Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good: Workshop Summary

ISBN
978-0-309-12060-9

342 pages
6 x 9
PAPERBACK (2010)

W. Alexander Goolsby, LeighAnne Olsen, and Michael McGinnis; IOM Roundtable on Evidence-Based Medicine (Series); Institute of Medicine

THE LEARNING HEALTH SYSTEM SERIES

ROUNDTABLE ON VALUE & SCIENCE-DRIVEN HEALTH CARE

CLINICAL DATA AS THE BASIC STAPLE OF HEALTH LEARNING

Creating and Protecting a Public Good

Workshop Summary

Claudia Grossmann, W. Alexander Goolsby,
LeighAnne Olsen, and J. Michael McGinnis

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

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.

This project was supported by the Agency for Healthcare Research and Quality, America's Health Insurance Plans, AstraZeneca, Blue Shield of California Foundation, Burroughs Wellcome Fund, California Health Care Foundation, Centers for Medicare & Medicaid Services, Charina Endowment Fund, Department of Veterans Affairs, Food and Drug Administration, Johnson & Johnson, Moore Foundation, National Institutes of Health, The Peter G. Peterson Foundation, sanofi-aventis, and Stryker. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the organizations or agencies that provided support for this project.

International Standard Book Number-13: 978-0-309-12060-9
International Standard Book Number-10: 0-309-12060-8

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu.

For more information about the Institute of Medicine, visit the IOM home page at: **www.iom.edu.**

Printed in the United States of America

The serpent has been a symbol of long life, healing, and knowledge among almost all cultures and religions since the beginning of recorded history. The serpent adopted as a logotype by the Institute of Medicine is a relief carving from ancient Greece, now held by the Staatliche Museen in Berlin.

Suggested citation: IOM (Institute of Medicine). 2010. *Clinical data as the basic staple of health learning: Creating and protecting a public good: Workshop Summary*. Washington, DC: The National Academies Press.

"Knowing is not enough; we must apply.
Willing is not enough; we must do."
—Goethe

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the Nation. Improving Health.

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

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. Ralph J. Cicerone 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. Charles M. Vest 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. Harvey V. Fineberg 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. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

ROUNDTABLE ON VALUE & SCIENCE-DRIVEN HEALTH CARE[1]

Denis A. Cortese (*Chair*), Emeritus President and Chief Executive Officer, Mayo Clinic; Foundation Professor, ASU
Donald Berwick, Administrator, Centers for Medicare & Medicaid Services (*ex officio*)
David Blumenthal, National Coordinator, Office of the National Coordinator for Health IT (*ex officio*)
Bruce G. Bodaken, Chairman, President, and Chief Executive Officer, Blue Shield of California
David R. Brennan, Chief Executive Officer, AstraZeneca PLC
Paul Chew, Chief Science Officer and CMO, sanofi-aventis U.S., Inc.
Carolyn M. Clancy, Director, Agency for Healthcare Research and Quality (*ex officio*)
Michael J. Critelli, Former Executive Chairman, Pitney Bowes, Inc.
Helen Darling, President, National Business Group on Health
Thomas R. Frieden, Director, Centers for Disease Control and Prevention (*designee*: **Chesley Richards**) (*ex officio*)
Gary L. Gottlieb, President and CEO, Partners HealthCare System
James A. Guest, President, Consumers Union
George C. Halvorson, Chairman and Chief Executive Officer, Kaiser Permanente
Margaret A. Hamburg, Commissioner, Food and Drug Administration (*ex officio*)
Carmen Hooker Odom, President, Milbank Memorial Fund
Ardis Hoven, Professor of Medicine, University of Kentucky; Chair-elect, American Medical Association
Brent James, Chief Quality Officer and Executive Director, Institute for Health Care Delivery Research, Intermountain Healthcare
Michael M. E. Johns, Chancellor, Emory University
Craig Jones, Director, Vermont Blueprint for Health
Cato T. Laurencin, Vice President for Health Affairs, Dean of the School of Medicine, University of Connecticut
Stephen P. MacMillan, President and Chief Executive Officer, Stryker
Mark B. McClellan, Director, Engelberg Center for Healthcare Reform, The Brookings Institution
Sheri S. McCoy, Worldwide Chairman, Johnson & Johnson Pharmaceuticals Group
Elizabeth G. Nabel, President, Brigham and Women's Hospital

