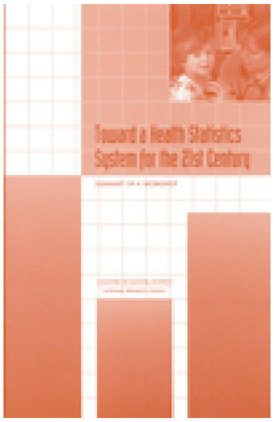

Toward a Health Statistics System for the 21st Century: Summary of a Workshop

Committee on National Statistics, Edward B. Perrin, William D. Kalsbeek, and Terri M. Scanlan, Editors, National Research Council

ISBN: 0-309-51097-X, 94 pages, 6 x 9, (2001)

This free PDF was downloaded from:
http://www.nap.edu/catalog/10171.html

Visit the National Academies Press online, the authoritative source for all books from the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council:

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Purchase printed books and PDF files
- Explore our innovative research tools – try the Research Dashboard now
- Sign up to be notified when new books are published

Thank you for downloading this free PDF. If you have comments, questions or want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, visit us online, or send an email to comments@nap.edu.

This book plus thousands more are available at www.nap.edu.

Copyright © National Academy of Sciences. All rights reserved.
Unless otherwise indicated, all materials in this PDF file are copyrighted by the National Academy of Sciences. Distribution or copying is strictly prohibited without permission of the National Academies Press <http://www.nap.edu/permissions/>. Permission is granted for this material to be posted on a secure password-protected Web site. The content may not be posted on a public Web site.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Toward a Health Statistics System for the 21st Century

SUMMARY OF A WORKSHOP

Committee on National Statistics

Edward B. Perrin, William D. Kalsbeek, and Terri M. Scanlan,
Editors

Division of Behavioral and Social Sciences and Education

National Research Council

NATIONAL ACADEMY PRESS
Washington, DC

NATIONAL ACADEMY PRESS • 2101 Constitution Avenue, N.W. • Washington, DC 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

The project that is the subject of this report was supported by funds provided by the Office of the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Human Services. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

Support of the work of the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (No. SBR-9709489) between the National Academy of Sciences and the National Science Foundation.

International Standard Book Number 0-309-07582-3

Additional copies of this report are available from National Academy Press, 2101 Constitution Avenue, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu

Printed in the United States of America

Suggested Citation: National Research Council (2001) *Toward a Health Statistics System for the 21st Century: Summary of a Workshop*. Committee on National Statistics, Edward B. Perrin, William D. Kalsbeek, and Terri M. Scanlan, Editors. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

THE NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

COMMITTEE ON NATIONAL STATISTICS
1999-2000

JOHN E. ROLPH (*Chair*), Marshall School of Business, University of Southern California
JOSEPH G. ALTONJI, Department of Economics, Northwestern University
LAWRENCE D. BROWN, Department of Statistics, University of Pennsylvania
JULIE DAVANZO, RAND, Santa Monica, California
WILLIAM F. EDDY, Department of Statistics, Carnegie Mellon University
HERMANN HABERMANN, Statistics Division, United Nations, New York
WILLIAM D. KALSBEEK, Survey Research Unit, Department of Biostatistics, University of North Carolina
RODERICK J.A. LITTLE, School of Public Health, University of Michigan
THOMAS A. LOUIS, RAND, Arlington, Virginia
CHARLES F. MANSKI, Department of Economics, Northwestern University
EDWARD B. PERRIN, Department of Health Services, University of Washington
FRANCISCO J. SAMANIEGO, Division of Statistics, University of California, Davis
RICHARD L. SCHMALENSEE, Sloan School of Management, Massachusetts Institute of Technology
MATTHEW D. SHAPIRO, Department of Economics, University of Michigan

ANDREW A. WHITE, *Director*

Acknowledgments

The Committee on National Statistics (CNSTAT) appreciates the time, effort, and valuable input of the many people who contributed to the workshop and to the preparation of this report. We would first like to thank those who made presentations identifying many of the key issues in this area. We would also like to thank all those who attended the workshop and participated in the discussions during which many important issues and concerns were raised and are captured in this summary report. Thanks are due especially to Edward Perrin, former CNSTAT member, who as workshop chair provided valuable advice during the planning stages and the leadership necessary for conducting a successful workshop. We are also grateful to Dr. Perrin for his immeasurable help in the writing of this summary report. We would also like to thank William Kalsbeek, CNSTAT member, for his assistance in the planning of the workshop and for his significant contributions to this report.

The agenda for the workshop was developed in consultation with Edward Sondik, Director of the National Center for Health Statistics, whose input was essential in identifying workshop objectives. Appreciation is also extended to the members and staff of the U.S. National Committee on Vital and Health Statistics (NCVHS) and the Department of Health and Human Services (DHHS) Data Council. We also thank the Office of the Assistant Secretary for Planning and Evaluation, DHHS, the funder of the workshop.

Particular appreciation is due to those who worked to organize the workshop and prepare this report. Jane Durch served as study director for the workshop until October 1999, and was instrumental in planning the workshop and developing the agenda. Miron Straf, former CNSTAT director, was responsible for early project development and workshop planning. Andrew White, also involved in early workshop development, oversaw the planning and execution of the workshop and the preparation of this report. Jonathan Zaff prepared the initial report draft. Terri Scanlan assisted in organizing the workshop, contributed to early report drafts, and responded to reviewers' comments. Agnes Gaskin was responsible for all of the details involved in organizing the workshop and preparing the report for final publication. Lorraine Ferrier edited the draft report. Eugenia Grohman, associate director for reports in the Division of Behavioral and Social Sciences and Education, guided the report through the review process, final editing, and publication.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the Report Review Committee of the National Research Council. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their participation in the review of this report: David Cutler, Center for Advanced Study in the Behavioral Sciences, Stanford, California; Richard A. Kulka, Research Triangle Institute, Research Triangle Park, North Carolina; Thomas A. Louis, RAND, Arlington, Virginia; Denise Love, National Association of Health Data Organizations, Salt Lake City, Utah; and Alan M. Zaslavsky, Department of Health Care Policy, Harvard Medical School.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by Barbara McNeil, Department of Health Care Policy, Harvard Medical School. Appointed by the National Research Council, she was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring panel and the institution.

John E. Rolph, *Chair*
Committee on National Statistics

Contents

1

Introduction

For decades, our nation has collected statistics that provide information on the health status and health care utilization of our population. This information is gathered from a variety of sources: national surveys, administrative records, vital statistics, and state- and local-level data sources. Several public agencies are responsible for the collection, compilation, analyses, and reporting of these statistics. The statistics are used by numerous different people and groups—policy makers in Congress and state legislatures, public health and health care researchers, health care providers, employers, insurers and other private sector stakeholders, consumers, and many others in the health community—for many different reasons. The data, the data collection methods, the data providers, and the data users are the components that combine to form what is considered to be our national health statistics system.

Recently, many new developments have brought to the forefront questions about the adequacy of our present health statistics system. For example, the growth of managed care, new information technologies, and changes in welfare policy will open up opportunities in administrative data; advances in genetics could develop new major health status data items and raise issues of privacy and confidentiality; the growing emphasis on state and local roles in health care creates a demand for greater geographic detail in data; efforts to develop and assess health promotion and disease prevention efforts will require population-based data; the aging of the population will draw greater attention to measures of functional status and disability;

and the boom in information technology will provide new opportunities for collecting and disseminating data. These and other imminent developments in medicine, health care, and public health are likely to influence interest in and need for health statistics as well as the ability of the health statistics system to produce useful information products in a timely fashion. The obvious question arises: What steps need to be taken to ensure that our nation's health statistics system will continue to meet the demands for health data in the twenty-first century?

It is this question that the U.S. Department of Health and Human Services (DHHS) posed to itself and its agencies, many of which devote resources to health statistics and are the source for the majority of our national health data. The National Center for Health Statistics (NCHS), part of the Centers for Disease Control and Prevention (CDC), is the principal vital and health statistics agency of the federal government. NCHS provides data on vital events as well as information on health status, lifestyle and exposure to unhealthy influences, the onset and diagnosis of illness and disability, and the use of health care. The Agency for Healthcare Research and Quality (AHRQ) supports research designed to improve the quality of health care, to reduce its cost, to improve patient safety, to address medical errors, and to broaden access to essential services. AHRQ sponsors and conducts research that provides evidence-based information on health care outcomes; quality; and cost, use, and access. The Substance Abuse and Mental Health Services Administration (SAMHSA) works to improve the quality and availability of substance abuse prevention, addiction treatment, and mental health services. SAMHSA provides national data on such health issues as drug abuse and drug treatment effectiveness. The National Institutes of Health (NIH) conducts extensive biomedical research and provides data on diseases and treatments. The Health Care Finance Agency (HCFA) administers Medicare, Medicaid, and the State Children's Health Insurance Program (SCHIP); and performs a number of quality-focused activities, such as regulation of laboratory testing and certification of nursing homes and other continuing care providers. HCFA provides state and national data on national health care indicators and expenditures, such as health care spending; employment and prices; and Medicare, Medicaid, and SCHIP enrollment. The Health Resources Services Administration (HRSA) directs national health programs that improve the nation's health by assuring equitable access to comprehensive, quality health care. HRSA provides data on minority health issues. The CDC is responsible for protecting the health of the American public through the monitoring of disease trends, investigation of outbreaks, assessment of health and injury risks, and implementation of illness and injury control and prevention interventions. When assessing the status of our nation's health statistics system, if changes or improvements need to be made, all the agencies that are responsible for the collection of health data must be involved in order to implement the changes effectively.

To help identify steps toward developing a new, more robust, relevant, timely, and secure health statistics system, DHHS adopted plans to develop a "21st century vision for health statistics." The vision process began in 1999 as a partnership between the National Committee on Vital and Health Statistics (NCVHS),[1] the National Center for Health Statistics (NCHS), and the DHHS Data Council.[2] Steps in the process involve a systematic effort to obtain input from experts from various related fields, state and local officials, data users, and other interested parties, beginning with a dialog on future health and health care trends, information and communications technology, and public policies. By its end, this process will identify forces that will shape health information needs and opportunities for the future, and will formulate a vision for the future that will help guide policy and planning for health statistics programs.

As part of this initiative, and at the request of NCHS, the Committee on National Statistics (CNSTAT) convened a workshop on November 4-5, 1999, to identify new directions for health statistics and the implications for health data of changes in the health arena faced by DHHS; state and local health departments; the consumers, developers, and providers of health care products and services; and other health policy makers. Changes in our understanding of health, in health care (managed care, Medicaid, Medicare), in welfare reform, in federal-state relations, in the availability of

---

[1] NCVHS serves as the statutory public advisory body to the Secretary of Health and Human Services in the area of health data and statistics. In that capacity, NCVHS provides advice and assistance to DHHS and serves as a forum for interaction with interested private sector groups on a variety of key health data issues.

[2] The DHHS Data Council coordinates all health and non-health data collection and analysis activities of DHHS, including an integrated health data collection strategy, coordination of health data standards, and health information and privacy policy activities. The Data Council consists of assistant secretary and agency administrator level DHHS officials who have a direct reporting relationship to the Secretary, the HHS Privacy Advocate, and the Secretary's Senior Advisor on Health Statistics. The Data Council serves as the DHHS liaison for the NCVHS.

administrative data, in advanced genetic data, in information technology, in confidentiality issues, and in data integration are examples of recent developments that may play a significant role for DHHS in making future policy decisions. Funding for the workshop was provided by the Office of the Assistant Secretary for Planning and Evaluation, DHHS.

The goals of the workshop were to:

1. provide an overview and historical perspective on the elements of the nation's health statistics system;
2. identify key assumptions and critical components of a health information system, in part by using examples of health statistics systems in other countries as models;
3. review the quantity, quality, and availability of the data currently provided the health statistics system;
4. identify current and future data gaps, particularly with respect to the health and policy questions the future health statistics system should be able to answer; and
5. discuss ways in which the current health data system must evolve to meet future information needs.

Before the health statistics system as a whole can be examined, particularly with respect to the goals of the workshop, the question of "what is health statistics" should be addressed. The term "health statistics" connotes different things to different people, from public health surveillance to vital records, and the definition and scope may change depending on the research or policy question that those statistics are being used to answer. Although the scope of this workshop did not directly encompass establishing a standard definition of health statistics, the presentations and discussions did fall into categories that are within the boundaries of what is commonly considered health statistics. The categories that emerged during the workshop discussions included:

- Health status: prevalence and burden of disease and epidemiological concerns
- Health services: service delivery, treatments, quality of care
- Health care financing: costs, access, insurance
- Outcomes of care and efficacy (using observational data)

The above list is in no way intended to be comprehensive or exclusionary, but more to set the context in which the term "health statistics" is used throughout the workshop discussions summarized in this report.

Participants at the workshop were diverse in background and perspective, representing both data providers and data users. There were representatives from the federal, state, and local agencies, as well as from the private sector, the research community, and academe. To offer examples as models of health statistics systems from other countries, there were several international health experts in attendance. To ensure that all the issues concerning health statistics were addressed, participants represented expertise in such areas as health policy; managed care; outcome measurement; children's health needs; population-based assessments of health status; surveillance of illness; assessing well-being; risk assessment; determinants of health (e.g., genetics, behavior, socioeconomic status, environment); organization, delivery, and financing of health services; health economics; technology, data policy, privacy, and confidentiality; and survey methods.

The meeting began with a keynote presentation and four invited background papers and concluded with four panel discussions. Following each session, time was reserved for discussion among all workshop participants. The keynote address, commissioned by CNSTAT and entitled "Health Statistics: Past, Present, and Future," is reproduced in its entirety as Chapter 2 of this report. The four background papers were commissioned separately by the NCVHS (see Appendix). The list of workshop participants and the agenda for the workshop are also provided in the Appendix.

Participants were asked to present their ideas of how changes in the health arena are affecting their needs for information. Representatives of the research and academic communities were asked to review and discuss scholarly work in relevant areas. Participants were asked to enumerate important changes in the health arena, to comment on their anticipated or observed impact on data needs and collection methods, and to suggest how DHHS could best contribute to meeting the nation's health data needs. Presentations by participants provided a base for the workshop's broader discussions of future directions for health statistics and the health statistics system. All participants were given the following questions to help guide the discussions:

1. Who have been the health information users, and how have their information needs changed in the past 25 years?
2. Who have been the information suppliers, and how have the skills

they need and the technology they use to produce and disseminate information changed during the past 25 years?

3. How are the data user and data provider communities, and their information needs, likely to change in the next 25 years?

4. What specific modifications must be made in the current health statistics system to adequately deal with these changes?

The overall goal of the workshop was to identify and briefly discuss as many pertinent and relative issues facing a new health statistics system as possible. Although this report is intended to thoroughly summarize the issues raised by participants during the presentations and discussions at the workshop, it is not intended to serve as a comprehensive reference on all of the issues involved in health statistics. The workshop agenda was designed to allow issues to be raised through formal presentations as well as in general discussion, during which all workshop participants could ask questions or raise issues of particular importance to them. However, due to time limitations and the flow of the discussion, it is important to note that some issues are covered in more detail than others, and some other issues in this area may not have been discussed at all. In particular, it is acknowledged that the discussions provide somewhat uneven coverage in identifying the focal issues and major developments in the health arenas of interest to DHHS. Although many are mentioned, only a few are fully articulated or explained. Also not discussed in detail are all of the potential ways in which the current data system must evolve to meet future information needs, particularly with respect to issues of overlap, integration, data sharing, and privacy and confidentiality issues. There is also little discussion of feasible alternatives and approaches to these issues. Furthermore, the report does not fully address how the current system evolved into a set of independent and inefficient "silos" spanning a very large number of federal agencies, and the bureaucratic and turf issues and constraints that make it very difficult to overcome problems. Despite the limited discussion or omission of these and other issues, it is hoped that this summary does offer the reader the true spirit of the discussions and debates among the participants at the workshop and that the ideas and issues raised in the discussions will be taken into consideration as plans for developing a new health statistics system unfold.

No attempt was made to establish a consensus among participants at the workshop, and the ideas presented do not necessarily represent those of any group, including CNSTAT and the National Academies.

2

Health Statistics: Past, Present, and Future*

Dorothy P. Rice, Professor Emeritus
Institute for Health and Aging
University of California, San Francisco

INTRODUCTION

The organization, delivery, and financing of health care services in the United States is complex, comprising an interdependence of the private and government sectors of the economy. This pluralistic health care economy, with its pragmatic mix of public and private organizations, has produced a wide range of databases that enable us to monitor the health of the nation.

