause of liberation yet is unable to enter into communion with the people, whom he or she continues to regard as totally ignorant, is grievously self-deceived. The convert who approaches the people but feels alarm at each step they take, each doubt they express, and each suggestion they offer, and attempts to impose his "status," remains nostalgic towards his origins.[22, pp. 60–61]
>
> At all stages of liberation, the oppressed must see themselves as women and men engaged in the ontological and historical vocation of becoming more fully human . . . action on the side of the oppressed . . . must be action with the oppressed. Those who work for liberation must not take advantage of the emotional dependence of the oppressed. . . . Using their dependence to create still greater dependence is an oppressor tactic.[22, pp. 65–66]

Unfortunately, often with the best of intentions, many of our public health professionals' efforts have been beset and greatly diminished by the sense that they are the "experts" bringing knowledge and enlightenment to those who have none of their own. Education in public health will have to teach its young would-be professionals that the new reality and techniques must be quite different than those of the past.

Problem-solving education of the kind I have described is essential for both public health professionals and individuals.. In such education, as Freire puts it, "people develop their power to perceive critically the way they exist in the world with which and in which they find themselves; they come to see the world not as a static reality, but as a reality in process, in transformation."[22, p. 83] For the public health professional, the conversion is a radical one in the direction of humility and listening. For the poor individual, such education is liberation. Both the professional and the oppressed are beings in the "process of becoming—as unfinished, uncompleted beings in and with a likewise unfinished reality."[22, p. 84] Mutual transformation in space and time is the common theme.

A second issue for public health as a field is its growing separation from mainstream curative or healing medicine. Although medicine has taken a decided turn toward the molecular, genomic, and highly specialized practice, public health has moved in the direction of dealing with populations and the associated societal factors that promote or inhibit the achievement of population health. The two disciplines must find new ways to work together. If this does not happen, the losses to the public benefit will

be much larger than they need to be and both fields will suffer. This is not to say that the two disciplines must do the same work. Instead, they must use the information and concepts each generates from their respective analytical platforms to enhance and make the work of the other more effective. For example, genomic information about populations with respect to metabolism and susceptibility to disease can be combined with the delivery of more appropriate and better-targeted public health information in mass campaigns.

Dr. Victor Herbert, the nutritionist, told me a story of his work in East Africa with folic acid to reduce the prevalence of anemia. To his surprise, he found that while he did indeed reduce anemia with his folate program, he also increased the incidence of malaria, because it turned out, perhaps not surprisingly in retrospect, that folate-poor erythrocytes were less attractive targets for the malarial parasite. The molecular science should have informed the public health campaign to reduce anemia, and the public health results should have helped drive the molecular understanding of real-world malaria. That is but one small example of a principle that requires emphasis in 21st-century public health. HIV/AIDS, hepatitis, diabetes, and cancer incidence and prevalence patterns offer other examples of the value of bringing public health and contemporary evidence-based medical practice and research together.

Exactly how to bring these two disciplines closer together, especially where the public can benefit in significant ways, is an important question. The first reaction to this question is to say that it is already happening, but the evidence does not indicate that this is so. There are at least two approaches to this. In the United States, the first is the utilization of an existing body, such as the IOM, to promote such cross-talk and cross-fertilization. This might be coupled with a programmatic effort by the U.S. Association of Schools of Public Health to promote such interchange. The American Association of Public Health and American Medical Association might provide another locus for interaction. The National Institutes of Health (U.S. Department of Health and Human Services) support for such a program could help put it on a sound, formal footing. A similar effort in the European Union or at the WHO level also could be valuable. At whatever level it takes place, there must be a solid, even passionate, commitment to identifying the areas in which such interactions could make a difference and staying with them until there is a practical product. It will not be easy, but it is crucial.