[1] Formerly the Roundtable on Evidence-Based Medicine. IOM forums and roundtables do not issue, review, or approve individual documents. The responsibility for the published workshop summary rests with the workshop rapporteurs and the institution.

Mary D. Naylor, Professor and Director of Center for Transitions in Health, University of Pennsylvania
Peter Neupert, Corporate Vice President, Health Solutions Group, Microsoft Corporation
Nancy H. Nielsen, Past President, American Medical Association
William D. Novelli, Former CEO, AARP; Professor, Georgetown University
Jonathan B. Perlin, Chief Medical Officer and President, Clinical Services, HCA, Inc.
Robert A. Petzel, Under Secretary, Veterans Health Administration (*ex officio*)
Richard Platt, Professor and Chair, Harvard Medical School and Harvard Pilgrim Health Care
John C. Rother, Group Executive Officer, AARP
John W. Rowe, Professor, Mailman School of Public Health, Columbia University
Susan Shurin, Acting Director, National Heart, Lung, and Blood Institute (*ex officio*)
Mark D. Smith, President and CEO, California HealthCare Foundation
George P. Taylor, Assistant Secretary for Health Affairs (Acting), Department of Defense (*designee*: **Michael Dinneen**) (*ex officio*)
Reed D. Tuckson, Executive VP and Chief of Medical Affairs, UnitedHealth Group
Frances M. Visco, President, National Breast Cancer Coalition

Workshop Planning Committee

David Blumenthal, Office of the National Coordinator for Health IT (formerly Massachusetts General Hospital, Harvard University)
Marc Boutin, National Health Council
Mary Durham, Kaiser Foundation Hospitals
Lynn Etheredge, George Washington University
George Isham, HealthPartners Inc.
Peter Juhn, Johnson & Johnson
Alexander Walker, Harvard School of Public Health, Harvard University

Roundtable Staff

Christie Bell, Financial Associate
Katharine Bothner, Senior Program Assistant (through July 2008)
Patrick Burke, Financial Associate (until December 2009)
Andrea Cohen, Financial Associate (until December 2008)
W. Alexander Goolsby, Program Officer (through August 2008)
Claudia Grossmann, Program Officer

Kiran Gupta, Research Assistant (until May 2009)
Katie Jakubs, NIH detailee
J. Michael McGinnis, Senior Scholar and Executive Director
LeighAnne Olsen, Program Officer
Daniel O'Neill, Research Associate (until December 2008)
Stephen Pelletier, Consultant
Brian Powers, Program Assistant
Ruth Strommen, Intern
Kate Vasconi, Senior Program Assistant
Pierre Yong, Program Officer
Catherine Zweig, Senior Program Assistant

Reviewers

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 National Research Council's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making its 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 wish to thank the following individuals for their review of this report:

Simon Cohn, Kaiser Permanente
John Lumpkin, Robert Wood Johnson Foundation
Douglas Peddicord, Oldaker, Belair, & Wittie, LLC
Alan Westin, Columbia University

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the final draft of the report before its release. The review of this report was overseen by **Daniel Masys,** Vanderbilt University School of Medicine. Appointed by the National Research Council and the Institute of Medicine, Dr. Masys 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 committee and the institution.

Institute of Medicine
Roundtable on Value & Science-Driven Health Care[1]
Charter and Vision Statement

The Institute of Medicine's Roundtable on Value & Science-Driven Health Care has been convened to help transform the way evidence on clinical effectiveness is generated and used to improve health and health care. Participants have set a goal that, by the year 2020, 90 percent of clinical decisions will be supported by accurate, timely, and up-to-date clinical information, and will reflect the best available evidence. Roundtable members will work with their colleagues to identify the issues not being adequately addressed, the nature of the barriers and possible solutions, and the priorities for action, and will marshal the resources of the sectors represented on the Roundtable to work for sustained public-private cooperation for change.