Health care expenditures have been rising rapidly in the United States and claiming a larger share of national resources during the past three decades. In 1965, $41.1 billion was spent for health care, comprising 5.7 percent of the gross domestic product (National Center for Health Statistics, 1999). In 1998, health care expenditures in the United States totaled $1.1 trillion, an average of $4,094 per person, comprising 13.5 percent of the nation's gross domestic product (Levit et al., 2000). Almost 11.5 million civilians were employed in the health services industry in 1998, com-

---

*The keynote address given at the workshop is presented here in its entirety. Rice is former Director of the National Center for Health Statistics and a former member of the Committee on National Statistics. The author extends a special note of thanks to Mary Grace Kovar, Harry Rosenberg, and Samuel Marcus, who offered helpful comments on earlier drafts.

prising 8.8 percent of employed civilians (National Center for Health Statistics, 1999).

The growth of the health care industry in the United States has been accompanied by significant achievements in public health, including advances in prevention and significant declines since 1950 in death rates for diseases of the heart (56 percent), and stroke (70 percent) (Morbidity and Mortality Weekly Report, 1999). We have been successful in monitoring these and other morbidity and mortality trends through the growth and development of our health data systems.

Health care is a pressing social, political, and economic issue in the United States. The American pluralistic health care economy presents special problems for data collection, analysis, and dissemination. Health statistics systems have grown rapidly with the growth of the industry and the expansion of private health insurance coverage and public health care programs.

There is general agreement that data are needed to monitor the health of the nation; to plan and develop better health services; to deliver those services in an effective, efficient, and equitable manner; to measure their effectiveness; to make decisions on resource allocation; and to conduct research. Data also are needed to facilitate effective policy making, planning, management, and evaluation. Private organizations of health professionals, health service providers, health insurance, and many others have important interests in the collection and use of health data. The federal government needs a variety of data to support its major role in improving health and medical care delivery systems throughout the nation. State and local government agencies also have key roles in disease prevention, delivery of health services, and health planning and evaluation that require timely and reliable health statistics.

This paper presents a brief historical review of how the health statistics system has evolved to its present configuration and the lessons to be learned that might guide the future evolution of the system. This review will focus on the changes during the past 35 years in the types and uses of health statistics, the constituencies, and changes in technologies supporting the health statistics system. Gaps in health statistics, as well as several cross-cutting issues, will be discussed. Special focus will be on the federal health statistical system, especially as it relates to the production, use, and need for health data at the federal, state, and local levels. The paper concludes with challenges for the future in producing a health statistics system for the twenty-first century.

BACKGROUND

The statistical needs of the American pluralistic health care economy have grown enormously in the past 35 years since the enactment of the Medicare and Medicaid programs in 1965, the rapid growth of private health insurance, the expansion of the health care industry, and the concomitant public health, medical, and technological advances to meet the needs of a growing population. The rapid aging of the population, the emergence of chronic illnesses to replace infectious diseases as the leading causes of morbidity and mortality, and the growing heath care needs of subpopulation groups (i.e., minorities, uninsured, immigrants, and persons with disabilities and low incomes) are current phenomena that require close monitoring in the future.

Health statistics often are obtained via sample surveys conducted through telephone, mail, or in-person interviews of individuals and/or households. Health surveys go back to the Hagerstown morbidity studies conducted by the Public Health Service in the early 1920s. However, sample surveys did not become dominant until the rise of probability sampling in the 1930s (Frankel and Stock, 1969). The Public Health Service conducted the first National Health Survey in 1935-1936, funded by the Works Projects Administration (Duncan and Shelton, 1978). In 1953 the National Opinion Research Center began a series of surveys separated by five-year intervals on the consumer's use of medical care, the degree of health insurance protection, and expenditures for care (Andersen and Anderson, 1967).

In October 1953, a subcommittee of the U.S. National Committee on Vital and Health Statistics (NCVHS) recommended that a national health survey be established on a permanent basis. The passage of the National Health Survey Act of 1956 called for a continuing survey and special studies on the nation's health. It also provided for studying methods and survey techniques for obtaining this statistical information and for disseminating results of these surveys and studies. The National Health Survey, later renamed the National Health Interview Survey (NHIS), began in 1957. In 1960, the National Center for Health Statistics (NCHS) was created by combining the National Health Survey and the National Office of Vital Statistics. Responsibility for vital statistics had been transferred to the Public Health Service from the Bureau of the Census.

NCHS is the federal government's principal health statistics agency (National Research Council, 1992; Office of Management and Budget, 1998). The NCHS congressional mandate addresses the full spectrum of concerns in the health field from birth to death, including overall health status, environmental, social and other health hazards, the onset and diagnosis of illness and disability, health resources, and the use, cost, and financing of health care. NCHS also has the mandated responsibility for assisting the states and local health agencies in meeting their costs of data collection.

Although NCHS is considered the main health statistics agency, many other federal agencies also have significant responsibilities for health data collection. For example, included in the NIH statistical budget are activities that support the design and implementation of epidemiologic studies, clinical trials, biomedical research, and laboratory investigations conducted by the various institutes. Other DHHS components, such as the Centers for Disease Control and Prevention (CDC), the Substance Abuse and Mental Health Services Administration (SAMHSA), the Office of the Assistant Secretary for Planning and Evaluation (OASPE), the Agency for Health Care Policy and Research (AHCPR), the Health Care Financing Administration (HCFA), and the Health Resources Services Administration (HRSA) also actively collect health data. The Office of Management and Budget (OMB), which reviews agencies' budgets and tracks the amount allocated toward "statistical activities" reports that 13 agencies of DHHS had direct statistical budgets amounting to $804 million in fiscal year 1999; of this total, NCHS's budget comprised only 10.7 percent (Office of Management and Budget, 1998). By comparison, the statistical budget for the National Institutes of Health (NIH) comprised more than two-fifths of the total DHHS budget—$347.7 million, or 43.2 percent.

Outside of DHHS, other federal agencies collect health-related data as part of their programs, such as Bureau of the Census, Bureau of Labor Statistics, Department of Veterans Affairs, Department of Agriculture, Department of Defense, Department of Commerce, Department of Transportation, and others.

The myriad of federal agencies, with their special needs related to their health programs, clearly use considerable resources each year on multiple, decentralized program-related health surveys and statistical activities. Most federal health data systems have traditionally been developed independent of each other. Despite the large amount of money and staff resources devoted to these statistical activities, we lack information necessary to adequately assess the health status of the population, and the determinants of risks to health, and the ability to analytically relate data across these areas.

Recognizing the inadequate coordination, and the inefficient and overlapping statistical activities within DHHS, the agency targeted the improvement of the analytic capacity of health and human services programs through Reinventing Government, Part II. A survey consolidation working group was formed in early 1995 to develop a consensus plan for consolidation of surveys (HHS Survey Consolidation Working Group, 1995). The recommendations of this group will be discussed following a review of the major types of health statistics collected and key constituents of these data.

VITAL STATISTICS

Birth, death, and fetal death statistics constitute the National Vital Statistics System (NVSS) of NCHS. This program, together with the Bureau of the Census decennial census and immigration and emigration data, provides information on the dynamics of population, its growth, and changes in its composition. They furnish the information basic for making population projections, fertility analysis, planning needs for health services, for projecting school needs, and for other purposes. They are essential in the teaching and application of demography, epidemiology, sociology, medicine, and public health.

Vital statistics are provided through state-operated registration systems. Standard forms for the collection of data and model procedures for the uniform registration of events are developed and recommended for use through cooperative activities of the states and NCHS. NCHS shares the costs incurred by the states in providing vital statistics for national use. Additional programs related to the NVSS include the Linked Birth and Infant Death Data Set, the National Maternal and Infant Survey, the National Mortality Followback Survey, and the National Death Index.

The historical roots of the vital registration system go back to the earliest American settlements, when such colonies as Massachusetts and Virginia, following the English custom, required that records be kept of christenings, weddings, and burials. In time, these records shifted to the more meaningful categories of births, marriages, and deaths. Lemuel Shattuck, the leading proponent of registration, demonstrated that the health of the residents of the city of Boston was deteriorating, as measured by mortality levels. The *Report of the Sanitary Commission of Massachusetts* recommended the creation of a state board of health based on complete registration and vital statistics (Rice, 1981). By 1933, all states were registering births and deaths. Currently, the marriage and divorce statistics program is limited to the publication by NCHS of monthly counts of marriages and divorces registered in each state. Detailed abortion statistics are not collected by NCHS due to budget constraints; aggregate statistics are available through CDC's abortion surveillance system established in 1969.

In the early 1970s, NCHS pioneered the development of automated systems to process cause-of-death data through its creation of the Automated Classification of Medical Entities (ACME) system. The purpose was to apply computer systems to the complex logic for selecting the underlying cause of death from among conditions that physicians reported on death certificates, which was a costly and complicated manual coding process. Beginning with data for 1968, all U.S. death statistics were based on the application of that computer algorithm at either the state or national level. These systems produced more consistent data as well as much greater detail than ever before with the exact diagnoses reported by certifying physicians. The effectiveness of the U.S. automated systems was affirmed by the growing adoption of the systems on an international basis.

The evolution of the vital statistics program is regarded by some as an example of a most successful program, providing full counts of births and deaths at the local, state, and federal geographic levels. Except for the important issue of timeliness, the reports emanating from the vital statistics program did an excellent job of meeting the demands of users, within the limits of timely availability of the data (Committee to Evaluate the National Center for Health Statistics, 1973). The availability of data electronically in recent years has gone a long way to improving the timeliness of vital statistics data, thereby enhancing the usefulness of the data.

The introduction of the National Death Index (NDI), a computerized index of death record information beginning with 1979 deaths, has made enormous contributions to more efficient epidemiologic and other health studies in which researchers can go to one source, NCHS, to obtain mortality information on their study participants. Prior to the establishment of the NDI, each state had to be contacted separately for such information on file in the state vital statistics offices.

PUBLIC HEALTH SURVEILLANCE

Public health surveillance is the "ongoing systematic collection, analysis, and interpretation of data on specific health events affecting a population, closely integrated with the timely dissemination of these data to those responsible for prevention and control" (Thacker et al., 1996:633). A feature of surveillance is the ability to identify individuals and groups of individuals for further action on prevention and treatment. This construct raises issues of privacy and confidentiality, rapidly evolving issues that reflect the complex interplay of personal rights, ethical concerns, legal responsibility, and societal interest in the general welfare of the population and public health. Privacy and confidentiality will be discussed further below.

The National Notifiable Disease Surveillance System illustrates traditional surveillance in which physicians, clinical laboratories, and other health care providers are required by state law to report all cases of health conditions, mainly infectious in origin, that are specified as being notifiable. The Council of State and Territorial Epidemiologists determines which notifiable conditions should be reported by state health departments to the CDC. The CDC and other federal agencies are involved in the collection of surveillance data, including, but not limited to the following:

- the National Institute of Occupational Safety and Health has maintained a sentinel health event verification system for occupational risk;
- the Food and Drug Administration conducts postmarketing surveillance of adverse reactions to drugs;
- the National Cancer Institute conducts the Surveillance, Epidemiology, and End-Results (SEER) Programs;
- the Behavioral Risk Factor Surveillance System, a telephone survey conducted in each of the 50 states, and supported in part by the CDC, provides data on health behaviors; questions can be added by the individual states;
- the Pregnancy Assessment Monitoring System developed by CDC to collect information on maternal behaviors that occur before, during, and shortly after pregnancy among women who deliver a live-born infant; and
- the Consumer Product Safety Commission conducts surveillance on product-related injuries.

Surveillance data vary in their quality and often are incomplete and unrepresentative, and they may vary in sensitivity and specificity (Stroup and Teutsch, 1998). Although the current programs provide essential data to monitor the incidence of communicable diseases and some chronic diseases, the system relies on voluntary physician reporting, which has been demonstrated to be variable and inconsistent. States differ in their authority to require physician reporting. Development of greater standardization in reporting from state to state, and obtaining improved physician cooperation are areas that need further exploration.

Population-based registries and national sample surveys have also been used for surveillance purposes. Registries are established to identify cases through several sources (e.g., schools, hospitals, and laboratories). Registries require extensive confirmation of cases, leading to longer lag times between a health event and the reporting of such an event. The National Cancer Institute SEER program covers about 10 percent of the U.S. population; it provides data that are used to monitor long-term trends of cancer incidence and mortality. Currently, approximately 30 states have population-based registries, but they may be limited by both under-registration and selection bias (Stroup and Teutsch, 1998).

HEALTH STATUS, HEALTH CARE UTILIZATION, AND MEDICAL CARE COSTS

Statistics abound on health status and use of medical care services at the federal, state, and local levels. The National Health Interview Survey (NHIS) and the National Health and Nutrition Examination Survey (NHANES) are the major national surveys for assessment of health status in the United States and are sponsored by NCHS. NHIS is a primary source of information on the health of the civilian, noninstitutionalized population of the United States. Conducted continuously since 1957, it provides national data on the annual incidence of acute illness and accidental injuries, the prevalence of chronic conditions and impairments, the extent of disability, the utilization of health care services, and other health-related topics. To provide data on special topic areas in addition to the basic NHIS data, extensive supplements have been conducted annually. Topics covered in the supplements vary from year to year. For example, in 1995 the supplements included questions on the following: immunization, children's and adults' disability; follow-up on persons with disabilities interviewed in the prior year, family resources (access to care, health care coverage, income and assets), year 2000 objectives (tobacco use, nutrition, clinical preventive services, physical activity and fitness, and AIDS knowledge and attitudes (National Center for Health Statistics, 1998). The NHIS sample design includes about 40,000 households interviewed, resulting in a sample of about 102,000 individuals, with oversampling of black and Hispanic persons.

NHANES was established in 1971 to collect the kinds of health data best obtained by direct physical examinations and physiological and biochemical measurements. NHANES is the cornerstone of the National Monitoring and Related Research Program, providing data needed for nutrition monitoring, food fortification policy, establishing dietary guidelines, and assessing government programs and initiatives such as Healthy People 2000 and 2010 objectives of DHHS. In the past, researchers sometimes had to wait as long as 10 years after data collection before gaining access to data based on the entire 6-year sample. Now, NHANES is a continuing, annual survey, linked to the NHIS, and data are being collected from a representative sample of the U.S. population, newborns and older, every year.

NHIS and NHANES are only two of the many national federal surveys that collect data on heath status, medical care utilization, and insurance coverage. Other important federal surveys collect similar data as well as data on medical care expenditures:

- The National Immunization Survey (NIS) is a continuing nationwide telephone sample survey to gather data on children 19-35 months of age. In 1997, data were obtained for 32,742 children to provide estimates of vaccine-specific coverage for national, state, and 28 urban areas considered to be high risk for under-vaccination.
- The Medical Expenditure Panel Survey (MEPS) conducted by the Agency for Health Care Policy and Research (AHCPR) is a study of approximately 9,000 households. MEPS is a subsample of NHIS participants, providing health status and other data for enhanced analytical capacity. Use of NHIS data in concert with the data collected in the 1996 MEPS provides the capacity for longitudinal analysis. Each sample panel is interviewed a total of five times over 30 months to yield annual use and expenditure data for two calendar years. The 1996 MEPS household component reflects an oversampling of households with Hispanics and blacks (Cohen et al., 1999). MEPS also has an institutional component.
- The National Household Survey on Drug Abuse (NHSDA), conducted by SAMHSA, focuses on the incidence, prevalence, consequences, and patterns of substance use and abuse. In 1997, the NHSDA was expanded from 18,000 respondents to about 25,000 respondents to generate estimates for the nation and for two states (California and Arizona). In 1999, the NHSDA was further expanded to 70,000 respondents to generate estimates for all 50 states.
- The Medicare Current Beneficiary Survey, conducted by HCFA, is

an ongoing rotating panel survey of approximately 12,000 aged and disabled Medicare beneficiaries, consisting of four overlapping panels of Medicare beneficiaries surveyed each year. Each panel contains a national representative sample of beneficiaries who are interviewed 12 times in the community or a long-term care facility to collect three complete years of utilization data. The survey provides comprehensive data on health and functional status, use of medical services, covered and noncovered health care expenditures, and health insurance for Medicare beneficiaries.

- The National Health Care Survey is a family of NCHS provider-based surveys that measure the utilization of health services through a series of surveys of providers. Included are hospitals (National Hospital Discharge Survey), physicians (National Ambulatory Care Survey), emergency and outpatient departments (National Hospital Ambulatory Medical Care Survey), ambulatory care centers (National Survey of Ambulatory Care Surgery), nursing homes (National Nursing Home Survey), and health agencies providing home health care services and hospice care (National Home and Hospice Care Survey).
- The National Survey of Family Growth (NSFG) is a periodic survey of women ages 15 to 44 years. The purpose of the survey is to provide national data on factors affecting birth and pregnancy rates, adoption, and maternal and infant health. In 1995, for the first time, the sample was obtained from households that had been interviewed in the NHIS. A total of 10,847 women were interviewed, and Hispanic and black women were oversampled. Cycle 6 of the NSFG will include a sample of men for the first time.
- The Healthcare Cost and Utilization Project, conducted by AHCPR, consists of the State Inpatient Database (SID) and the Nationwide Inpatient Sample (NIS). SID contains all hospitals and all discharges from 22 participating states. AHCPR receives the data from each statewide data organization, processes the data into a uniform format, and then returns the uniform SID files to the statewide data organization. The NIS database contains a sample of hospitals selected from SID. The NIS comes with weights that can be used to produce national estimates, regional estimates, and state estimates for participating states.
- The Current Population Survey (CPS) is a monthly sample survey of about 50,000 households conducted by the U.S. Bureau of the Census for the Bureau of Labor Statistics. The CPS is the primary source of information on labor force characteristics of the U.S. population. Monthly estimates from the CPS include employment, unemployment, earnings,

hours of work, and other indicators. The annual March supplement produces national and state estimates of health insurance coverage, including private health insurance, Medicare, Medicaid, the Civilian Health and Medical Program of the Uniformed Services (CHAMPUS), and military health care.