Perhaps more practical and important for the long run is an increased emphasis on integrating public health issues and thinking into medical and nursing school education. From the other side, more of contemporary medicine's approaches to health could be injected into the education of public health students. Based on our experience with PSBH® around the world, it could be a tool to bring medical and nursing and public health students together and also introduce both to the thinking of the other discipline. There is no reason why every medical and nursing student, however specialized, cannot contribute to the public health aspects of the diseases treated, especially where prevention and health promotion are possible. Practical problem solving can show future doctors and nurses what they can do beyond the usual clinical therapeutics to promote family and public health. The public health professional's awareness of the generally more narrow clinical treatment concerns of medicine means that data collection and policy issues may be much better addressed over time. There is everything to gain and nothing to lose.

Yet another challenge for 21st century public health is an enhanced ability to deal with the complex interactions among various sectors of society that end up affecting health. Health is a product of society. A well-functioning community or society will produce better health than one that is marginally functional or fully dysfunctional. Public health and medicine, for understandable and practical reasons, have more often than not focused on particular habits, diseases, or conditions, such as tobacco, lipid profiles, or currently, the obesity epidemic and type 2 diabetes, and their accompanying knowledge, attitudes, beliefs, and practices in the public.

Twenty-first century public health should begin to look more rigorously at the multiple factors in a society that predict health outcomes. These factors include economics, housing, nutrition, sports and recreation, education, spirituality, family structure, gender relations, childcare, transportation, and whatever other factors make up a whole, integrated human life. A chart of the multiple influences that ultimately define the level of health and quality of life for one community in the Mississippi Delta,[2] a DHF PSBH® program site, makes these multiple influences and resources clear (see figure 14.2). The fact is that people cannot be described by a disease label, such as diabetes, HIV/AIDS, tuberculosis, peripheral vascular disease, or whatever else comes from the diagnoses medicine provides. The total human being is much

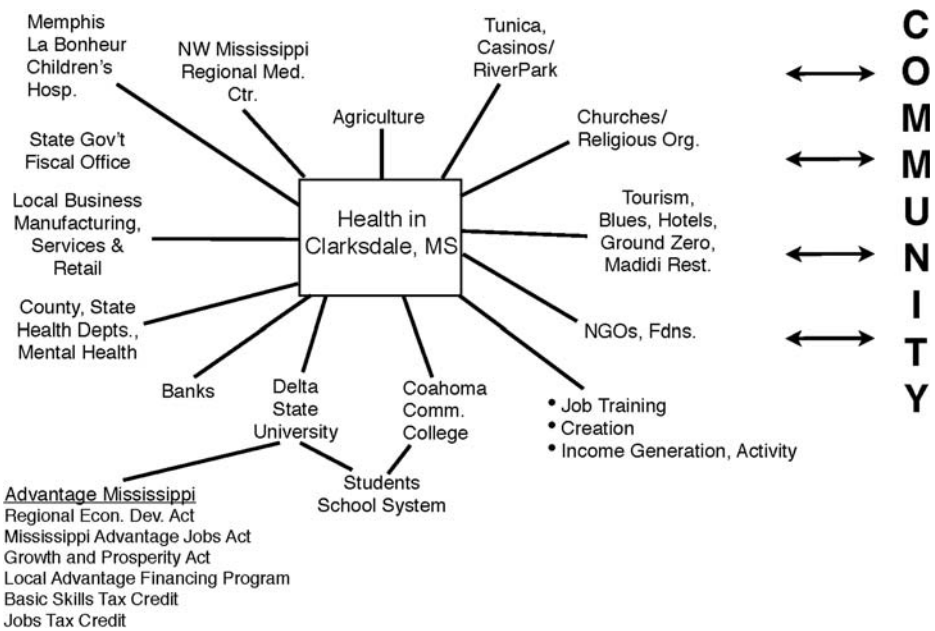

Figure 14.2 Multiple Influences that Define Level of Health.