****

The Institute of Medicine's Roundtable on Value & Science-Driven Health Care has been convened to help transform the way evidence on clinical effectiveness is generated and used to improve health and health care. We seek the development of a ***learning health system*** that is designed to generate and apply the best evidence for the collaborative healthcare choices of each patient and provider; to drive the process of discovery as a natural outgrowth of patient care, and to ensure innovation, quality, safety, and value in health care.

Vision: Our vision is for a healthcare system that draws on the best evidence to provide the care most appropriate to each patient, emphasizes prevention and health promotion, delivers the most value, adds to learning throughout the delivery of care, and leads to improvements in the nation's health.

Goal: By the year 2020, 90 percent of clinical decisions will be supported by accurate, timely, and up-to-date clinical information, and will reflect the best available evidence. We feel that this presents a tangible focus for progress toward our vision, that Americans ought to expect at least this level of performance, that it should be feasible with existing resources and emerging tools, and that measures can be developed to track and stimulate progress.

Context: As unprecedented developments in the diagnosis, treatment, and long-term management of disease bring Americans closer than ever to the promise of personalized health care, we are faced with similarly unprecedented challenges to identify and deliver the care most appropriate for individual needs and conditions. Care that is important is often not delivered. Care that is delivered is often not important. In part, this is due to our failure to apply the evidence we have about the medical care that is most effective—a failure related to shortfalls in provider knowledge and accountability, inadequate care coordination and support, lack of insurance, poorly aligned payment incentives, and misplaced patient expectations. Increasingly, it is also a result of our

[1] Formerly the Roundtable on Evidence-Based Medicine.

limited capacity for timely generation of evidence on the relative effectiveness, efficiency, and safety of available and emerging interventions. Improving the value of the return on our healthcare investment is a vital imperative that will require much greater capacity to evaluate high priority clinical interventions, stronger links between clinical research and practice, and reorientation of the incentives to apply new insights. We must quicken our efforts to position evidence development and application as natural outgrowths of clinical care—to foster health care that learns.

Approach: The IOM Roundtable on Value & Science-Driven Health Care serves as a forum to facilitate the collaborative assessment and action around issues central to achieving the vision and goal stated. The challenges are myriad and include issues that must be addressed to improve evidence development, evidence application, and the capacity to advance progress on both dimensions. To address these challenges, as leaders in their fields, Roundtable members will work with their colleagues to identify the issues not being adequately addressed, the nature of the barriers and possible solutions, and the priorities for action, and will marshal the resources of the sectors represented on the Roundtable to work for sustained public-private cooperation for change.

Activities include collaborative exploration of new and expedited approaches to assessing the effectiveness of diagnostic and treatment interventions, better use of the patient care experience to generate evidence on effectiveness, identification of assessment priorities, and communication strategies to enhance provider and patient understanding and support for interventions proven to work best and deliver value in health care.

Core concepts and principles: For the purpose of the Roundtable activities, we define science-driven health care broadly to mean that, ***to the greatest extent possible, the decisions that shape the health and health care of Americans—by patients, providers, payers, and policymakers alike—will be grounded on a reliable evidence base, will account appropriately for individual variation in patient needs, and will support the generation of new insights on clinical effectiveness***. Evidence is generally considered to be information from clinical experience that has met some established test of validity, and the appropriate standard is determined according to the requirements of the intervention and clinical circumstance. Processes that involve the development and use of evidence should be accessible and transparent to all stakeholders.