In addition to the federal health statistics surveys and programs briefly discussed above, each of the 50 states and the private sector maintain data systems and conduct many surveys of hospitals, health professionals, and health care organizations. The private health sector includes organizations of health service providers, health professionals, health insurance payers, consumers, industry, and private philanthropy. Many national and state data collection activities are conducted by these private organizations, but their quality is variable. The results of all these statistical efforts are duplicative and overlapping data systems in the public and private sectors. Hospital inpatient data, for example, are collected in the public sector by NCHS, AHCPR, HCFA, SAMHSA, Veterans Administration, and others. Most states have their own hospital discharge data systems conducted by state rate-setting, planning, and health systems agencies. In the private sector, hospital data are collected by the American Hospital Association, many abstracting organizations, Blue Cross, professional standard review organizations, and health maintenance organizations. It is recognized that hospital data are necessary to understand, monitor, and evaluate programs related to hospital-based delivery of health care. The reporting burden on hospitals, however, is great; recording, storing, abstracting, and processing medical records is expensive for both the institution and the users. The rationale for these overlapping and duplicative hospital inpatient data is difficult to justify.

USERS OF HEALTH STATISTICS

Few people now dispute the need for data as a basic requirement for the development of policy in every area of our national life, whether it be health, environment, education, employment, delinquency and crime, defense, agriculture, transportation, welfare, housing or any important area (Hauser, 1975). In the health area, the need to plan for appropriate levels of health resources, to protect the public from hazards in the workplace and environment, and to evaluate the effectiveness of health programs led to increased demands for data on health status, health resources and their utilization, costs, and financing. The HHS Survey Consolidation Working Group (1995) listed the following key customers and constituents of HHS.

- HHS agencies that use data for program management, actuarial projections, health policy (see Kronick, 1999), and evaluation;
- HHS researchers who use data in epidemiologic studies, identification of hypotheses for biomedical research, demographic studies, etc.;
- analysts in other executive and legislative organizations, including the Office of Management and Budget, the Congressional Budget Office, and the General Accounting Office;
- state and local governments involved in public health and health care financing;
- academic researchers working independently and with HHS support;
- advocacy groups, including advocates of public health, disease-specific medical research, children, the aged, Americans with disabilities, and those advocating strategies for health reform;
- the private sector, including providers, managed care organizations, and associations.

The following may be added to the above (Hauser, 1975):

- patients, families, and community residents in general;
- social service workers;
- payers, purchasers, and regulators with interest in and responsibility for evaluating quality and outcomes of care;
- community and consumer organizations, charitable and volunteer groups in the health area;
- legal professionals bringing criminal or civil charges, and others.

GAPS IN HEALTH STATISTICS

State-Level Data

Many federal health data systems have evolved from specific program needs as the federal role in health care has expanded in the last three decades. The needs for health data at the state and local levels, however, have not adequately been met. The Cooperative Health Statistics System, a nationwide cooperative network of public and private agencies linked together to meet their respective needs for health statistics, is no longer supported by the federal government (National Center for Health Statistics, 1980). While the responsibilities of the states for administering health care programs are being expanded greatly, there has not been a commensurate increase in resources devoted to their statistical functions and their needs for health data. For example, the growing racial and ethnic diversity within the states raises important issues with respect to monitoring the health status and health care costs and financing of minority groups. However, sample sizes in national data sets are not large enough to disaggregate at the state level. In addition, there is a wide range of ability and capacity among state statistical programs to collect, analyze, and interpret health data.

Health Statistics for Subpopulation Groups and Minorities

As efforts continue to reduce health disparities among special population groups of low-income persons, racial and ethnic minorities, and persons with disparities, it is recognized that data are needed to monitor our progress toward eliminating these disparities. Except for the data derived from the decennial census and from the vital registration system (birth and death statistics), the existing sources of health data do not permit examination of socioeconomic differences for any but the three largest race and ethnic categories: non-Hispanic white persons, non-Hispanic black persons, and persons of Hispanic or Mexican origin. Data shown for broad groupings usually mask significant differences among subgroups. For example, "Asian or Pacific Islander" includes persons with ancestry in such countries as China, Vietnam, the Phillippines, Japan, and Samoa, while "Hispanic" combines persons whose origins were Cuba, Puerto Rico, Mexico, or any other countries of Central or South America. These subgroups often have very diverse health status and risk behavior. It is essential that our health statistical systems at the national and state levels capture this diversity.

Longitudinal Data

The surveys, surveillance, and vital statistics programs meet many of the current needs for health data. The cross-sectional survey data give a "snapshot" at a point in time of the health status of people at different stages in their lives and allow periodic examinations of changes over time. Still needed, however, are large-scale longitudinal efforts that record in sequence the health events of life. Longitudinal efforts in the health area are limited. Recent examples of relatively short-term follow-ups of survey participants include the NHANES I Epidemiologic Followup Study, the Longitudinal Study on Aging, NHIS Disability Supplement, MEPS, and the Medicare Current Beneficiary Survey (MCBS).

The needs for additional longitudinal studies have been specifically addressed in at least 10 of the 62 reports dealing with some aspects of health data and data systems published by the National Research Council and Institute of Medicine since 1985 (Jane Durch, personal communication, September 9, 1999). These reports have specifically recommended the need for longitudinal studies, such as the recommendation that NCHS develop and implement a continuous, longitudinal survey of health care utilization and expenditures, and their health care providers, using cohorts of individuals selected from among NHIS survey respondents (National Research Council and Institute of Medicine, 1992)

Other Identified Gaps

Testimony to the many other identified gaps in current health statistics and the needs for specific health data is documented in the recent compilation of the conclusions and recommendations regarding health data and data systems in the published reports of the National Research Council and Institute of Medicine since 1985 (Jane Durch, personal communication, September 9, 1999). A total of 62 studies made 249 recommendations relating to health data and data systems, averaging about 4 recommendations per study. The number of recommendations ranged from 1 per study (14 studies) to 17 in one study, *Toward a National Health Care Survey: A Data System for the 21st Century* (National Research Council and Institute of Medicine, 1992). The specifics of these recommendations are too many to enumerate. The recommendations are varied, including the proposed establishment of a surveillance system to detect, monitor, and warn of adverse effects in the recipients of blood and blood products (Institute of Medicine, 1995); the collection of person-based longitudinal information in the National Health Care Survey, expanding the data collected to include, but not be limited to, information on the health care received, costs and gross expenditures for health care, and outcomes (National Research Council and Institute of Medicine, 1992); recommendations for additional resources for current or new data systems, and many others.

On one hand, implementation of the many health data recommendations enumerated in the National Research Council reports would proliferate the existing fragmented data systems, especially at the federal level; on the other hand, there are many identified gaps in the existing data systems and many needs for improved health statistics to fill them.

CROSS-CUTTING ISSUES

Overlap and Consolidation

In 1995, DHHS recognized that it had a considerable investment in surveys and other data systems to support broad analytic and program objectives, and that the operation of these data systems and surveys was decentralized with limited central strategic planning and direction, resulting in overlaps with respect to populations of interest, analytic capabilities, sample and questionnaire designs, and collection efforts. Through Reinventing Government Part II, DHHS formed an interagency survey consolidation working group that on April 11, 1995, reported its *HHS Plan for Consolidation of Surveys* (HHS Survey Consolidation Working Group, 1995). Key features of the plan included 15 specific proposals, including establishing the capability to use NHIS as a sampling frame for other surveys such as MEPS, NHANES, NSFG, and NHSDA. By moving to this consolidated, annual household data collection effort, the analytical capabilities of these surveys have been significantly expanded and enhanced. The NHIS household interview core questionnaire provides population-based statistics on health status and health care utilization with sufficient sample size to allow for analyses based on disaggregation of detailed age, race, sex, income, and other sociodemographic characteristics, and allows for the collection of data on a broad range of topics provided by NHIS.

Implementation of some of the above recommendations has been a significant step forward in data collection and a very exciting development in health statistics. In 1996, the MEPS sample was composed of the fourth-quarter NHIS sample. When MEPS data are linked to NHIS data, the microanalytical potential for studies of health status, prevalence of chronic conditions, health care coverage, and utilization of and expenditures for medical care services is greatly enhanced. Likewise, NHANES and NSFG also are now using the sampling frame of the NHIS.

Unfortunately, no progress has been made on the coordination of the National Household Survey on Drug Abuse (NHSDA) with NHIS. As noted earlier, the sample size of the NHSDA has been expanded to 90,000 persons in 1999 to enable the production of the state estimates. The Survey Consolidation Working Group had recommended that a design framework in the NHSDA could be consolidated with NHIS in several respects, including (1) closer coordination between the questionnaires of the two surveys; (2) using NHIS as the sampling frame for NHSDA; and (3) conducting NHSDA as a supplement to the NHIS. One wonders why there are sufficient resources for the expansion of NHSDA, but not for NHIS, to provide state estimates.

Other important proposals for consolidation included: (1) merging the National Nursing Home Survey and the MEPS institutional component into an integrated, periodic survey of nursing home capacity, services, utilization, and expenditures; (2) the Medicare Current Beneficiary Survey should be closely coordinated with MEPS in terms of greater questionnaire coordination and analytical linkages; and (3) design of a state-level telephone survey to obtain basic health status, access to care, insurance, and expenditure data of importance for national policy analysis, performance evaluation, and modeling. By using an expanded NHIS sample and questionnaires from the consolidated national surveys, state-level data can be obtained efficiently and will be directly comparable to national data. Information on access to health care and health insurance coverage is needed by the states. To date, such information has been available from the CPS March supplement. Questions may be raised as to why such important health data at the state level have to come from a labor force survey rather than from a health survey.

It is clear that considerable progress has been made within DHHS in survey consolidation. But we have a long way to go to eliminate fragmentation and overlap in DHHS surveys as well as in other federal and state agencies and the private sector.

Computer Technology

The information sector of society is rapidly changing with the evolution of computer technology. The development and widespread use of the computer unquestionably has been one of the great technological changes in the past 50 years. One of its effects on statistics has been very large reductions in clerical personnel requirements and consequent reductions in total costs. The most pervasive effects of computers on health statistics have been in the dramatic changes in all aspects of data collection, data analysis, and data dissemination. We now have the ability to do things in all these areas that could not be done at all without computers, either because they could not be done in time to be useful or because they would have cost too much to be practical.

We have witnessed significant changes in methods of sample survey data collection in recent years, from personal household interview surveys to random-digit-dialing telephone surveys, from computer-assisted telephone interview surveys to computer-assisted personal interviews. Expanded access to the Internet and more powerful computing hardware for management and processing of data have had positive effects on the accuracy of data collected and disseminated as well as on the timeliness of available data. Access to health data via computers has clearly increased, thereby changing and expanding the user pool and thus the uses of the data.

The amount of data about individuals and their use of the health delivery system has grown exponentially as a direct result of advances in computer technology. The ability to capture and retain information on individual records and the use and disclosure of personally identifiable medical information have been the subjects of substantial discussion by government agencies, professional associations, and others. This issue is discussed below as it relates to health statistics.

Privacy Protection, Confidentiality, and Data Sharing

The conflict between freedom of information and invasion of privacy in relation to data collection has received increasing attention in recent years. A balance must be struck between the public's right to know and the right of individuals and institutions to protect their privacy. Even in those programs where strong legal safeguards and technical procedures protect the confidentiality of the information collected, there remains a persistent fear that this vast complex of information might be used as an instrument of social control, if not for commercial purposes.

Advances in technology and the increasing collection of personal data for public and private decision making are raising concerns among many Americans about the confidentiality of the information they provide for use in government surveys. Both individuals and businesses are questioning how the information is used and who has access to it. At the same time, data users, especially those outside government, are increasingly frustrated by limits on the amount of detailed information they can obtain from statistical agencies.

A CNSTAT report on confidentiality and accessibility of government statistics offered principles and specific recommendations for managing data for research and policy making and the confidentiality of information (National Research Council, 1993). Another Institute of Medicine report offered recommendations related to public disclosure of quality-of-care information and protection of the confidentiality of personal health information (Institute of Medicine, 1994). A more recent study, *For the Record: Protecting Electronic Health Information* (National Research Council, 1997), dealt with the need for the health care industry to create the infrastructure necessary to support the privacy and security of electronic health information. These reports recognized that diverse groups of researchers, business leaders, and policy makers have developed databases to permit increasingly sophisticated analyses of community health needs, practice patterns, costs, and quality of care. Greatly enhanced electronic capabilities for data management create opportunities for easy linkage of health data files, resulting in concerns about misuse of the information and how well the privacy and confidentiality of personal health information will be guarded.

Data Sharing and Data Linkage

It is beyond the scope of this paper to deal with ways to protect patient- and person-identifiable health and medical data collected in the public and private sectors. One aspect of the privacy and confidentiality issue, however, that could have a significant impact on reduction of duplicative and overlapping reporting systems is data sharing and data linkage among government agencies. It has long been recognized that the development of comprehensive data systems concerning the interrelations among various aspects of social and economic patterns sometimes requires that various data sets be combined. Recommendations have been made for exchange of statistical data under legislatively mandated “protected enclaves” for selected statistical and research agencies within the federal government (Office of Federal Statistical Policy and Standards, 1978). The House Government Reform Committee approved H.R. 2885, the Statistical Efficiency Act, which designates eight federal statistical agencies as statistical data centers and allows for limited sharing of statistical information by other agencies with these data centers and sharing among the statistical data centers. If approved by the Senate, this Act will go a long way to facilitate sharing of data among federal agencies. For example, NCHS could use business data from the Census Bureau or the Bureau of Labor Statistics to construct sampling frames for surveys of employers or health providers, or use census data from the Census Bureau to augment population samples.

Linking public and private data is an area with tremendous potential for analysis if issues of confidentiality can be overcome. Linking such data is especially important for medical effectiveness and outcomes research, which examines the effects of alternative treatments of a given medical condition on the eventual outcomes realized by the patients (Agency for Health Care Policy and Research, 1991).

Data integration among current large data collection activities should be carried out to maximize the results of separate efforts. Linkage of data files should be encouraged when there is good reason to believe that the results of a specific linkage program will be sufficiently complete for the specific purpose and that biases and limitations of linkage studies will not be so severe as to vitiate results. Linkages should be carried out only when there is a hypothesis that can be investigated through linked data and if suitable safeguards of confidentiality can be applied when the research benefits exceed any potential risks to subjects.

Quality and Reliability of Data

Many organizations, especially federal statistical agencies, such as NCHS, have made and continue to make considerable and commendable efforts to maintain and improve the quality of their major statistical series. In the private sector, however, the quality and reliability of data are uneven and unknown in many ongoing databases. Survey results are subject to sampling, reporting, processing, and nonresponse errors; the data cannot be fully understood and properly used unless these errors are reported. Standard errors are routinely reported in federal statistical reports on survey results but are unavailable in most reports emanating from facility and manpower surveys conducted in the private sector. The improvement of the quality and reliability of health statistics in the private sector is most urgently needed.

Standardization of Data Elements, Uniform Definitions and Coding, Minimum Data Sets

Health data are collected by many organizations and at multiple geopolitical levels for a variety of uses. Standardization of data elements across programs is necessary to permit comparisons and to avoid duplicative efforts. Considerable progress has been made at the federal level in providing standards for data collection, analysis, and distribution by the Office of Management and Budget. For example, a standard classification for race and ethnicity has been promulgated and implemented by federal agencies in all their data collection activities. In subject areas of health insurance and disability, it is essential that questions be standardized across federal surveys. Currently, the surveys often produce different estimates because of the lack of standardization.

Some progress has been made in the development of uniform minimum data sets under the auspices of the NCVHS, but this effort must be continued as the needs for data at the state and local levels continue to grow.

Administrative Simplification

The Health Insurance Portability and Accountability Act (HIPAA) of 1996 includes a section on administrative simplification aimed at the reduction of administrative costs and burdens in the health care industry. It requires the DHHS to adopt national uniform standards for electronic transmission of certain health information. It is understood that the adoption of uniform national standards for electronic processing of insurance claims and related transactions will improve information flow and help generate significant savings, while improving efficiency and enhancing the quality of health care services. The law may remove some roadblocks that now impede access to some data or make it difficult to link benefits, services, and outcomes.