2. The Mississippi Delta, which is the geographic region extending from just below Memphis, Tennessee, to Vicksburg, Mississippi, is one of the poorest regions in the United States, with Mississippi itself ranking 50th in both economic and health terms of all the U.S. states.[23] It is also a place of great people and great promise.

more than one or more medical diagnoses or behavioral habits and practices. A more effective public health discipline can bring the multiple factors contributing to people's total well-being, health, and quality of life together to predict outcomes both qualitatively and quantitatively. As a result, it can plan interventions that are more likely to be successful in promoting health and preventing disease. It also will draw on multiple resources, many of which do not seem to have a direct relationship to health, to achieve the health goals that have been set. Achieving this new analytical and predictive power will require new mathematical approaches, perhaps with elements taken from chaos theory [24] and the sophisticated navigational systems involving satellites and computers for data collection and analysis program that already have been developed.[25] Developing and embracing these techniques must be a priority for public health. Without such techniques, public health cannot fulfill its responsibilities for assessment, policy development, and assurance.

Public health of the 21st century will be an increasingly important part of the approach to ensuring better health and better quality of life for all people. The challenges are tremendous, but so are the opportunities. To achieve its full potential, public health must bring the "public" into the process. Schools of medicine, dentistry, nursing, and public health need to integrate their curricula better and reach out to schools in other sectors to find new and better ways to achieve better health. All concerned with the improvement of health on both a local and national and global scale need to work collectively rather than in isolation from one another. Health, after all, is a product of the multiple facets of society, and, as such, requires a multifaceted approach to health promotion, the prevention and treatment of disease, and, most important of all, the improvement of the quality of life for all people.

REFERENCES

1. Winslow C-EA. The untilled fields of public health. *Science.* 1920; 51(1306):23–33.
2. New York City Department of Health and Mental Hygiene. Summary of vital statistics 2007: the City of New York. http://www.nyc.gov/html/doh/html/vs/vs.shtml. Published 2008.
3. New York City Department of Health and Mental Hygiene. Summary of vital statistics 2007: the City of New York. http://www.nyc.gov/html/doh/html/vs/vs.shtml.
4. World Health Organization. World health statistics 2008. http://www.who.int/countries/ind/en.
5. Mississippi State Department of Health, Office of Health Data and Research. 2008 Mississippi infant mortality report. http://www.msdh.state.us.ms/msdhsite/_static/resources/3109.pdf. Published 2008.
6. World Health Organization. World health statistics 2008. http://apps.who.int/whosis/database/core/core__select__process.
7. World Health Organization. Constitution of the World Health Organization. 45th ed. http://www.who.int/governance/eb.constitution/en. Published 2006.
8. World Health Organization. Social determinants of health. http://www.who.int/social_determinants/en/. Published 2008.
9. Raphael D, ed. *Social Determinants of Health: Canadian Perspectives*. 2nd ed. Toronto, ON: Canadian Scholars' Press; 2008.
10. McKinlay J, McKinlay SM. Medical measures and the decline of mortality. In: Schwartz HD, ed. *Dominant Issues in Medical Sociology*. New York: Random House; 1987.