A common commitment to certain principles and priorities guides the activities of the Roundtable and its members, including the commitment to: the right health care for each person; putting the best evidence into practice; establishing the effectiveness, efficiency, and safety of medical care delivered; building constant measurement into our healthcare investments; the establishment of healthcare data as a public good; shared responsibility distributed equitably across stakeholders, both public and private; collaborative stakeholder involvement in priority setting; transparency in the execution of activities and reporting of results; and subjugation of individual political or stakeholder perspectives in favor of the common good.

Foreword

The Institute of Medicine (IOM) was established in 1970 as part of the National Academies to serve as the foremost adviser to the nation on issues related to medicine, health, and the biomedical sciences. Essential to fulfilling this role is a commitment to drawing upon the best evidence to guide policy development. Indeed, the capacity to excel, innovate, and advance in any field of inquiry is predicated upon the quality and availability of information—whether to guide decision making, suggest new areas of research, or confirm hypotheses. Clinical data represent an enormous opportunity to improve the quality and efficiency of the nation's healthcare system. Whether captured as part of research, through delivery processes, or at the point of care, these data will be central to advance our understanding of which medical practices and treatments work best for different patients, and drive continual improvements in the delivery of health care. This potential will only be enhanced as health information technology enables the capture and analysis of vastly larger quantities of data.

While important data are currently collected by a number of healthcare organizations and sectors, their capture and use reflect the fragmentation of the healthcare system, with much of this information siloed in different databases and repositories. In addition, broader sharing, aggregation, and use of these data are often restricted due to proprietary or privacy concerns. Taking better advantage of these resources, while maintaining appropriate privacy protections, requires the engagement of all healthcare sectors in discussions of key challenges and opportunities in a neutral and trusted forum, such as that provided by the IOM Roundtable on Value & Science-Driven Health Care. Established in 2006 as the Roundtable on Evidence-Based

Medicine, the IOM Roundtable is composed of the nation's top leaders in the public and private sectors who have a keen interest in transforming our nation's healthcare system. Roundtable members have established a common vision for a learning health system and a goal that by the year 2020, 90 percent of clinical decisions will be supported by accurate, timely, and up-to-date clinical information and will reflect the best available evidence. Since its inception, the Roundtable has convened a series of workshops and publications, as well as initiated a variety of projects and activities to help accelerate progress toward a learning health system.

This publication represents the sixth in the Learning Health System series and summarizes the discussions at the Roundtable's 2-day workshop titled Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good. The workshop explored the range of activities that constitute the system of healthcare data in the United States, as well as the challenges and opportunities associated with efficiently leveraging data. Discussion identified many challenges, barriers, and policy issues that must be engaged to move to the next generation of data utility. As reflected by participants' comments, broader use of health information technology for knowledge development, particularly electronic health records, holds significant potential for healthcare advancement through collaborative data mining and improving transparency of and access to data. Also evident is the central importance of stakeholder leadership, such as that provided by the Roundtable, for maximizing the use of clinical data for continuous learning and improvement in health care.

I would like to offer my personal thanks to the Roundtable members for the leadership that they bring to these important issues, to the Roundtable staff for their skill and dedication in coordinating and facilitating the activities, and importantly, to the sponsors who make this work possible: Agency for Healthcare Research and Quality, America's Health Insurance Plans, AstraZeneca, Blue Shield of California Foundation, Burroughs Wellcome Fund, California Health Care Foundation, Centers for Medicare & Medicaid Services, Charina Endowment Fund, Department of Veterans Affairs, Food and Drug Administration, Johnson & Johnson, Moore Foundation, National Institutes of Health, The Peter G. Peterson Foundation, sanofi-aventis, and Stryker.