HIPAA also requires DHHS to develop a unique health identifier for each individual, employer, health plan, and health care provider. The NCVHS is charged with offering technical support and advice to the Secretary of DHHS on the development of this unique identifier. The Committee is currently working with standard-setting organizations to identify, define, agree on, and then implement uniform standards. At the same time, insurers and providers will have to review and revise existing data infrastructures. Also, important and difficult issues relating to privacy of individually identifiable health information are being addressed. With the implementation of the Administrative Simplification subtitle of HIPAA, for the first time in its history the United States will have the means to monitor the health, health care, and health care costs of its entire population (Pollock and Rice, 1997).

CHALLENGES FOR THE FUTURE

From this brief historical review of health statistics in the United States, we reluctantly conclude that despite improvements, health statistics production in this country presents a picture of fragmented data collection, lack of common definitions and uniformity of reporting, duplicative and overlapping systems, and resistance to data sharing. We are encouraged that progress has been made along some fronts, but we have a long way to go to fill the data gaps and to provide the health statistics needed for the twenty-first century.

As former director of one of the federal statistical agencies, I believe that the best way to provide objective high-quality information on the demographic, economic, social, and health characteristics of our population, and trends in those characteristics, is through agencies specifically established for that purpose. These agencies have “no axe to grind,” can usually guarantee confidentiality to respondents, and hence are able to produce unbiased quality information acceptable to a wide array of users both within and outside government. However, even in the best of economic times it is difficult to obtain adequate budgets to support the necessary data collection and analysis activities. Recognizing the philosophy that the federal government should only be in the business of doing things that cannot be adequately done by states and/or the private sector, it may be necessary to reassess the core programs of the federal statistical system.

Regardless of what changes must be made in the core programs, we must ensure that an information base continues to be available that will provide baseline data, be useful for monitoring trends, and have the ability to quickly detect any changes or aberrations in the economic, social, or health characteristics of the nation. The appropriate federal role in statistics is to produce national-level data useful for those purposes as well as to provide norms to which subnational data can be compared. The data must be of high quality, produced in a timely manner, and relevant to issues of the day.

Federal statistical agencies must assume responsibility for activities that cannot reasonably or feasibly be assumed by individual states, local governments, and the private sector. The federal role must include the development and promulgation of standards and procedures for assuring the validity, reliability, comparability, and quality of statistical products and the provision of technical assistance in these areas. Federal statistical agencies also must anticipate future needs for information and design today's systems to meet those needs.

In considering future prospects for improved health statistics to meet the needs of the twenty-first century, we must recognize that resources will not grow parallel to demands for data and services. The demands for health data are greater than our ability to produce them. Budgetary pressures are requiring assessment of current data collection and dissemination procedures. Statistical agencies must make choices between data collection, research, and analysis, and among needed data sets.

As we move closer to our objective of a national and systematic approach to meeting the information needs for health policy development and program evaluation, we also need to coordinate our data collection activities, both within the federal establishment and between government and the private sector. Although considerable progress has been made in coordination, we must continue to avoid unnecessary and costly duplication, to encourage comparability of information collected by different systems, and to use the ongoing data collection programs to provide specific information for many organizations. More effort is needed to provide essential data, yet reduce the burden on individual and institutional respondents. We must develop and articulate a twenty-first century vision for health statistics.

REFERENCES

Agency for Health Care Policy and Research
1991 *Report to Congress: The Feasibility of Linking Research-Related Data Bases to Federal and Non-Federal Medical Administrative Data Bases.* AHCPR Pub. No. 91-0003. Washington, DC: U.S. Government Printing Office.

Andersen, R., and O.W. Anderson
1967 *A Decade of Health Services: Social Survey Trends in Use and Expenditures.* Chicago, IL: The University of Chicago Press.

Cohen, S.B., R. DiGaetano, and H. Goksel
1999 *Estimation Procedures in the 1996 Medical Expenditure Panel Survey Household Component.* MEPS Methodology Report No. 5: May. AHCPR Pub. No. 99-0027. Rockville, MD: Agency for Health Care Policy and Research.

Committee to Evaluate the National Center for Health Statistics
1973 Health Statistics Today and Tomorrow: A Report of the Committee to Evaluate the National Center for Health Statistics. *Vital and Health Statistics*, Series 4, No.15. DHEW Publication No. (HRA) 74-1452: September.

Duncan, J.W., and W.C. Shelton

1978 *Revolution in United States Government Statistics.* Office of Federal Statistical Policy and Standards, U.S. Department of Commerce. Washington, DC: U.S. Government Printing Office.

Frankel, L.R., and J.S. Stock

1969 On the sample survey of unemployment. *Journal of American Statistical Association;* December 1969:77-80.

Hauser, P.M.

1975 *Social Statistics in Use.* New York: Russell Sage Foundation

HHS Survey Consolidation Working Group.

1995 *HHS Plan for Consolidation of Surveys:* April 11, 1995. Washington, DC: U.S. Department of Health and Human Services.

Institute of Medicine

1994 *Health Data in the Information Age: Use, Disclosure, and Privacy.* Committee on Regional Health Data Networks. Molla S. Donaldson and Kathleen N. Lohr, editors. Division of Health Care Services. Washington, DC: National Academy Press.

1995 *HIV and the Blood Supply: An Analysis of Crisis Decisionmaking.* Committee to Study HIV Transmission Through Blood and Blood Products. Lauren B. Leveton, Harold C. Sox, Jr., and Michael A. Stoto, editors. Division of Health Promotion and Disease Prevention. Washington, DC: National Academy Press.

Kronick, R.

1999 Numbers We Need: Health Statistics and Health Policy. Paper presented at the Committee on National Statistics Workshop, Toward a Health Statistics System for the 21st Century, November 4, Washington, DC. Available: <http://www.ncvhs.hhs.gov/hsvision/visiondocuments.html>. [July 12, 2001]

Levit, K., C. Cowan, H. Lazenby, A. Sensenig, P. McDonnell, J. Stiller, A. Martin, and the Health Accounts Team

2000 Health Spending in 1998: Signals of Change. *Health Affairs* 19: 124-132.

Morbidity and Mortality Weekly Report

1999 Achievements in Public Health, 1990-1999. August 6, 1999. 48(30): 649-656.

National Center for Health Statistics

1980 Directions for the '80s: Final Report of the Panel to Evaluate the Cooperative Health Statistics System. DHHS Pub. No. (PHS) 80-1204. Hyattsville, MD: U.S. Public Health Service.

1998 Current Estimates From the National Health Interview Survey, 1995. *Vital and Health Statistics* 10(109). DHHS Pub No. (PHS) 98-1527. Washington, DC: U.S. Government Printing Office.

1999 Health United States, 1999 With Health and Aging Chartbook. DHHS Pub. No.(PHS) 99-1232. Washington, DC: U.S. Government Printing Office.

National Research Council

1992 *Principles and Practices for a Federal Statistical Agency.* Committee on National Statistics. Margaret E. Martin and Miron L. Straf, editors. Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

1993 *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics.* Panel on Confidentiality and Data Access. George T. Duncan, Thomas B. Jabine, and Virginia A. de Wolf, editors. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

1997 *For the Record: Protecting Electronic Health Information.* Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure. Computer Science and Telecommunications Board, Commission on Physical Sciences, Mathematics, and Applications. Washington, DC: National Academy Press.

National Research Council and Institute of Medicine

1992 *Toward a National Health Care Survey: A Data System for the 21st Century.* Panel on the National Health Care Survey. Gooloo S. Wunderlich, editor. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, and Division of Health Care Services, Institute of Medicine. Washington, DC: National Academy Press.

Office of Federal Statistical Policy and Standards

1978 *A Framework for Planning U.S. Federal Statistics for the 80's.* U.S. Department of Commerce. Washington, DC: U.S. Government Printing Office.

Office of Management and Budget

1998 *Statistical Programs of the United States Government: Fiscal Year 1999.* Washington, DC: Executive Office of the President.

Pollock, A.M., and Rice, D.P.

1997 Monitoring health care in the United States: A challenging task. *Public Health Reports:* 112:108-113.

Rice, D.P.

1981 Health statistics: Past and present. Editorial. *The New England Journal of Medicine* 305(4): 219-220.

Stroup, D.F., and Teutsch, S.M.

1998 *Statistics in Public Health: Qualitative Approaches to Public Health Problems.* New York: Oxford University Press.

Thacker, S.B., Stroup, D.F., Parrish, R.G., and Anderson, H.A.

1996 Surveillance in environmental public health: Issues, systems, and sources. *American Journal of Public Health* 86(5):633-638.

3

Workshop Discussion

DEFINING HEALTH, MEASURES OF HEALTH AND HEALTH CARE, AND DATA NEEDS FOR A FUTURE HEALTH STATISTICS SYSTEM

Through a myriad of methods, health data have been and continue to be collected that provide information on the health status and health care utilization of our population. Although these data are used to answer important and necessary research and policy questions, they are also often used to answer questions that they may never have been intended to answer. The current health statistics system, in particular, has evolved to meet many needs, but perhaps, with minimal vision to the evolution. One of the goals of the workshop was to identify current and future data gaps with respect to the health and policy questions the future health statistics system should be able to answer. In her introductory remarks at the workshop, Margaret Hamburg, Assistant Secretary for Planning and Evaluation at the U.S. Department of Health and Human Services (DHHS), stated that the focus on data for decision making was one of three top priorities for the Secretary of Health. Edward Sondik followed with the charge to workshop participants, "we really need to be looking ahead and not looking back. If we can't anticipate the future, we don't know what information it is that we need to collect." Continuing his introductory comments, Sondik reminded participants that "the goal in all of this is ultimately to improve health status, whether we are talking about issues of disparities, issues of health care organization, or health care research." So, in addition to identifying existing data needs inherent to the current health statistics system, many workshop participants suggested new data that be should be collected in a health statistics system that perhaps better defines the vision of health for the future.

If a goal of a health statistics system is to improve health status, then the question of what is health is raised. It was argued by some workshop participants that, as a starting point for determining data content, it is important to develop a common understanding of the concept of health and of the factors that contribute to health. Prior to discussions that outlined specific data needs, several participants offered their perspectives on this issue. Charlyn Black, of the Manitoba Centre for Health Policy and Evaluation at the University of Manitoba, emphasized in her presentation that health is more than the product of an interaction with the medical care system. She considered health to be influenced by a variety of interactions within the larger ecosystem. Black suggested that if the focus is health and improving health, then the broad range of factors that influence health must be considered. Whereas a major model for health statistics has been to examine the utilization of health care as an indicator of overall health, the Canadian model, as expressed by Evans and Stoddard (1990), focuses more on other factors that influence health status and health-related quality of life. An overreliance on data concerning the utilization of health care, they argue, also results in a lack of appreciation for and knowledge of the outcomes of care. Black stated that it is incumbent upon a national health information system to provide information on both the impact of health care on the health status of patients and the effects of other factors on this health status (see Figure 3-1). She labeled this approach a "population health perspective." Black suggested that it will be necessary to provide data to respond to a change in the policy discussions from a focus on what health care services are being provided to a focus on what is being done to improve health. The concern that health information systems are relying too heavily on medical care utilization information to describe the health of a population was suggested by others in attendance at the workshop. Richard Kronick, from the University of California at San Diego, commented that without knowing if medical care has much effect on health status, researchers will not know if it is worthwhile to continue to investigate medical care as a major determinant of health. He cited as an example the fact that no one has been able to produce a reasonable estimate of the effects of the Health Security Act on the health status of the U.S. population.

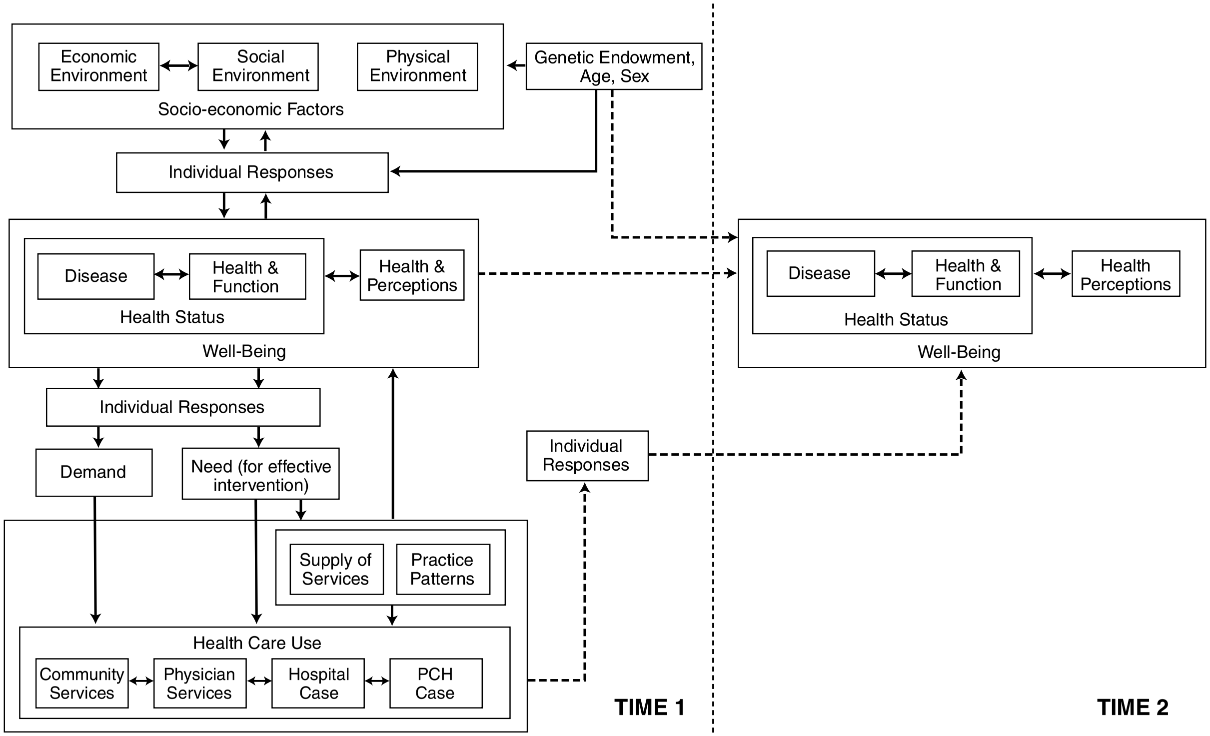

FIGURE 3-1 Conceptual model underlying the population health information system.
SOURCE: Black et al. (1999).

In describing her experience with measuring children's health, Lorraine Klerman, of the University of Alabama, Birmingham, emphasized the need to convey to legislators that other things affect the health of children, in particular, besides medical care. Klerman described the State Children's Health Insurance Program (SCHIP), and noted that the policy goals of SCHIP were focused more on ensuring participation by low-income uninsured children in SCHIP without extensive displacement of private coverage, coverage that includes benefits appropriate for children, continuity of coverage, and effective coordination with Medicaid, rather than goals emphasizing the improvement of the health of children. She noted that the focus on health promotion and disease prevention has not been accompanied by the development of indicators that would measure positive health, rather than illness or injury. Klerman outlined the need for data on children's health and emphasized the need for measures of child health rather than disease. However, even the measures of disease should be reconsidered when children are being_studied. For example, mortality data do not really help us figure out how healthy children are; hospitalization data are not sufficient because hospitalization of children is relatively rare, and individual diseases are relatively infrequent. She noted that several national surveys that contain questions on children's health, such as the National Hospital Discharge Survey, the National Ambulatory Medical Care Survey, and the National Hospital Ambulatory Medical Care Survey, are not able to distinguish child health from child ill-health. Therefore, it would appear necessary, according to Klerman, to refocus attention toward positive measures of child health such as quality of life and level of functioning. Only then will an accurate portrayal of the effects of a program such as SCHIP on child well-being be achieved.

Other participants also suggested the idea that health is more than presence or absence of a negative state. For example, Dorothy Rice, from the University of California at San Francisco, in her opening remarks at the conference commented that a refocusing on positive health measures is needed (see Chapter 2). Burton Singer, from Princeton University, proposed a model that represented health as a more positive state than has generally been described in the past. Referring to data that he and his colleague, Carol Riff, have collected and analyzed, Singer focused on six dimensions of psychological well-being: positive relationships, autonomy, personal growth, self-acceptance, purpose in life, and environmental mastery. Singer acknowledged that although psychological well-being is a subjective indicator of health status, he explained how researchers could use his model to appropriately measure the effect of psychological well-being on overall health status, thus using it as an indicator of physical health.