11. Institute of Medicine, Committee for the Study of the Future of Public Health, Division of Health Care Services. *The Future of Public Health*. Washington, DC: Institute of Medicine; 2002.
Institute of Medicine. *The Future of Public Health in the 21st Century*. 2002. Washington, DC: Institute of Medicine; 2002.
12. Taylor H. Introduction to public health practice. http://jhsph.edu/courses/publichealthpractice101/PDF/Session4.pdf. Published 2008.
13. Institute of Medicine. *The US Commitment to Global Health: Recommendations for the Public and Private Sectors*. Washington, DC: Institute of Medicine; 2009.
14. Bill and Melinda Gates Foundation. http://www.gatesfoundation.org/globalhealth.
15. United Nations Department of Economic and Social Affairs, Population Division. World population to 2300. http://www.un.org/esa/populations/longrange2//WorldPop2300final.pdf. Published 2004.
16. World Health Organization. The world health report 2008. http://www.who.int/entity/whr/current//en/. Published 2008.
17. Sen A. *Development as Freedom*. New York: Anchor Books, Random House; 1999.
18. Smith BH, Barnett S, Collado D, et al. Optimizing the use of available resources: problem solving for better health. *World Health Forum.* 1994;15:9–15.
19. Dreyfus Health Foundation. Problem solving for better health handbook. http://www.dhfglobal.org. Published 2009.
20. Hoyt P. An international for approach to Problem Solving Better Health Nursing.™ (PSBHN). *Int Nurs Rev.* 2207; 54:100–106.
21. Hoyt P, Fitzgerald J, Smith B, eds. *Global Experience with Problem Solving for Better Health®*. New York: Springer New York; 2009.
22. Freire P. *Pedagogy of the Oppressed*. 30th anniversary ed. New York: Continuum; 2000.
23. StateMaster.com. http:/www.statemaster.com/state/MS-mississippi/. Published 2009.
24. Resnicow K, Page SE. Embracing chaos and complexity: a quantum change for public health. *Am J Pub Health.* 2008;98(8):1382–1389.
25. Scotch M, Parmanto B, Gadd CS, Sharma RK. Exploring the role of GIS during community health assessment problem solving: experiences of public health professionals. *Int J Health Geogr.* 2006;5:39–49.

This page intentionally left blank

About the Editor

MADELON L. FINKEL, PhD, is professor of clinical public health and director of the Office of Global Health Education at the Weill Cornell Medical College in New York City. She also is course director in the Department of Public Health, coordinating the epidemiology/biostatistics, evidence-based medicine, and health care systems courses as well as the public health clerkship. Dr. Finkel's interests focus on women's health issues. Research projects include a cervical cancer screening program in rural India, reduction of preterm birth in Brazil, and better management of hypertension in Peru. She has served as consultant to numerous organizations, including law firms, pharmaceutical companies, and health care organizations. She serves on several boards of organizations whose missions are to promote public health. Her book *Understanding the Mammography Controversy: Science, Politics, and Breast Cancer Screening* was published by Praeger (2005). Praeger also published her book *Truth, Lies, and Public Health: How We Are Affected When Science and Politics Collide* (2007).

This page intentionally left blank

About the Contributors

Erika Abramson, MD, is a pediatrician and health services researcher at Weill Cornell Medical College in New York City. Her research focus is on the effect of health information technology on medication safety. Her work has examined such questions as whether the transition between different electronic prescribing systems affects medication safety.

Jessica S. Ancker, MPH, PhD, is an informatics researcher at Weill Cornell Medical College in New York City who studies the impact of health information technology on patients, providers, and public health. She is a member of the Health Information Technology Evaluation Consortium (HITEC), which is an academic collaborative designated by the state of New York to perform evaluation research on New York's state-funded health IT initiatives.

Holly G. Atkinson, MD, is assistant professor of medicine and co-director of the Advancing Idealism in Medicine (AIM) Program at Mt. Sinai School of Medicine and adjunct assistant professor of public health at Weill Medical College of Cornell University in New York City. She is chief medical officer for HealthiNation, and board member and past president of Physicians for Human Rights, an organization that shared in the 1997 Nobel Peace Prize. Dr. Atkinson focuses her work on the intersection of health and human rights, especially as it applies to advancing the well-being of women and girls.

Heejung Bang, PhD, is an associate professor, Division of Biostatistics and Epidemiology in the Department of Public Health at Weill Medical College of Cornell University, New York City.

Carla Boutin-Foster, MD, MS, is an associate professor of medicine and associate professor of public health at the Weill Cornell Medical College in New York City. She is the Nanette Laitman Clinical Scholar in Public Health and Community Health as well as the co-director of the Comprehensive Center of Excellence in Disparities Research and Community Engagement at Weill Cornell Medical College and associate director of the Center for Multicultural and Minority Health at Weill Cornell. Dr. Boutin-Foster's research focuses on the psychological and social determinants of health outcomes and on the social epidemiology of health disparities.