Harvey V. Fineberg, M.D., Ph.D.
President, Institute of Medicine

Preface

The Institute of Medicine convened the Roundtable on Value & Science-Driven Health Care (formerly the Roundtable on Evidence-Based Medicine) in 2006 to provide a trusted forum in which multiple healthcare sectors—including patients, health providers, payers, employees, health product manufacturers, information technology companies, policy makers, and researchers—could share perspectives on key opportunities to help transform how evidence is generated and applied to drive improvements in the efficiency and effectiveness of health care, to guide healthcare decisions, and to improve the nation's health. An early outcome of the Roundtable activities is a shared vision of a healthcare system that draws on the best available evidence to appropriately tailor care to individual patients and to continuously add to the healthcare knowledge base. Salient topics, addressed through public workshops and collaborative groups, engage critical aspects of concepts essential to achieving this vision and the Roundtable's goal that by 2020, 90 percent of clinical decisions will be supported by accurate, timely, and up-to date clinical information, and will reflect the best available evidence.

The perspectives, themes, and insights from each workshop are disseminated through the Learning Health System series of publications. The *Clinical Data as a Basic Staple of Health Learning: Creating and Protecting a Public Good* publication is a summary of the proceedings of the sixth workshop in the Learning Health System series. Held on February 12–13, 2008, this workshop was designed to explore leading perspectives on clinical data as a transformative agent in health care, as well as possible strategies for their implementation in the delivery of evidence-based care. Issues motivating the discussion include:

1. Discovering what works best in medical care—including for whom and under what circumstances—requires that clinical data be carefully nurtured as a resource for continuous learning.
2. Transformational opportunities are presented by evolving large and potentially interoperable clinical and administrative datasets.
3. Clinical data are recorded and held in multiple activities and many institutions, including medical records, administrative and claims records, and research studies.
4. Public policy and public awareness lag behind the technical, organizational, and legal capacity for reliable safeguarding of individual privacy and data security in mining clinical data for new knowledge.
5. A significant challenge to progress resides in the barriers and restrictions that derive from the treatment of medical care data as a proprietary commodity by the organizations involved.
6. Even clinical research and medical care data developed with public funds are often not available for broader analysis and insights.
7. Broader access and use of healthcare data for new insights require not only fostering data system reliability and interoperability but also addressing the matter of individual data ownership and the extent to which data central to progress in health and health care should constitute a public good.

During the 2-day workshop, participants explored a variety of relevant technical, economic, legal, and policy issues important to addressing these issues. Invited speakers and panelists highlighted opportunities to advance elements of clinical data and identified areas in need of greater attention and focus. The following pages summarize the workshop discussion, including the review of characteristics of clinical data, the marketplace for healthcare data, legal issues related to data access and use, and the role of privacy and security concerns. Throughout the workshop, participants identified specific policy issues in need of engagement to move to the next generation of data utility. The need for broader public engagement was also discussed, as participants noted that public policy and awareness often lag behind the technical, organizational, and legal capacity for reliable safeguarding of individual privacy and data security. Key opportunities identified included realigning incentives to promote real-time use of clinical data in evidence development, correcting the market failure for expanding electronic health records, and greater engagement of the public in shaping evidence development strategies.

We wish to acknowledge the individuals and organizations who offered their time and guidance in the development and execution of the workshop and summary. Individuals presenting at the workshop and authoring

manuscripts for the summary are acknowledged in particular. Also vital to the success of the workshop were the planning committee members, representing a cross-section of stakeholders and thought leaders in clinical data issues: David Blumenthal (Massachusetts General Hospital), Mary Durham (Kaiser Permanente), Lynn Etheredge (George Washington University), George Isham (HealthPartners, Inc.), Peter Juhn (Johnson & Johnson), and Alexander Walker (Harvard University). Roundtable staff members, including Katharine Bothner, Alexander Goolsby, Claudia Grossmann, Kiran Gupta, LeighAnne Olsen, Daniel O'Neill, Kate Vasconi, Pierre Yong, and Catherine Zweig, were instrumental to workshop coordination and summary production. We also wish to acknowledge the Roundtable members for their guidance and the sponsors of Roundtable activities: Agency for Healthcare Research and Quality, America's Health Insurance Plans, AstraZeneca, Blue Shield of California Foundation, Burroughs Wellcome Fund, California Health Care Foundation, Centers for Medicare & Medicaid Services, Charina Endowment Fund, Department of Veterans Affairs, Food and Drug Administration, Johnson & Johnson, Moore Foundation, National Institutes of Health, The Peter G. Peterson Foundation, sanofi-aventis, and Stryker.