Robert Kaplan, from the University of California at San Diego, suggested that there are two main approaches to measuring health and illness: the psychometric approach and the weighted approach. Currently, the most widely used psychometric measurement instrument is the SF36, which provides measurements of a person's physical limitations and how those limitations may or may not affect how they function physically, socially, and emotionally. The SF36 also measures changes in health status as well as overall well-being, including energy level, incidence of pain, and emotional well-being. The SF36 is considered an easy measure to use, and it provides a quality-of-life profile with the complexity that researchers desire. However, the SF36 cannot be employed easily in an economic analysis and may not be sensitive to minor variations in wellness. Kaplan described a weighted approach to measuring health and illness. In general, life expectancy and traditional survival analysis have been good generic ways to look at population health, but they have limitations. For example, in traditional survival analysis, if a person is alive, a point of credit is given or a score of 1.0 is entered in the computer; if the person is dead, 0 is entered. The difficulty is that survival analysis doesn't make very obvious distinctions; the healthy person scores 1.0, as does the person in a coma. Kaplan proposed an adaptation of traditional survival analysis that accounts for the difference in quality of life and tries to fill in this continuum between optimal functioning and death. Dorothy Rice mentioned, to general assent, that the standardization of whatever measures and definitions we use is extremely important.

Clyde Hertzman, from the University of British Columbia, contended that health is determined in an important way by factors outside of the health care system. To explain, he described an ecological model of the determinants of health that basically conceives of a given society in three levels. At the macro-level is the national socioeconomic environment, consisting of the level of wealth of a society and how resources are distributed. In the meso-level, there is the civil society, which would encompass issues of voluntary associations, questions of institutional responsiveness, neighborhood safety and cohesion, occupational health and safety, and access to organized child care. At the microsystem level is the informal social support network. Intersecting through that is the individual human life cycle, and what we understand about differences in health status between populations in other countries. Hertzman stressed that it would be important for researchers to examine the determinants of health, health status, and health services utilization on these various socio-levels before they could draw any firm conclusions about what is driving the health outcome measures. Using this model, the population health effects could be measured, for example, when welfare policies change, when unemployment insurance changes, or when the unemployment rate changes. Such a health system would have to allow simultaneous analysis of multiple outcomes, because health status has multiple influences, such as health services utilization. Hertzman also suggested that having too many health indicators can have the negative effect of polluting the health information system. Instead, a few good-quality measures that get at the essence of health status are needed.

The need for data on subpopulations, such as children and minorities, was expressed in the workshop discussions. As mentioned earlier, Lorraine Klerman emphasized the need for data on children's health. In fact, she stated that she found it discouraging that in the early part of the workshop discussions the issue of the measurement of children's health had not been mentioned. A question was then posed in the discussions as to how the data for determining the health status of children are different from those required for determining the health status of adults. Klerman responded by pointing out that ignoring the special physical and psychological developmental problems and needs that children and adolescents face could raise serious questions about the validity of conclusions that are reached about this special subpopulation, noting that other subpopulations may also have different subjective notions of what health status means. Olivia Carter-Pokras, from the Office of Minority Health in the Office of the Secretary, U.S. Department of Health and Human Services, followed up this thought by stressing the importance of the system to routinely provide data on ethnic minority groups.

Researchers and policy makers should identify beforehand what the research or policy questions are, before deciding what data will be needed to effectively answer those questions. Similarly, when defining a new vision for a health statistics system, the desired goals of the system should also be identified and be used to generate the appropriate questions to be answered. The concept of what it is we should measure—the definition of health and the measures of health that are appropriate to employ—is an important first consideration in the development of the components of a health information system. With the idea that a proposed goal of a health statistics system is to improve health status, there was much discussion at the workshop on the definition of health and measures of health. Some defined health from a negative consequence perspective, focusing on illness, whereas others want to move away from a deficit model and toward a positive health model that would involve research of how various factors interact to influence the health status and well-being of patients as well as nonpatients. Although there was no real consensus among workshop participants on whether or not to collect data on illness versus measures of good health, participants did agree that there are many dimensions of health that should be collected, and then those data can be used to answer different research and policy questions. However, there should be a reasoned agreement concerning what data are to be gathered together in a national health data system prior to focusing on the issues of data collection and infrastructure.

HEALTH DATA:
WHO ARE THE USERS AND WHAT ARE THEIR NEEDS?

Participants at the workshop noted that the potential users of health data represent numerous areas of the health industry: governments at the local, state, and federal levels; employers who supply health benefits for their workers; insurers; medical care providers; and consumers of health care, to name a few. Government officials at the federal, state, and community levels all need health information to guide them in their public policy decision making. In general, there was a recognition that health data are frequently not given as important a role in public policy decision making as they often merit and that there is a need to interest legislators and elected officials in the value of data collection and evaluation.

Other important nongovernmental players in the health care industry, such as insurers, have their own agendas. Insurance companies seek to run the most cost-efficient program possible. Therefore, they will look for cost-benefit analyses, combined with other measures such as medical care usage. The consumers of medical care usually want to know what the most effective treatments are for their particular needs (e.g., the best treatment for diabetes or cancer). Regardless of the differences in these agendas, each of these sets of decision makers should have data to inform their decisions.

David Fleming suggested that, in many situations, there is a "fundamental disconnect between the data and information side and the policy and program side of our system in the United States." At the workshop, the SCHIP program was cited as an example of the inconsistent use of data-driven information by Congress when considering legislation. Data came from several sources: government reports, such as from the Congressional Budget Office, the Congressional Research Service, and the Government Accounting Office; staff- and member-level support from the Congressional Research Service; congressional committee hearings; and fact sheets and other materials from nongovernment organizations. It was suggested that the problem with these information sources was that they often did not contain timely and data-driven material. For instance, some of the sources did not cite the NCHS's National Health Interview Survey (NHIS), which has included family health coverage questions for several years. The information that was used revolved mainly around cost and access to health care, not whether increased access would result in better health status.

Deficits in health information are present in the private sector, as well. Jacqueline Kosecoff, from Protocare, Inc., discussed the types of data that are needed for her clients, the private sector of the health care industry. Of the various types of data sets that are available (see Table 3-1), each has its advantages and disadvantages. Joseph Newhouse, from Harvard University, made the point that certain national-level surveys, such as the NHIS, Medical Expenditure Panel Survey (MEPS), and Medicare Current Beneficiary Survey (MCBS), are used to monitor big-picture national trends, but they may not be as relevant for state- or community-level needs. This deficit in local and community data underscores the fact that, as many at the workshop noted, there is not presently a single data set that can be used to answer all health policy questions.

An issue that arose repeatedly in the discussion of potential data users was the proper identification of the data needs of these users. One need that appeared to be agreed upon widely by the workshop participants was for data to estimate the effects of programs on health status. As mentioned above, Congress did not, for example, have this information when deciding how to craft an effective child health protection program. A similar lack of appropriate data was evident in the process of developing the Health Security Act, proposed by President Clinton in 1993. Although there were data contributing to the estimates of how many people would be covered by the Health Security Act, how much it would cost the government, whether the cost of health care would increase or decrease, and whether the proposal would affect employment; there were few, if any, data contributing to the question of whether or not the HSA would actually improve the health status of Americans.

There was general agreement among workshop participants on the need for more research to better inform policy makers, but it became a matter of debate at the workshop as to what level of data needs to be collected and reported in order for that research to be useful to the health policy decision makers. Participants suggested that there had been, to some extent, a shift away from the federal government to the state and local levels, and some mentioned that national-level data are not especially relevant to community- and state-level policy decisions even if aggregated by state or region. Indeed, John Lumpkin, from the Illinois Department of Public Health, suggested in his presentation that not even state data are specialized enough for the public health field. Local community-level data are needed, because public health issues generally occur at that level. The presentation by Alonzo Plough, from the Seattle-King County Health Department, emphasized that community-level data are needed to investigate, among other things, disparities in health outcomes that might be present in the local level. Jacqueline Kosecoff suggested that the data collection system as it presently exists in the private sector has turned into a two-class system, in which those that can afford high-quality data, like pharmaceutical companies, have access to good information, while the less fortunate researchers and data users do not. Kosecoff proposed an idea of a population-based claims database managed by a private, publicly funded group, to which all health plans would be required to submit 5 percent of their data annually so that private data would be available to everyone. Though participants seemed to like this idea, some questioned if health plans have enough of the requisite data, particularly in the capitated sector.

Several participants at the workshop argued, without denying the importance of state-level data, that national surveys give valid information that can be generalized to all states or be aggregated to specific settings and populations. These speakers pointed out that states and communities do not have the infrastructure to handle large-scale surveys and that the cost would be too high to create such a system. At a broader level, some workshop participants mentioned international comparisons as an important component of health information analysis. To do this well, we have to collect and coordinate data on an international level, requiring a large infrastructure.

Several workshop participants stated that regardless of the level at which health data are collected, if the information is not gathered and disseminated in a timely manner, it will cease to be relevant to decision makers—not only the government decision makers but also insurers, health care providers, and consumers.

The completeness and coverage of the data are also important factors

TABLE 3-1 Data from the Private Sector

Type of Data	Description	Characteristics
Administrative Data	Routinely submitted billing data (Cost, Dx, Px, Rx, Beneficiary)	- Longitudinal, population-based complete view of care -Available without additional collection - Complete view of all utilization/costs * office visits * referrals * tests * procedures * medications
Enhanced Administrative Data	Administrative data + lab results	- Longitudinal administrative data - Linkage of key laboratory values to patients' diagnoses and services (e.g., hemoglobin, WBC, liver/renal function)
Medical Records	Routine patient charts	- Indepth clinical information, complete patient history (all inpatient, outpatient, R_x, and beneficiary) * physical exam findings * laboratory results * diagnostic tests * surgical interventions * medications - Typically smaller data sets than administrative data

Uses	Limitations
- Quantify cost/utilization - Population level outcomes and quality information	- No information about symptoms, functional status, physical exam, or laboratory findings: can document extent of anti-depressant switching, but not why it occurs (e.g., sexual dysfunction, weight gain) - Capitated payment systems may "hide" certain services patients receive - No capture of OTC medication use
- Answer more clinically sophisticated questions requiring understanding of where drug or disease impacts specific lab values (relationship of hemoglobin A1c to adverse outcomes and drug use) - Show how quickly Lipitor impacts cholesterol compared to its competition	- Difficult to create (payers do not have lab data, providers do not have data from the full "episode" - hospital, ER, specialists, pharmacy) - Unable to answer questions about symptoms or those requiring detailed clinical information
- Track how care is currently provided and changing patterns of care - Document clinical outcomes and adverse events	- Minimal data on patient symptoms, quality of life, or compliance - Variability of documentation ("what you see is what you get") - Hard to find complete data - Not standardized (can't get data you need on all patients collected in same way)

continued on next page

TABLE 3-1 Continued

Type of Data	Description	Characteristics
Patient "Living" Registries	Ideal data set	- Collect specific data in a standardized way * trained data collectors * richer and more accurate than medical records * can readily revise data collected - Ability to continuously track patients across providers and insurance plans - Simulates clinical trials quality data

SOURCE: Data from Jacqueline Kosecoff, Protocare, Inc. (personal communication)

in using data for decision making. For example, as many workshop participants suggested, data about subpopulations, including children with special problems and ethnic minority groups, are needed to give an accurate depiction of the health status and health care needs of all Americans. Furthermore, the lack of longitudinal data can call into question the validity of health care usage and expenditure decisions. Several workshop participants recommended conducting large-scale longitudinal studies composed of several age cohorts, thus controlling for historical and cohort differences. It was noted that that there have been longitudinal studies in the past, including the NHANES I Epidemiologic Followup Study and the Longitudinal Study on Aging, but those were conducted on a relatively short-term basis. Longer series of observation would be more useful.

Dorothy Rice also made the point that, although the quality of government-based data has continually improved, private-sector health data have been erratic. Sampling, reporting, processing, and nonresponse errors plague private-sector data and therefore "the improvement of the quality and reliability of health statistics in the private sector is most urgently needed." Moreover, peer-reviewed studies are not immune from drawing conclusions from possibly invalid data. In a review of studies dealing with managed care and quality of care, Miller and Luft (1997) found that many of those studies used data that were several years old, based on a relatively

Uses	Limitations
- Document drug's impact on symptoms, work loss, and cost of care - Determine which symptoms and QOL issues matter most to patients (target DTC advertising) - Assess perceived and actual side effects of drugs on an ongoing basis - Help plan clinical trials - Answer fundamental questions about diseases and therapies - Relate findings back to other data sets	

small sample size, and based on "ad hoc data collection." In addition, participants suggested that the quality of insurance coverage and health care can vary greatly, so that simple utilization rates of health care do not necessarily provide a good indicator of health status. Although some of these data can be audited, some cannot, and therefore it is important to provide adequate incentives to those who are collecting data on all levels.

In summary, many different decision makers need health data, e.g., government officials, legislators, insurers, health care providers, and consumers. Simply providing data to these groups of decision makers is not enough, however. The data must be relevant to the questions being asked and must be given in a timely manner. The relevance of data is often affected by the level at which data are collected. Meeting the needs of all users is, in fact, a gigantic task, and one in which the data collection and integration procedures are very important.

HEALTH DATA: COLLECTION AND INTEGRATION

Two schools of thought emerged from the workshop concerning the collection of health data. One group of participants felt that there are major gaps in the present federal, state, and private sector data sets and that new data collection methods and strategies are needed to fill these gaps.

The second group agreed that there are gaps in the data, but believed that the answers did not necessarily lie in developing new data collection activities. This second group felt that by examining the data that already exist and consolidating and integrating existing data one could compensate for most of what is lacking.

There certainly is currently a great deal of health data in the public domain. Several federal agencies and systems take part in epidemiological and other studies that accumulate health-related data, many of them in the form of surveys. The National Institute of Occupational Safety and Health, the Food and Drug Administration, the National Cancer Institute, the Centers for Disease Control, and the Consumer Product Safety Commission are just a few of the agencies that conduct national surveys. According to the presentations and comments of workshop participants, it was clear that via systems such as these, information is gathered on a broad range of health topics. For example, the National Immunization Survey provides data on the immunization rates of children 19-35 months of age; the Medical Expenditure Panel Survey provides data on health status; the National Household Survey on Drug Abuse provides data on the incidence and prevalence of drug use; the National Health Interview Survey provides national data on noninstitutional samples for acute illness, accidental injury, illness prevalence, health utilization; and the National Health and Nutrition Examination Survey provides national data obtained from physical examinations and physiologic and biomedical measurements. When you take into account all of the available sources that provide health data, you have an imposing collection of health data at the federal level (see Table 3-2). Furthermore, the federal government is not the only entity conducting these types of data collection. The private sector, including corporations, consulting firms, and hospitals, also conduct their own surveys.

Although having such a wealth of varied information might seem to be a positive outcome, Dorothy Rice along with several other workshop participants expressed concern that the overlap and duplication of data results in inefficiency with regard to cost and information. She stated, “Despite improvements, health statistics production in this country presents a picture of fragmented data collection, lack of common definitions, duplicative and overlapping systems, and resistance to data sharing.” Several participants at the workshop referred to this model of data collection and storage as creating a collection of separate silos, each housing fragmented statistical data and not easily relating to the others. The problem at the federal level, at least, was well characterized by Janet Norwood, who said, “The important thing, I think, is that there is not a single place in the United States government to which one can turn for a definitive answer [to questions concerning health statistics] and explanation of the different data sets."

Dorothy Rice reported on a program within DHHS created as part of the Reinventing Government, Part II, initiative, that resulted in the formation of an interagency working group whose purpose was to explore ways to consolidate various DHHS surveys into a single framework. As a result of this activity, for example, MEPS, NHANES, and the National Survey of Family Growth (NSFG) are all now using the sampling frame of the NHIS. Although this represents a major step forward in reducing overlap and increasing efficiency in DHHS data systems, according to Dr. Rice, there is still a long way to go before fragmentation and overlap are eliminated. Several workshop participants echoed these sentiments. H.R. 2885, the Statistical Efficiency Act, currently under consideration in Congress, would be another step in increasing efficiency in federal statistical systems. This act, if passed by the Congress, would designate eight federal statistical agencies as data centers and allow for limited sharing of that data. Janet Norwood cited another step toward data consolidation and coordination taken by DHHS—the creation of a Data Council reporting to the Office of the Secretary. Jennifer Zelmer, from the Canadian Institute for Health Information, pointed out that many countries (e.g., Canada, England, and Australia) regularly perform analyses that "take advantage of those data that can already be integrated consistently." For example, the population-based information systems known as POPULIS employed in Manitoba and described in the presentation by Charlyn Black demonstrates the advantage of considering the possibilities for data linkages when planning population based health surveys.