Ross C. Brownson, PhD, is professor of epidemiology at Washington University in St. Louis and co-director of the Centers for Disease Control and Prevention–-funded Prevention Research Center in St. Louis whose focus is to develop innovative approaches to chronic disease prevention. He is associate editor of the *Annual Review of Public Health*.

Denise Castronovo, MS, is the president of Mapping Sustainability, LLC, a scientific and technical consulting firm for spatial data analysis and a research partner with Tufts Initiative for the Forecasting and Modeling of Infectious Diseases (InForMID). She is interested in the development of novel geospatial methodologies and visualization for infectious disease surveillance, including dynamic mapping and remote sensing analysis.

Ryan Cauley, MD, is a Weill Cornell Medical College graduate. He was elected to the Alpha Omega Alpha Medical Honor Society. He is a first year resident in general surgery at the Brigham and Women's Hospital of Harvard University in Boston, Massachusetts. His areas of research focus on the intersection of surgery and public health and international medicine.

Kenneth K. H. Chui, MS, MPH, PhD, investigates the seasonal patterns of hospitalizations related to various infectious gastrointestinal diseases, and their possible association with extreme weather events. His research interests include applications of time series analyses and geographic information system in nutrition and public health and development of data visualization for efficient communication.

Karen Scott Collins, MD, MPH, is vice president, quality and patient safety at the New York-Presbyterian Hospital in New York City.

Joseph T. Cooke, MD, is an associate professor of clinical medicine and public health at the Weill Cornell Medical Center and chairman of the General Faculty Council. Dr. Cooke is board certified in internal medicine, pulmonary disease, and critical care medicine, and a fellow of the American College of Chest Physicians. He serves as the chief quality and patient safety officer for the New York Presbyterian Hospital Weill Cornell Medical Center. He is the co-chairman of New York Presbyterian Hospital Quality and Patient Safety Committee.

Anthony Dawson, RN, is the vice president for operations, New York-Presbyterian/Milstein Hospital. He supports Milstein's senior vice president and chief operating officer in managing the capital budget and plays a leadership role in implementing the hospital's strategic vision for clinical services.

Inmaculada de Melo-Martín, PhD, MS, is associate professor of public health in the Division of Medical Ethics in the Department of Public Health at the Weill Cornell Medical College in New York City. She holds a doctorate in philosophy and a master's of science degree in molecular biology. Her research interests include bioethics and philosophy of science.

Jonathan E. Fielding, MD, MPH, MA, MBA, is director of the Los Angeles County Department of Public Health and the County Health Office. He is a professor in the Schools of Medicine and Public Health at UCLA, and chairs the U.S. Community Preventive Services Task Force and the Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives for 2020.

Madelon L. Finkel, PhD, is professor of clinical public health and director of the Office of Global Health Education at the Weill Cornell Medical College in New York City. Her research focuses on women's health issues and global public health.?

Paul Glare, MD, is chief of pain and palliative care in the Department of Medicine at Memorial Sloan Kettering Cancer Center in New York City. He is trained in internal medicine, palliative medicine, and pain management.

Phillip L. Graham III, MD, MPH, is an assistant professor of pediatric infectious diseases at Columbia University School of Medicine, an adjunct assistant professor of pediatrics at Weill Cornell Medical College, and the pediatric hospital epidemiologist and quality and patient safety officer for women and children's health at New York-Presbyterian Hospital in New York City.

Robert A. Green, MD, MPH, is assistant professor of clinical medicine and associate director of emergency services at New York-Presbyterian/Columbia University Medical Center in New York City. He also is chief patient safety officer of the Columbia University faculty practice organization and the New York-Presbyterian Hospital/Columbia Campus.

Dan Hyman, MD, MMM, a board certified pediatrician, is chief quality officer at New York-Presbyterian Hospital in New York City. He has been the chief medical officer for the Hospital's Ambulatory Care Network and chief children's quality officer for the Morgan Stanley Children's Hospital and the Komansky Center for Children's Health.