Denis A. Cortese, M.D.
Chair,
Roundtable on Value & Science-Driven Health Care

J. Michael McGinnis, M.D., M.P.P.
Executive Director,
Roundtable on Value & Science-Driven Health Care

Contents

Summary

INTRODUCTION AND OVERVIEW

Because of their potential to enable the development of new knowledge and to guide the development of best practices from the growing sum of individual clinical experiences, clinical data represent the resource most central to healthcare progress (Arrow et al., 2009; Detmer, 2003). Whether captured during product development activities such as clinical research trials and studies, or as a part of the care delivery process, these data are fundamental to the delivery of timely, appropriate care of value to individual patients—and essential to building a system that continually learns from and improves upon care delivered. The opportunities for learning from practice are substantial, from improved understanding of the effects of different treatments and therapies in specific patient subpopulations, to developing and refining practices to streamline or tailor care processes for complex patients, to the development of a delivery system that can advance the evidence base on novel diagnostic and therapeutic techniques (Hrynaszkiewicz and Altman, 2009; Nass et al., 2009; NRC, 2009; Safran, 2007). Furthermore, U.S. per capita healthcare costs are now nearly double that of comparable nations (*Health care spending in the United States and OECD countries*, 2007), and broader access and use of existing and future clinical data may be a key opportunity to better understand and address system-wide factors—such as waste and inefficiencies—that contribute to rising healthcare expenditures.

Clinical data now reside in many often unconnected and inaccessible repositories, making linkage, analysis, and interpretation of these data

challenging on an individual or population level. The increase in potentially interoperable electronic and personal health datasets—integrated with laboratory values, diagnostic images, and patient demographic information and preferences—and development of approaches to link and network these data offer even greater opportunity to create and use rich data resources to help transform healthcare delivery and improve the public's health. Concerns about privacy of health data, as well as the treatment of medical data—even those generated with public funds—as proprietary goods pose additional challenges to data use (Blumenthal, 2006; Nass et al., 2009, editors, *Nature* 2005, Ness, 2007; Piwowar et al., 2008).

The utility of clinical data as a transformative agent in the U.S. healthcare system was the focus of the February 2008 Institute of Medicine (IOM) workshop, Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good. Issues motivating discussion include the potential for clinical data as a resource for continuous learning and key component of an efficient healthcare system; the opportunities presented by vastly larger and potentially interoperable data resources—particularly those developed with public funds; the challenges and barriers to more appropriate use of these resources (e.g., related to the fragmentation of data, proprietary nature of data, and privacy concerns); the lag of public policy development and public awareness of and attention to these issues; and the need to address key issues, including the extent to which data constitute a public good (Box S-1).

During the 2-day workshop, participants representing a variety of healthcare perspectives, reviewed current use of data for benchmarking and generating new clinical and operational insights, and discussed a sampling of innovative efforts to aggregate data for greater insights. In evaluating the current marketplace for care data, participants presented opportunities to increase access to and sharing of health information as private and public goods, while devoting particular attention to legal and social aspects of privacy and security of healthcare data. The workshop addressed multiple health-sector perspectives in the identification of specific policy areas for developing strategies and next-generation health data systems. Engaging the public in the advances necessary to develop a learning health system was viewed as a particularly important area for further discussion.