Private sector databases generally revolve around health care information obtained from administrative data, medical records, and patient registries. Jacqueline Kosecoff cited the need for enhanced administrative data in the private sector. An example would be data that incorporate lab results with administrative data and allow for longitudinal analyses and linkages of key lab data to patients' diagnoses and services, as a means of attaining more efficient information collection and the ability to answer more clinically sophisticated questions. For instance, can using lab results from diabetics answer such questions as whether the diabetic is healthier in a capitated plan, or healthier when treated by a diabetologist versus a family physician, or whether the age of the physician impacts care, or whether female diabetics do better with female physicians? Kosecoff and others at

TABLE 3-2 Data Sources for Health Indicators Included in the Candidate Sets

Health Indicator	Data Source	Level of Availability
Mortality		
Infant	Vital Statistics	Local, State, National
Maternal	Vital Statistics	Local, State, National
Motor vehicle crash	Vital Statistics, FARS	Local, State, National
Alcohol-related motor vehicle crash	FARS	State, National
Work injury	CFOI	State, National
Suicide	Vital Statistics	Local, State, National
Homicide	Vital Statistics	Local, State, National
Firearm fatality	Vital Statistics	Local, State, National
Lung cancer	Vital Statistics	Local, State, National
Breast cancer	Vital Statistics	Local, State, National
Cardiovascular disease	Vital Statistics	Local, State, National
Stroke	Vital Statistics	Local, State, National
Diabetes	Vital Statistics	Local, State, National
Unintentional injury	Vital Statistics	Local, State, National
Residential fire	Vital Statistics	Local, State, National
Morbidity		
HIV	NNDSS	Local, State, National
AIDS	NNDSS	Local, State, National
TB	NNDSS	Local, State, National
Measles	NNDSS	Local, State, National
Syphilis	NNDSS	Local, State, National
Gonorrhea	NNDSS	Local, State, National
Hypertension	BRFSS, NHANES, NHIS	State, National
Hypercholesterolemia	BRFSS, NHANES, NHIS	State, National
End-stage renal disease	HCFA	State, National
Asthma hospitalization	NHDS	Some State, National
Cumulative trauma disorders	ASOII	Some State, National
Depression	NCS, ECAS	National
Reported disability	BRFSS, NHIS	State, National
Hospital days/100,000	NHIS	Some State, National
Years potential life lost	Vital Statistics	Local, State, National
Emerging infectious diseases	NNDSS	State, National
Food/water-borne diseases	NNDSS	State, National
Hospital admissions	NHDS	Local, State, National

TABLE 3-2 Continued

Health Indicator	Data Source	Level of Availability
Service Delivery		
Childhood immunizations	NIS, NHIS	MSA, State, National
Pneumonia/flu immunization	BRFSS, NHIS	State, National
Cervical cancer screening	BRFSS, NHIS	State, National
Mammography	BRFSS, NHIS	State, National
Preventive services delivery	PCPS	National
Primary care linkage	PCPS	National
Other Risk Conditions and Factors		
Low birth weight incidence	Vital Statistics	Local, State, National
Teen intercourse	NSFG,YRBS	National
Teen pregnancy	Vital Statistics, NSFG	Local, State, National
Teen births	Vital Statistics	Local, State, National
Condom use	NSFG,YRBS	National
First trimester prenatal care	Vital Statistics	Local, State, National
Breastfeeding	Ross Labs, NSFG	State, National
Cigarette smoking/sales	NHSDA/NHIS/YRBS/MFS	National
Smokeless tobacco	NHSDA/NHIS/YRBS/MFS	National
Alcohol misuse/emergency room visits	NHSDA/NHIS/YRBS/MFS	National
Illicit drug use/emergency room visits	NHSDA/NHIS/YRBS/MFS	National
Seatbelt use	NHIS	State, National
Firearm storage	BRFSS, NHIS	State, National
Overweight	BRFSS, NHANES	State, National
Sedentary pattern	BRFSS, NHANES, NHIS	State, National
Untreated dental caries	NHANES	National
Air quality exposure	AIRS	(Non-attainment areas)
Health insurance/loss	NHIS, Census, MEPS	State, National
High school graduation rate	NCES	State, National
Childhood poverty	Census	State, National

continued on next page

TABLE 3-2 Continued

Key:
AIRS = National Air Quality and Emissions Trends Report/Aerometric Information Retrieval System
ASOII = Annual Survey of Occupational Injuries and Illnesses
BRFSS = Behavioral Risk Factor Surveillance System
CFOI = Census of Fatal Occupational Injury
ECAS = Epidemiologic Catchment Area Study
FARS = Fatality and Analysis Reporting System
HCFA = Health Care Financing Administration
MCBS = Medicare Current Beneficiary Survey
MEPS = Medical Expenditure Panel Survey
MFS = Monitoring the Future Study (University of Michigan)
NCES = National Center for Educational Statistics, Department of Education
NCS = National Comorbidity Survey
NHANES = National Health and Nutrition Examination Survey
NHDS = National Hospital Discharge Survey
NHIS = National Health Interview Survey
NHSDA = National Household Survey on Drug Abuse
NIS = National Immunization Survey
NNDSS = National Notifiable Disease Surveillance System
NSFG = National Survey of Family Growth
NSWHPA = National Survey of Worksite Health Promotion Activities
PCPS = Primary Care Provider Survey
YRBS = Youth Risk Behavior Survey

SOURCE: Department of Health and Human Services (1998).

the workshop encouraged the development of a greater connection between federal agencies and the private sector, so that each sector could have additional, useful information that it would be unable to collect on its own, and so that costly overlap among the public and private sectors could be reduced.

Discussion of linkage of government and private sector resources highlights an interesting question: At what level or levels should linkages of health data take place? Interagency linkage and consolidation is in its infancy in the United States and is used considerably more in other nations. The private/public partnership that several workshop attendees mentioned is another level of linking data, as is linkage among community, state, and federal agencies, not just interagency connections on the federal level. By linking data at all three levels of governmental organization, a more complete picture can be had and less fragmentation of health data will result. Clyde Hertzman emphasized the importance of data integration at the international level. The power and diversity of the data that would result from international partnerships would be valuable. Hertzman also suggested that an important potential partnership is that between universities and government agencies, such that sample data collected in university settings and broad, population data collected in agency settings could be effectively integrated.

According to several workshop participants, a variety of problems can arise with data partnerships. There may be concern that one organization or agency will control the data in opposition to the desires of the other data partners. As an example, many workshop participants mentioned that state-level government agencies might be—or are—hesitant to share data with federal agencies because of fear that the federal agencies will hold all the data and impede the dissemination process. Another major obstacle to integration and linkage of data cited was the lack of standardization in data collection among the levels of data collection mentioned above. Many workshop participants agreed that some sort of standardization in data collection is needed in terms of procedure, definitions of terms, and populations sampled. Some argued that the diversity resulting from not having such standardization can be beneficial, but acknowledged that lack of standardization leads to many inefficiencies. The differing responsibilities of agencies can also inhibit data collaboration. Dr. Kenneth Thorpe from Emory University stated, "It is not a tremendous surprise that, when you see the proliferation of data sets and overlap of data systems, these are often driven by the agency's differing set of responsibilities." Not discusssed at the workshop was the topic of legal restrictions that apply to the flow of certain data away from the government and that go hand in hand with the government's legal authority to collect particular data.

New technologies to facilitate data collection are continuously emerging. As William Eddy from Carnegie Mellon University stated in his presentation, "Communication is the key." New data collection procedures continually improve the possibilities for gathering data from hard-to-reach populations and hard-to-measure variables, and provide opportunities for improving response rates. Dorothy Rice provided a brief overview of the changes that have occurred over the years in regard to data collection. From in-home interviews to random-digit telephone dialing, computer-assisted telephone surveys, and computer-assisted interviews, the progress of technology in general has had positive effects on health statistics. With the increase in use of the Internet, further improvements in data collection (e.g., Web-based surveys) are probably not far behind. Participants cautioned, however, that there is always the possibility that more advanced technology may not produce the most valid information. For instance, Lorraine Klerman mentioned that experts tend to agree that household interviews are more valid than surveys conducted by telephone or questionnaire, but questioned if there would be any difference when using the Internet. An added benefit of increased computer technology is, of course, the ease with which data can be shared. Websites can be designed to disseminate timely data analyses rapidly. Complex databases could be formed that consolidate several different surveys and automatically remove any overlapping or duplicative data. Workshop participants noted that a major consideration in the use of technology for these purposes will be the ability of the architects of the system to assure that the privacy and confidentiality of the data in the system are appropriately protected at all times.

PRIVACY AND CONFIDENTIALITY

The protection of privacy and the assurance of confidentiality are important for both the individual and the data collector in any data system. In addition to the obvious ethical and legal responsibilities on the part of the data collector for the assurance of privacy and confidentiality, a lack of secure feelings among the survey participants regarding their privacy and confidentiality can adversely affect the data collection process. As many workshop participants mentioned, obtaining sensitive information is not an easy task. When there is a distrust among survey participants concerning how their information will be used by federal, state, and local governments or by private organizations, the data collection process will inevitably become more difficult and incomplete. Dorothy Rice summarized these difficulties by stating that “both individuals and businesses are questioning how the information is used and who has access to it. At the same time, data users, especially those outside of government, are increasingly frustrated by limits on the amount of detailed information they can obtain from statistical agencies” (see Chapter 2). Not only might survey participants be reluctant to provide accurate data if they feel their privacy would be compromised, but, more important to the health of each individual, they might not cooperate in a treatment setting for fear that information obtained would be made available to others.

The ethical background of health care professionals plays an important role in protecting the privacy of data. For example, workshop participants pointed to the Hippocratic oath (i.e., "Whatever in connection with my professional practice or not in connection with it I may see or hear in the lives of my patients which ought not be spoke abroad I will not divulge, reckoning that such should be kept secret") or to the American Medical Association's Principles of Medical Ethics (i.e., "A physician shall respect the rights of patients, of colleagues, and of other health professionals, and shall safeguard patient confidences within the constraints of the law"). However, not all health researchers, of course, are physicians, and although several professional organizations, such as the American Psychological Association, have their own standards of ethics, many do not. Furthermore, George Duncan of Carnegie Mellon University pointed out in his remarks these two statements of ethical standards are based on the assumption that the paradigm will be one of physician and patient only. These standards are not as relevant to complex systems in which multiple data sources containing information about patients outside any one doctor's domain are available via various communication streams. Duncan suggested that a more relevant ethical standard to have would be "autonomy and respect for the individual patients." Whatever the standards, though, the increase in information technology makes data theft a more plausible outcome. Linkages from various databases, easy dissemination using the Internet and other communications technology, and the ease of finding information through user-friendly search engines exacerbate the potential problems.

The storage of data can be made more secure by technological innovations. Just as improved technology creates the potential of nonsecured data sets being broadcast over various communication lines, other technology is being developed to help prevent security breaches. Federal legislation is aimed at insuring that privacy and confidentiality are protected. John Eisenberg, Director of the Agency for Healthcare Research and Quality (formerly known as the Agency for Health Care Policy and Research), reported to the workshop attendees that the DHHS has been developing privacy regulations in connection with the Health Insurance Portability and Accountability Act (HIPAA) legislation.[1] These regulations would,

---

[1] The Health Insurance Portability and Accountability Act of 1996, Public Law 104-191, was signed into law on August 21, 1996. Having its roots in the 1993 Clinton health care reform proposals, the primary intent of HIPAA is to provide better access to health insurance, to limit fraud and abuse, and to reduce administrative costs. The Administrative Simplification aspect of that law requires DHHS to develop standards and requirements for maintenance and transmission of health information that identifies individual patients.

among other things, govern the electronic transfer of data used in federal health programs, such as Medicare. George Duncan presented several ideas regarding potentially useful methods for the protection of the privacy and confidentiality of research participants. One method would be to limit the access to and disclosure of the data. A consensus on the appropriate limits to access and disclosure is often very difficult to reach, however, given the potentially conflicting objectives of the interested parties. For example, J. Michael Fitzmaurice, from the Agency for Healthcare Research and Quality, suggested that a possible problem with the HIPAA legislation might be overly strict rules that would have the unwanted effect of keeping important data away from ethical researchers. Furthermore, Jacqueline Kosecoff suggested that in many circumstances health plans need to submit data that are scrambled and "scrambled data are really hard to link with previously scrambled data." John Eisenberg commented that he considered the proposed regulations to be "a terrific document. It really does bring together control and assurance to the public that their data is being held confidential, and yet a recognition that there are certain goods that we have to keep in mind."

George Duncan summarized the problem of protecting the privacy and confidentiality of health data by stating that the goal is to "satisfy the customer and deter the shoplifter." Full data access would result in high-disclosure risk and go above the maximum level of tolerance that an individual would allow (see Figure 3-2). The managers of the data system "have to provide access to the data. At the same time, they have to provide confidentiality protection. If one of these pillars disappears, the whole edifice collapses. That is the framework in which we operate."

Confidentiality can be protected in ways other than by simply not allowing people to access the information.[2] For instance, data masking can be employed to make the data anonymous. This can be accomplished by

---

[2] In October 1999, the Committee on National Statistics convened a workshop that reviewed the benefits and risks of providing public-use research data files and explored alternative procedures for restricting access to such data, particularly longitudinal survey data that have been linked to administrative records. Please see the summary report of that workshop, in which the tradeoffs between researcher and other data user needs and confidentiality requirements are well articulated, as are the relative advantages and costs of data perturbation techniques versus restricted (physical) access as tools for improving security (National Research Council, 2000).

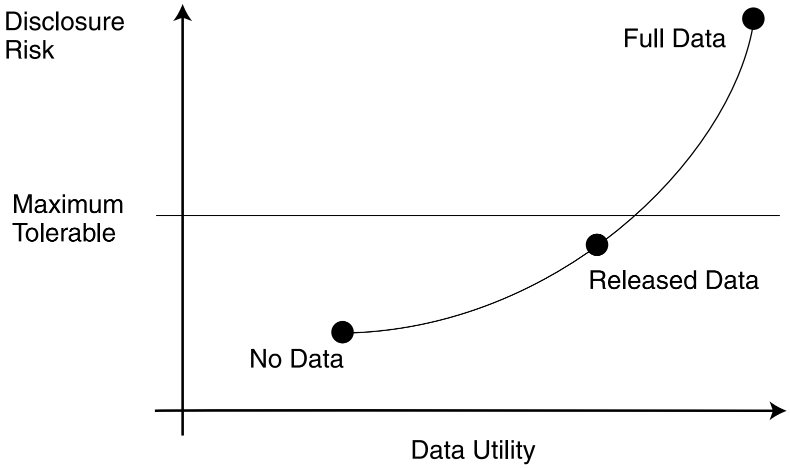

FIGURE 3-2 Data disclosure risk and individual tolerance level.
SOURCE: Data from George Duncan (personal communication).

removing identifiers, restricting reporting of cases and variables, systematically altering the data, and matrix masking. Matrix masking involves transforming the data by recoding or releasing only subsets of the data, but leaving the essence of the data in tact. One of the difficulties with this approach is that anonymity is not always assured by simply removing identifiers or variables from the data sets. The use of synthetic data provides another option for restricting access. In this process, the key idea is that descriptive models (usually parametric) are estimated from the data, and then samples are generated from the model. The main problems of this method are the complexity of putting it into practice and the fact that there is no direct link between the original data and the data that are eventually released. However, the essence of the information remains, and attaching data to individual patients is nearly impossible.

In summary, issues of confidentiality are important for both patients and other individuals who supply the data and for health information analysts, decision makers, and others who use the data. If either set of interests is not appropriately taken into account, then the health information system will be filled with frustrations and inefficiencies. Steps have to be taken to ensure that privacy and confidentiality will remain intact while making the data accessible for legitimate uses. Approaches such as using masked or synthetic data, restricting access to a particular database, or requiring passwords for Web-based access are examples of the methods that can be used, either solely or in combination, to protect confidentiality.

HEALTH DATA SYSTEMS: LESSONS FROM OTHER COUNTRIES

When reviewing the health systems in other countries, some of the characteristics and methodologies regarding the collection and utilization of health data could serve as examples for developing an improved system in the United States. Several presenters at the workshop discussed health information systems that are being implemented in other countries, and suggested that these systems might be viewed as potential models for the American system. Each of these systems featured as core components a central coordinating body, linkage of data sources, and standardization of data and data collection.

Charlyn Black described the POPULIS program, a population-based health information system housed in the Manitoba Centre for Health Policy and Evaluation, which is a university-based research group. The purpose of POPULIS is to provide "accurate and timely information to managers, decision-makers, and providers, in order to support them in offering health care services…." Through its system of data sets, organization, and analysis models, POPULIS is designed to allow evaluation of the components that "influence health, the state of health, and the availability and utilization of health and other services," thus enabling researchers to address complex questions. Another important feature of using a population-based system is that it "enables the user to simultaneously relate characteristics that affect a population's need for health care to that population's use of health care, to that area's supply of health care resources, and finally, to the health status of a population." Black suggested that a prerequisite of a strong health information system is a strong health data system. The data system used by POPULIS is a population-based research registry that captures data provided by the Manitoba provincial health insurance administration system. Black pointed out that a data system is not sufficient by itself, however, to produce a quality information system. An appropriate conceptual model for analysis of the data is very important, as well. The model the POPULIS program employs recognizes that there are a multitude of factors that can contribute to the health status of the population. Health status can, in turn, have its own impact on each of those factors—for example, poor health status can affect socioeconomic status, and vice versa. This model allows for "rich and complex population-based analysis." POPULIS uses data obtained from the administration and payment for services delivered to estimate the key concepts under investigation such as health status, health care utilization, and supply. Black noted, "there is a unique identifier within the research registry to maintain privacy and confidentiality. It is scrambled, it is anonymized, and we maintain a parallel and different set of data." Data from many files, such as hospital data, data from the medical community-based services, pharmaceutical data, vital statistics, and aggregate census data, exist in an unlinked format, and the use of the unique identifier allows for very rich microdata at the level of individualized, anonymized people. Black emphasized that when designing population-based health surveys, the possibility for linking data and conducting validity studies using administrative data should be foreseen and incorporated into study design. In Manitoba, with the expansion of national, longitudinal, population health services, there has been explicit consideration given to the potential to link detailed data from surveys to the population-based study that comes from the administrative registry information. Through the work of the POPULIS program, Black added, they have been able to transform an administrative medical care data system into an information system that focuses on the population's health, use of the health care system, and cost of the health care system. The system is much more responsive and useful in anticipating policy needs, monitoring trends, and understanding connections, and has been instrumental in changing the focus of policy discussions from what health care services are provided to what is being done to improve health.