Jyotsna S. Jagai, **MS, MPH, PhD,** is currently pursuing postdoctoral training at the U.S. Environmental Protection Agency. Her area of interest is characterizing the impacts of climate and environment on human health.

Steven Kaplan, MD, a Board Certified Emergency Physician, is currently the Chief Medical Director and Quality and Patient Safety Officer for Ambulatory Care at New York-Presbyterian Hospital in New York City.

Rainu Kaushal, MD, MPH, is an associate professor of public health and pediatrics at the Weill Cornell Medical College and is chief, Division of Quality and Medical Informatics in the Department of Public Health. She is also the director of Health Information Technology Evaluation Consortium (HITEC), an academic collaborative designated by the state of New York to evaluate its health IT initiatives in communities around the state. Her research focuses on evaluating the value and impact of health information technology, particularly interoperable health IT.

Lisa M. Kern, MD, MPH, is an assistant professor of public health in the Division of Quality and Medical Informatics at the Weill Cornell Medical College. She is a health services researcher and internist whose research focuses on measuring the value of health information technology in both clinical and economic terms.

Eliot J. Lazar, MD, MBA, is vice president and chief quality officer at the New York Presbyterian Hospital in New York City. He is a board certified cardiologist.

Rika Maeshiro, MD, MPH, is the director of Public Health and Prevention Projects at the Association of American Medical Colleges (AAMC) where her responsibilities include serving as the deputy principal investigator for AAMC's cooperative agreement with the Centers for Disease Control and Prevention and the principal investigator for the Regional Medicine-Public Health Education Center initiative, an effort to improve population health, public health, and prevention content in medical education.

Christopher M. Maylahn, MPH, is an epidemiologist in the Office of Public Health Practice of the New York State Department of Health. His focus of interest is chronic disease prevention.

Jesse C. McEntee, MA, is a doctoral candidate at Cardiff University's Center for Business Relationships, Accountability, Sustainability, and Society. He is president of McEntee Research and Consulting, a firm focused on the development of public health applications of remote sensing and geographic information system data.

Elena N. Naumova, PhD, is a professor in the Department of Public Health and Community Medicine, Tufts University School of Medicine and director of the Initiative for Forecasting and Modeling Infectious Diseases (InForMID). Her expertise is in the development of analytical tools for time series and longitudinal data analysis for disease surveillance, exposure assessment, and environmental epidemiology.

Vaishali Patel, PhD, MPH, is a health services researcher in the Department of Public Health at the Weill Cornell Medical College in New York City whose current research focuses on a number of different community-based interoperable health information technology initiatives being implemented across New York State, including personal health records, virtual health records, and electronic prescribing applications.

Brian K. Regan, PhD, is director of clinical affairs for the New York-Presbyterian Healthcare System and director of quality and patient safety, external alliances, for New York-Presbyterian Hospital located in New York City. Dr. Regan is a leader for system development and management, including the promulgation and implementation of member performance standards for clinical operations, the facilitation of linkages for patient safety improvements, and administrative and operational efficiencies.

Barry H. Smith, MD, PhD, is clinical professor of surgery and director of The Rogosin Institute at Weill Cornell Medical College in New York City. He also is the director of the Dreyfus Health Foundation, a Division of the Rogosin Institute, where he has worked to further the enhancement of basic and clinical science health care among health professionals in more than 30 countries around the world.

Roma Tickoo, MD, MPH, is an assistant attending in the Palliative Care Service in the Department of Internal Medicine at Memorial Sloan Kettering Cancer Center in New York City. She is trained in internal medicine, geriatric medicine, and pain management.

Paula Trushin is a medical editor and writer who has worked with researchers and physician educators on projects covering a wide range of therapeutic areas, including biostatistics, global health, and medical ethics.

William D. White, PhD, is director of the Sloan Program in Health Administration and professor in the Department of Policy Analysis and Management at Cornell University.

Index