Jennifer Zelmer presented a summary of the health information systems that have been created in Australia, Canada, France, Finland, England, and Denmark. She spoke of tenets that have been followed in each country's efforts. A common theme in each case is that the activities are coordinated by an independent agency that is responsible not only for coordinating the statistical activities at the national level but also for coordinating the health information. For instance, in Australia, the task of oversight has been given to the Institute of Health and Welfare; in Canada, this responsibility is handled by the Institute for Health Information; in France, it is the Direction de la recherche, des études, de l'évaluation, et des statistiques; and, in Finland, the efforts are coordinated by the National Research and Development Centre for Health and Welfare. These coordinating bodies do not act alone, but actively collaborate with federal, provincial, and territorial governments, and other relevant decision makers. In Canada, for instance, Statistics Canada is an important partner in the health information process. The Danish, Canadian, and Finnish systems integrate microlevel data across time, which allows for complex analyses. More localized efforts can be seen in Canada, where several provincial research institutes have collaborated in order to bring together a wide range of data. Collaborations also exist on the international level. For instance, the European office of the World Health Organization has acted as the coordinating body to bring together national and subnational statistics in that region in order to facilitate international analysis. Other collaborations include the European and international standardization committees, joint health statistics meetings between the World Health Organization and the European Economic Community (EEC), and several initiatives that fall under the Global Healthcare Applications Project.

Efforts to facilitate the collection and linkage of administrative and respondent data are continuously being explored and the information technology continuously upgraded in most developed countries. Data integration and linkage, of course, carry with them a risk of infringement on individuals' privacy. Finland has attempted to solve this problem by encryption of identifiers and data integration by a third party. Switzerland uses anonymous linkage codes for each hospital patient and destroys the data after 10 years. As in the United States, many countries have implemented specific legislation to protect the rights to data privacy and confidentiality. The legislation is generally modeled on the principles outlined by the Organization for Economic Cooperation and Development. However, with the progression of information technology, countries have realized that more should be done in this area and have sought to develop innovative technologies to aid in the protection of privacy.

Merwyn Greenlick, from the Oregon Health Sciences University, spoke about his Utopian view of what an American health information system should look like in the twenty-first century. Termed a humanistic health care system, Greenlick's ideal would link each individual to the health care system based on the individual's needs, desires, aspirations, risks, disease condition, and health functional status. In order to meet this need, the health care system must ensure a sense of trust, overcome knowledge deficits, and have a powerful reimbursement method in place. Greenlick pointed to Kaiser Permanente as an example of an organization that has developed a micro image of what a system should possess. Using innovative information technologies, doctors can communicate using computers to refer, schedule, and even treat patients. Patients could even be given information that they can use to help treat themselves. Greenlick stated that this would place a piece of the treatment of decision making into the hands of the patient. This type of system would also allow for doctors to practice within a population-based clinical practice. Eileen Peterson, from United Health Care, a private-sector health care organization, proposed a private managed health care data system for the United States. The system would be a managed care data repository and reporting system based on administrative data that would combine financing and delivery data. Ideally, in such a system, an electronic medical record would be linked into a practice management system in a provider's office, and with the press of a button, a claim is sent and a medical record is created. Other routinely collected data would be incorporated into the system, with surveys (e.g., the Consumer Assessment of Health Plans and the Health of Seniors), disease management, contracts databases, and support databases adding important information. Mining the existing data for all relevant information would be a good cost-efficient procedure. An important step of this process would be the one in which the data stakeholders convene "to define the rules of the road" (i.e., who can and cannot have access) and to explore and map linkages to data resources. However, Peterson emphasized that before any of these goals and strategies are discussed, the stakeholders would need to define the products to be produced. Once these steps are taken, Peterson said that it would be necessary to demonstrate the capabilities of the system through monitoring and surveillance to assure there is a quick turn-around of data to inform policy and to demonstrate that a full-scale research project can be undertaken. An evaluation of the opportunity costs—both monetary and nonmonetary, such as the improvements in health—would be essential to determine the "return on capital" to consumers and the health industry. Although integrating such a system with the public-sector systems would bring about many challenges, Peterson reported that new technologies present an exciting opportunity to do just that. Dorothy Rice echoed this sentiment when she said that the private sector should be an important contributor to future health information systems. Jacqueline Kosecoff stated that it is not necessarily optimal to have all the data that a particular hospital or health care provider has collected. Instead, it is important to be able to track all care "delivered by a cohort of providers." Kenneth Thorpe, along with several others, suggested that the federal government, more specifically the executive branch, should take the lead in coordinating such a system. This is the case in part, according to Robert

Murphy from Westat, because the federal government has a more adequate infrastructure than that available at the state and local levels. This sentiment was echoed by Miron Straf from the National Research Council, who spoke of a vision in which local and state public health communities would take advantage of the infrastructure of NCHS. David Fleming, from the Oregon Health Division, and others questioned whether federal agencies such as NCHS can, in fact, provide the necessary leadership to state and local agencies.

In summary, a number of countries have strong health information systems. Canada, with several collaborations among local and national organizations, Australia, Finland, France, and several others, have central coordinating bodies for their health information systems. Researchers in these countries believe they are able to successfully obtain data from a broad range of sources and across a wide variety of variables, yet are still able to maintain a level of privacy and confidentiality that has helped to ensure trust among their public. Several workshop participants praised the efforts of these national systems and asked if it was possible to build such a system in the United States. The question was raised whether a government or nongovernment entity should act as the coordinator if the United States were to build such a system. Jacqueline Kosecoff suggested that whichever organization could make data available cheaply and efficiently should take the lead regardless of whether it is a public- or private-sector organization. She continued by saying that researchers should stop arguing over who will house the data sources and worry more about getting a system in place. Daniel Friedman, from the Massachusetts Department of Public Health, echoed the comments of several at the workshop when he said that the United States can achieve what Canada has done and that the societal differences between the countries should not be seen as an insurmountable road block. He suggested that the health information community needs to start thinking that the United States can succeed in creating a comprehensive and effective health information system and that it is time to decide how to get it done.

FINANCING A HEALTH INFORMATION SYSTEM

Dorothy Rice reported that federal spending for health and human services statistics had reached $804.1 million in 1999. Despite the size of this investment, several workshop participants either suggested that more money is needed in U.S. data systems or wondered how a new and innovative health information system could be financed in the present political climate (the fall of 1999). Rice commented that states need help and that the federal government needs to do more to take a leading role. The central governments in countries with successful health information systems appear to have provided greater financial support to these systems. Jennifer Zelmer reported, for instance, that the government in the United Kingdom has promised to contribute £1 billion of new money to their health information system over a seven-year period, while the Canadian government has estimated that they will spend over $1.5 billion on information technology for the health field in the year 2000, with the majority of the money going to support the infrastructure at the provincial, territorial, and local levels.

An issue of financial importance to the private sector is the appropriate compensation of medical care personnel for taking the time and resources to supply agencies and organizations with patient data. Elliot Stone wondered if the Health Care Financing Administration reimbursement would be the answer for those who follow national standards of data collection. Jacqueline Kosecoff stated that she would not want to have an agency, such as a federal body, tell her what and how to collect data unless she was compensated for that work, further noting, however, that she would rather see a system of tax credits than a system of reimbursement. There was a sentiment among participants that more efficient spending of available funds could be at least a partial answer to financing a health information system. Kenneth Thorpe, for example, noted that the federal government is spending hundreds of millions of dollars on health data and that private foundations are doing the same thing, but with much overlap among disparate surveys. Consolidation and integration might be a way to defragment the data collection process.

REFERENCES

Black, C., N. Roos, and L. Roos
1999 From Health Statistics to Health Information Systems: A New Path for the 21st Century. Paper commissioned by the National Committee on Vital and Health Statistics and presented at the Committee on National Statistics Workshop, Toward a Health Statistics System for the 21st Century, November 4, Washington, DC. Available: <http://www.ncvhs.hhs.gov/hsvision/visiondocuments.html> [July 12, 2001].

Evans, R.G., and G.L. Stoddard
1990 Producing health, consuming health care. *Social Science and Medicine* 31:1347-1363.

Miller, R.H., and H.S. Luft
1997 Does Managed Care Lead to Better or Worse Quality of Care? *Health Affairs* 16:7–25.

National Research Council
2000 *Improving Access to and Confidentiality of Research Data: Report of a Workshop*. Committee on National Statistics, Christopher Mackie and Norman Bradburn, eds. Washington, DC: National Academy Press.

U.S. Department of Health and Human Services
1998 Leading Indicators for Healthy People 2010. A Report from the Working Group on Sentinel Objectives. Office of Disease Prevention and Health Promotion, U.S. Department of Health and Human Services. Available <http://odphp.osophs.dhhs.gov/pubs/LeadingIndicators/ldgindtoc.html> [July 12, 2001].

4

Some New Directions

William Kalsbeek, Member, Committee on National Statistics, and
Edward Perrin, Workshop Chair

It is clear from the presented papers and ensuing discussion at this workshop that the nation's health statistics system, while reasonably comprehensive in content, is unable to fully meet the nation's current and anticipated health information needs. It is also apparent from the workshop discussions that existing health information initiatives elsewhere in the world may provide a springboard of ideas for improvement.

Comments from workshop participants suggest that dealing with the current system's shortcomings will require more major redirection than minor incremental changes in the system's organization and function. Indeed, the workshop discussions point to a search for innovative ways to gather health data and to transform it into useful health information, not just shoring up existing components of the system. Participant comments also seem to endorse the encouraging trend toward interagency cooperation that has expanded the utility of several health data systems in the past 10 years.

Many important shortcomings of the current health data system emerged in the course of the workshop. Some of the more provocative shortcomings mentioned by attendees (named in parentheses) were the following:

- Content overlap among existing data generating sources (Norwood, Rice)

- Uneven access to proprietary data generated by the private sector of the health system (Kosecoff)
- Limited standardization in reported data items among data sources (Kosecoff, Rice, Wolfson)
- Failure to broadly disseminate health information to public health professionals (Kronick)
- Limits in the amount of data sharing among federal statistical agencies as a way to broaden the utility of the data they generate individually (Rice)
- The absence of broadly focused coordination and vision in the gathering and dissemination of health data (Norwood, Rice, Wolfson)
- An insufficient capacity to generate needed information at the state and local levels, as well as for other small but important subgroups of the U.S. population, such as young children (Klerman, Rice, Sondik, Wolfson)
- Limited ability to longitudinally link disease prevention and treatment to health outcomes in the general population (Kosecoff, Kronick, Rice, Sondik, Wolfson)
- Several notable content gaps in the present health statistics system including:

—health system performance (Kronick)

—health-related quality of life in general (Kronick), and in children in particular (Klerman)

—indicators of functional capacity and other indicators of "positive health" in children (Klerman)

—health assessment for children who are disabled, institutionalized, or possessing other special problems (Klerman)

Papers and discussion from this workshop seem to suggest that some form of integrated planning is needed to improve the usefulness of the current decentralized system. There are at least two possible initiatives that could address important gaps in the nation's health statistics system. One is the creation of a separate, broadly focused, and semi-independent *health information planning board* with the responsibility and authority to provide overall focus, direction, and coordination to the national health data system. Patterned after the Australian Institute of Health and Welfare, the Canadian Institute for Health Information, and comparable panels in other countries, such a board might serve a central role in forging an ongoing partnership of health providers and insurers, as well as public and private users and producers of health data. The board's main function could be to orchestrate and broker the production and dissemination of population-based health data in the United States. Some of its major specific aims might therefore be to:

1. identify and organize information needs at all levels of the health care system;
2. establish a working collective partnership with all producer agencies that contribute to the health data system;
3. negotiate a clear set of information gathering roles for each member of the partnership;
4. develop and implement a long-term plan to promote the collection of standardized, high-quality health data;
5. develop and implement innovative ways to gather needed health data;
6. promote the privacy and confidentiality of all publicly acquired data, and explore ways to defeat accompanying threats;
7. support the development of new methods to effectively use collected data to answer important health-related questions being asked by the health community at all levels;
8. determine who decides the content and standards for health indicators for national health and health systems performance comparisons;
9. promote access to all existing health data by public and private researchers and policy makers; and
10. explore new outlets and uses for health information based on current information technology.

At first glance, some of the goals of this proposed planning board may seem to overlap with the goals of two existing advisory bodies, namely, the U.S. National Committee on Vital and Health Statistics (NCVHS) and the Department of Health and Human Services (DHHS) Data Council. It is important to note that there are clear distinctions in the roles. NCVHS serves as the statutory public advisory body to the Secretary of Health and Human Services in the area of health data and statistics. In that capacity, NCVHS provides advice and assistance to DHHS and serves as a forum for interaction with interested private sector groups on a variety of key health data issues. NCVHS is composed of individuals from the private sector. The DHHS Data Council coordinates all health and nonhealth data collection and analysis activities of DHHS, including an integrated data collection strategy, coordination of health data standards and health information and privacy policy activities. The Data Council consists of all assistant secretary and agency administrator-level DHHS officials who have a direct reporting relationship to the secretary. The proposed planning board differs from these two bodies in many ways. First, the planning board would be "semi-independent," meaning it would be a partially public body, not solely an internal federal body. Second, the planning board would be composed of people from both outside and inside the federal government, but not federal employees who could be affected by the recommendations of the board. Finally, the planning board would ideally operate at an integrated national system level, and thus could identify, suggest, and coordinate plans for all data producers, and not be limited to impacting the programs and plans of DHHS.

Another new initiative suggested by this list of shortcomings of the current system is the development of a *comprehensive national longitudinal health study* of significant magnitude to better understand biological, behavioral, treatment, and policy effects on whole-life health outcomes. This might be accomplished largely by applying existing health data gathering apparatuses to a general population sample based on a rolling cohort of births. A "birth" in this context might be interpreted initially as participation in a population-based household sample, and then subsequently to natural birth and immigration. The sample of "births" could be followed continuously until "death" occurs naturally or by migration to another country. After gathering baseline data this rolling cohort sample could be followed up to obtain a full range of needed data tied to health-related attitudes, behavior, and provider care. Besides continuously monitoring the health experience of cohort members and their families, linkable health services data could also be continuously gathered from their health insurers and providers. An example of this sort of comprehensive information system is the one currently under development in the United Kingdom (National Health Service, 1998).

An underlying purpose of an ambitious longitudinal sample-derived portrait of the American public would be to expand the budding synergy among existing health surveys into a broad and significantly more important source of health information. Each member of this reoriented confederation of the nation's most important health surveys would become an important component of a more comprehensive design paradigm, thereby creating a larger whole whose information benefits are significantly more than the sum utility of these surveys by themselves. A key to the success of such a longitudinal birth cohort study would be, of course, the creation of a reasonably nonthreatening division of information gathering roles among those agencies with a vested interest in these roles.

As with any longitudinal study that contains data linkages, careful attention must be paid to the issues of data privacy and confidentiality. The protection of the health data of study participants must be a priority, but will be a challenge nonetheless. In addition to the obvious needs of privacy legislation and technological protection solutions, efforts must be made to demonstrate to the public the benefits to be gained (or the information opportunities to be missed) from this study and to assure participants of the privacy of the data. Other countries that have already faced these obstacles and implemented solutions can again serve as models for our nation's system. For example, as mentioned above, the United Kingdom's National Health Service (NHS) outlined a strategy for their health system that includes lifelong electronic health records for every person in the country and round-the-clock on-line access to patient records and information about best clinical practices for all NHS clinicians (National Health Service, 1998). Similarly, a goal of the Canadian Advisory Council on Health Infostructure is to create information resources on factors affecting the health of Canadians and to improve privacy protection within the health sector (Advisory Council on Health Infostructure, 1999). The privacy and confidentiality of health data are very important and are not just a statistical issue. Efforts would need to be undertaken by all levels of government and society to ensure that the challenge of privacy protection is overcome without compromising the knowledge to be gained from a national longitudinal health study.

General responsibility for developing survey instrumentation, data analysis, and information dissemination for surveys addressing each information focus would be the responsibility of an agency partner; developing the design and gathering the data in this longitudinal study would fall under the purview of the health information planning board. By dividing responsibility this way, an agency with an information focus need not be consumed by data production, but rather can focus its efforts on data measurement and information generation and use. Beyond the health of particular relevance to their focus, each agency would benefit by having ready access to lifetime person-level data that would enable them to explain health outcomes or to evaluate efforts designed to improve these outcomes.

Challenging but solvable problems would be faced in fielding and producing findings from a national longitudinal study. At the planning stage the content of existing health surveys would need to be reevaluated and perhaps prioritized around established information needs. Some survey measures would need to be developed or improved to meet new or emerging information needs (e.g., quality of life, gauging the quality of provider care). Creative technological solutions would be needed to preserve the privacy and confidentiality of all gathered data. To minimize followup nonresponse and the risk of a "Hawthorne effect" on the respondents' health experiences, new ways to easily, efficiently, and innocuously extract data would need to be developed. To avoid measurement softness in portions of the respondent data record, uniform standards and definitions would need to be established and enforced for all study questionnaires. The sample would need to be large enough to facilitate the production of acceptable estimates for states and other "small areas," including specific race/ethnicity groups of interest. Finally, new or improved statistical methods to produce small-area estimates would be needed.

Another possibility to consider when thinking of the existing national surveys is a redesign and standardization that would allow for state and local estimates. Such a redesign would benefit states and permit flexibility in data collection and use. Furthermore, statewide data collection infrastructures are evolving, and as standardized sources of non-inpatient data become more readily available, the ability to link and aggregate these data into multi-state and national statistics increases, thus suggesting future reevaluation of the current national survey systems.

Although not explicitly included on the formal agenda of the workshop for discussion, there are valid historical and political reasons why the current system evolved as a set of independent "silos" spanning a large number of federal agencies. The many bureaucratic and "turf" issues and constraints that make it difficult to achieve the goals of an efficient data system were also not discussed. A health information planning board and a comprehensive longitudinal study would require a "surrendering of turf" by our major health agencies along with significant changes to remove practical and political barriers before they could be implemented. Those complex issues merit further work on how to motivate the desired organizational and individual behavior, perhaps drawing on expertise in organizational structure, human behavior, and societal and economic incentives. Without the insight of such work, a history of failed reorganization and cooperation attempts will continue.

Continuing growth in health care costs, in the complexity of the health care delivery process, and in the public's focus on matters of health, makes having timely, relevant, high-quality health data an essential part of the nation's future information landscape. In addition to raising many key issues we must face, this workshop points to an even greater need to search for and critically evaluate new ways to assemble and use health information. In failing to do so, we are reminded of Goethe's words that "nothing is more terrible than ignorance in action."

REFERENCES

Advisory Council on Health Infostructure
1999 Canada Health Infoway: Paths to Better Health. Final Report. Health Canada Publications, Ottawa. Available: <http://www.hc-sc.gc.ca/ohih-bsi> [July 12, 2001].

National Health Service
1998 Information for Health 1998-2005: An Information Strategy for the Modern NHS. National Health Service Executive, United Kingdom. Available: <http://www.nhsia.nhs.uk> [July 12, 2001].

Appendix:
Workshop Materials

AGENDA

THURSDAY, NOVEMBER 4, 1999

8:30 a.m.	**WELCOME / INTRODUCTIONS / GOALS OF THE WORKSHOP** *Edward Perrin, University of Washington, Workshop Chair* *Margaret Hamburg, Assistant Secretary for Planning and Evaluation, DHHS, and Co-chair, DHHS Data Council* *Edward Sondik, Director, National Center for Health Statistics and DHHS Statistical Advisor*
8:45-9:30 a.m.	**OVERVIEW PAPER/PRESENTATION**
8:45 a.m.	**Health Statistics: Past, Present, and Future** *Presenter: Dorothy Rice, University of California, San Francisco* This session will provide an overview and historical perspective on the elements of the nation's health statistics system. • Who has used health statistics, and what have their information needs been? • How do the various producers of health data

and statistics collectively comprise the nation's health statistics system?

- What have been some of the main challenges (technical, operational, resources, etc.) in producing health statistics? Have health statistics been fully utilized? Have they been responsive to information needs? What efforts have been made to deal with limitations in the health statistics system? How successful have they been?

9:10 a.m.	Comments from invited discussant *Janet Norwood, former U.S. Commissioner of Labor Statistics*
9:20 a.m.	Questions / Discussion
9:30-11:45 a.m.	**BACKGROUND PAPERS/PRESENTATIONS**
9:30 a.m.	**Health Statistics Systems in Other Countries** *Jennifer Zelmer, Canadian Institute for Health Information* This presentation will be based on a paper comparing how different countries organize their health statistics systems. The focus will be on Canada, Australia, and the United Kingdom. The paper will identify common models, trends, and future directions in the evolution of the vision for and management of national health statistics in these countries. *(NCVHS commissioned paper)*
9:45 a.m.	**From National Health Statistics to Health Information Systems** *Charlyn Black, University of Manitoba* This presentation will be based on a paper discussing key assumptions and critical components of a health information system; outlines the importance of being able to make

links across key areas and to examine various types and levels of investment in medical care for different populations; describes the Manitoba experience with the POPULIS system; and discusses possibilities for developing such systems in the future. *(NCVHS commissioned paper)*

10:00 a.m. Discussion

10:30-10:45 a.m. **Break**

10:45 a.m. **The Numbers We Need: Health Statistics and Health Policy**

Richard Kronick, University of California at San Diego

This presentation will focus on the health statistics needed for improving health policy, emphasizing the financing and delivery of care to the uninsured, and policy concerning managed care. The need for timely and detailed measurement of health-related quality of life will be considered. *(NCVHS commissioned paper)*

11:00 a.m. **The Children's Health Insurance Program and Other Efforts to Respond to Children's Health Needs**

Lorraine Klerman, University of Alabama at Birmingham

This presentation will use the State Children's Health Insurance Program (SCHIP) to illustrate the role health statistics played in determining a need for SCHIP and developing its provisions. Also discussed will be the health statistics that policy makers at the state and federal levels believe will be essential to demonstrate the impact of SCHIP. The need for additional statistics about children's health status and use of services will be considered. *(NCVHS commissioned paper)*

11:15 a.m. Discussion

11:45 a.m.-1:00 p.m. **Lunch**

1:00-5:00 p.m. **DEVELOPMENTS SHAPING FUTURE NEEDS FOR HEALTH STATISTICS**

Moderator: Edward Perrin, Workshop Chair

In each session, these 15-minute presentations will highlight new developments, changing priorities, newly emerging concerns, and specific opportunities and challenges for the health statistics system in the 21st century. The presentations are intended to combine a "visionary" consideration of these issues with observations on how current research or operational experience relates to these views of the future.

1:00 p.m. **Health Status and the Determinants of Health**

- Population-based assessments of health status, surveillance of illness, assessing well-being (positive health, psychosocial aspects of health, physical environment)
- *Presentations:*
  Determinants of health: Clyde Hertzman, University of British Columbia
  Positive health: Burton Singer, Princeton University
  Disease and risk factor surveillance: David Fleming, Oregon Health Division
  Health status assessment/instruments: Robert Kaplan, University of California at San Diego

2:00 p.m. Discussion

2:20 p.m. **Organization, Delivery, and Financing of Health Services**

- Supporting the management and analysis of the organization, delivery, and financing of personal and population-based health services and their relationship to health status and health outcomes
- *Presentations:*
  Health economics: Joseph Newhouse, Harvard University
  Health system change: Merwyn Greenlick, Oregon Health Sciences University
  Proprietary databases: Jacqueline Kosecoff, Protocare, Inc.

3:20 p.m. Discussion

3:40-4:00 p.m. **Break**

4:00 p.m. **The Data Policy and Technology Environment**

- Innovations in information technology for collecting, analyzing, and disseminating health data; issues of privacy, confidentiality, and security of health data
- *Presentations:*
  Survey methods: Sandra Berry, RAND
  Information technologies: William Eddy, Carnegie Mellon University
  Data policy issues: John Eisenberg, Agency for Healthcare Research and Quality
  Privacy and confidentiality issues: George Duncan, Carnegie Mellon University

5:00 p.m. Discussion

5:20 p.m. **ADJOURN**

FRIDAY, NOVEMBER 5, 1999

8:15-9:45 a.m. **PANEL DISCUSSION**

Changing Roles and Responsibilities in the Health Statistics System
Moderator: Edward Perrin, Workshop Chair
This panel will bring together panelists from several segments of the health statistics system to address priorities for

- changes in the health statistics system to meet new and continuing information needs and to respond to changing roles and responsibilities;
- research to respond to changing needs in the health statistics system.

Panelists will be asked to present for discussion specific proposals regarding steps that might be taken to meet evolving needs for health statistics. Panelists should consider needs/requirements of users and producers of health statistics and should draw on the background papers, the previous day's discussions, and their own experiences in framing their proposals.

The moderator will guide the discussion toward specific conclusions that can contribute to framing an overall agenda for steps to ready the health statistics system for the 21st century.

***Panelists** (from federal, state, local, and private sectors)*
Edward Sondik, Director, National Center for Health Statistics, and DHHS Statistical Advisor
John Lumpkin, Illinois Department of Public Health

Kenneth Thorpe, Rollins School of Public Health, Emory University
Eileen Peterson, United Health Group

9:45-10:00 a.m. **Break**

10:00 a.m.-1:15 p.m. **WRAP-UP SESSION**
Developing an Agenda for the Health Statistics System in the 21st Century
Moderator: Edward Perrin, Workshop Chair

10:00-10:15 a.m. **Reviewing the Issues**
Presenter: Miron Straf, Director (on leave), Committee on National Statistics

This session will provide a brief overview of the key issues covered during the workshop sessions, setting the stage for discussion of opportunities for change and research needs.

10:15-11:15 a.m. **Comments from Invited Discussants**

Invited discussants will reflect on implications of emerging health data needs for the health statistics system (15 minutes each).

Discussants:
Dorothy Rice, University of California, San Francisco
Richard Kronick, University of California at San Diego
Daniel Friedman, Massachusetts Department of Public Health
Michael Wolfson, Statistics Canada

11:15-1:15 p.m. **Identifying Opportunities for Change and Research Needs**

Discussion by all workshop participants to outline

- an agenda for key changes to respond to needs across the health statistics system

- a research agenda to respond to needs across the health statistics system
- priorities for HHS and its agencies engaged in health statistics activities

1:15 p.m. **ADJOURN**

PAPERS PRESENTED

The workshop papers were commissioned by the U.S. National Committee on Vital and Health Statistics (NCVHS). They are available at: <http://www.ncvhs.hhs.gov/hsvision/visiondocuments.html>.

"Recent Developments in Health Information: An International Perspective," Jennifer Zelmer, Shazeen Virani, and Richard Alvarez, Canadian Institute for Health Information.

"From Health Statistics to Health Information Systems: A New Path for The 21st Century," Charlyn Black, Noralou Roos, and Leslie Roos, University of Manitoba.

"Numbers We Need: Health Statistics and Health Policy," Richard Kronick, Department of Family and Preventive Medicine, University of California, San Diego.

"The State Children's Health Insurance Program: A Case Study," with additional material on data needs in the area of child health, Lorraine V. Klerman, Department of Maternal and Child Health, School of Public Health, University of Alabama at Birmingham.

"Building Robust Statistical Systems for Health," Daniel Melnick, Dan Melnick Research, Inc.

PARTICIPANTS

Presenters and Discussants

Edward B. Perrin (*Workshop Chair*), University of Washington, and Member, Committee on National Statistics
Sandra H. Berry, RAND, Santa Monica, California
Charlyn Black, University of Manitoba
George Duncan, Carnegie Mellon University
William F. Eddy, Carnegie Mellon University, and Member, Committee on National Statistics
John M. Eisenberg, Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services
David W. Fleming, Oregon Health Division
Daniel J. Friedman, Massachusetts Department of Public Health
Merwyn R. Greenlick, Oregon Health Sciences University
Margaret Hamburg, U.S. Department of Health and Human Services
Clyde Hertzman, University of British Columbia
Robert M. Kaplan, University of California at San Diego
Lorraine V. Klerman, University of Alabama, Birmingham
Jacqueline Kosecoff, Protocare Inc., Santa Monica, California
Richard Kronick, University of California, San Diego
John R. Lumpkin, Illinois Department of Public Health
Joseph P. Newhouse, Harvard University
Janet L. Norwood, former U.S. Commissioner of Labor Statistics
Eileen Peterson, Center for Health Care Policy and Evaluation, United Health Group
Alonzo Plough,* Seattle-King County Health Department
Dorothy P. Rice, School of Nursing, University of California, San Francisco
Burton H. Singer, Office of Population Research, Princeton University
Edward Sondik, National Center for Health Statistics, U.S. Department of Health and Human Services
Kenneth E. Thorpe, Rollins School of Public Health, Emory University
Michael C. Wolfson, Statistics Canada
Jennifer Zelmer, Canadian Institute for Health Information

---

*Unable to attend workshop, but graciously sent presentation materials that were distributed and discussed among workshop particpants.

Invited Guests

Ross Arnett, Agency for Health Care Policy and Research
Ruth Berkelman, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services
Linda Blankenbaker, National Center for Health Statistics, U.S. Department of Health and Human Services
James Bost, National Committee for Quality Assurance
Lisa Broitman, National Center for Health Statistics, U.S. Department of Health and Human Services
Claire Broome, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services
Brett Brown, Child Trends, Inc.
Olivia Carter-Pokras, Division of Policy and Data, Office of Minority Health, National Institutes of Health, U.S. Department of Health and Human Services
Kathryn Coltin, Harvard Pilgrim Health Care
Linda K. Demlo, National Center for Health Statistics, U.S. Department of Health and Human Services
John Drabok, U.S. Department of Health and Human Services
Rachel Duguay, The Lewin Group
J. Michael Fitzmaurice, Agency for Healthcare Research and Quality, U.S. Department of Health and Human Services
Manning Feinleib, Johns Hopkins School of Public Health
Daniel J. Friedman, Bureau of Health Statistics, Research and Evaluation, U.S. Department of Health and Human Services
Dan Gaylin, U.S. Department of Health and Human Services
Marsha Gold, Mathematica Policy Research
Marthe Gold, City University of New York Medical School
Donald Goldstone, Substance Abuse and Mental Health Services Administration, U.S. Department of Health and Human Services
Marjorie S. Greenberg, National Center for Health Statistics, U.S. Department of Health and Human Services
Jan Heinrich, U.S. General Accounting Office
Gerry Hendershot, National Center for Health Statistics, U.S. Department of Health and Human Services
Betsy L. Humphreys, National Library of Medicine, U.S. Department of Health and Human Services

Edward Hunter, National Center for Health Statistics, U.S. Department of Health and Human Services
William Hven, The Lewin Group
Raynard Kington, National Center for Health Statistics, U.S. Department of Health and Human Services
Mary Grace Kovar, National Opinion Research Center, Washington, DC
Andrew Kram, National Committee on Vital and Health Statistics
Thomas Louis, RAND, Arlington, VA, and Member, Committee on National Statistics
Jennifer Madans, National Center for Health Statistics, U.S. Department of Health and Human Services
Daniel Melnick, Dan Melnick Research, Inc.
Michael Millman, U.S. Department of Health and Human Services
Leyla Mohadjer, Westat, Rockville, MD
Robert S. Murphy, Westat, Rockville, MD
Paul Newacheck, University of California, San Francisco
Roy Gibson Parrish, Centers for Disease Control and Prevention, Atlanta
Steve Perente, University of Minnesota
Mark Rothstein, National Committee on Vital and Health Statistics
Patricia Ruggles, U.S. Department of Health and Human Services
James Scanlon, U.S. Department of Health and Human Services
Susan Schechter, U.S. Office of Management and Budget
Paul Scheible, National Compensation Survey
Elliot M. Stone, Massachusetts Health Data Consortium, Inc.
Michael Stoto, The George Washington University
Daniel Waldo, Health Care Financing Administration, U.S. Department of Health and Human Services
Elizabeth Ward, National Committee on Vital and Health Statistics
Robert Weinzimer, National Center for Health Statistics, U.S. Department of Health and Human Services

National Academies' Staff

Miron L. Straf, Committee on National Statistics, National Research Council
Andrew A. White, Committee on National Statistics, National Research Council
Terri Scanlan, Committee on National Statistics, National Research Council

Agnes Gaskin, Committee on National Statistics, National Research Council
Jane Durch, Institute of Medicine
Maria Hewitt, Institute of Medicine
Margarita Hurtado, Institute of Medicine
Gooloo Wunderlich, Institute of Medicine
Jonathan Zaff, Committee on Population, National Research Council