The IOM Roundtable and the Clinical Data Utility

Convened by the IOM in 2006, the Roundtable on Value & Science-Driven Health Care (formerly the Roundtable on Evidence-Based Medicine) serves as a mechanism for bringing stakeholders from multiple sectors together to evaluate means through which improving the generation and application of evidence will accelerate progress toward an efficient, effective

BOX S-1
Issues Motivating Discussion

1. Discovering what works best in medical care—including for whom and under what circumstances—requires that clinical data be carefully nurtured as a resource for continuous learning.
2. Transformational opportunities are presented by evolving large and potentially interoperable clinical and administrative datasets.
3. Clinical data are recorded and held in multiple activities and many institutions, including medical records, administrative and claims records, and research studies.
4. Public policy and public awareness lag behind the technical, organizational, and legal capacity for reliable safeguarding of individual privacy and data security in mining clinical data for new knowledge.
5. A significant challenge to progress resides in the barriers and restrictions that derive from the treatment of medical care data as a proprietary commodity by the organizations involved.
6. Even clinical research and medical care data developed with public funds are often not available for broader analysis and insights.
7. Broader access and use of healthcare data for new insights requires not only fostering data system reliability and interoperability but also addressing the matter of individual data ownership and the extent to which data central to progress in health and health care should constitute a public good.

U.S. medical care system. These stakeholders span the realm of health care, and include patients, employers, health product manufacturers, payers, policy makers, providers, and researchers. As a guiding principle for the Roundtable, decisions shaping American health and health care will draw from a proven evidence base, appropriately accommodate patient variation, and simultaneously generate additional insight into clinical effectiveness.

Roundtable participants established a goal that, by the year 2020, 90 percent of clinical decisions will be supported by accurate, timely, and up-to-date clinical information and will reflect the best available evidence. Central to this goal is the development of a learning health system designed to generate the best evidence for the collaborative healthcare choices of each patient and each provider; to drive the process of discovery as a natural outgrowth of patient care; and to ensure innovation, quality, safety, and value in health care. The broader availability and use of clinical data is an essential component of a learning system given the large potential for gains

in the efficiency, quality, and safety of the care delivered; however, such a shift will have systemwide implications: drawing upon resources in each sector, and requiring cross-sector cooperation and discussion to ensure the appropriate development, support, and use of these resources.

The Roundtable's Learning Health System series of workshops and publications are opportunities to foster the broad cross-sector discussions needed to better characterize the key elements, barriers, and needs of a transformed healthcare system. Each workshop is summarized in a publication available through the National Academies Press. Workshops and publications in this series since 2006 include:

- The Learning Healthcare System
- Judging the Evidence: Standards for Determining Clinical Effectiveness
- Leadership Commitments to Improve Value in Health Care: Finding Common Ground
- Redesigning the Clinical Effectiveness Research Paradigm: Innovation and Practice-Based Approaches
- Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good
- Engineering a Learning Healthcare System: A Look at the Future
- Learning What Works: Infrastructure Required to Learn Which Care Is Best
- Value in Health Care: Accounting for Cost, Quality, Safety, Outcomes, and Innovation
- The Healthcare Imperative: Lowering Costs and Improving Outcomes—A Four-Part Workshop Series

This publication summarizes the proceedings of the sixth workshop in the series, Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good. A summary chapter includes highlights from each workshop session; manuscripts submitted by each speaker and panel discussion summaries can be found in the subsequent chapters. Two keynote presentations, included in Chapter 1, titled "Clinical Data as the Basic Staple of the Learning Health System" and "Creating a Public Good for the Public's Health," offered critical context for the workshop. The first day of the 2-day workshop also featured presentations that profiled data in the current healthcare system (Chapter 2), provided an overview of innovative efforts to use data (Chapter 3), evaluated the public and private natures of healthcare data (Chapter 4), and discussed issues related to privacy and security (Chapter 5). The second day featured a panel discussion on policy opportunities (also in Chapter 5) and presentations and discussions that identified next-generation data utilities (Chapter 6). The

workshop concluded with a focus on engaging the public in efforts to use clinical data for insights (Chapter 7) and some final observations on meeting themes and potential follow-on activities (Chapter 8). The workshop agenda, biographical sketches, and a list of participants are located in the appendixes.

COMMON THEMES

Apart from shedding light on the issues that impede or challenge improved data utility, the discussion identified a rich array of ideas for accelerating progress toward better application of data. Across the 2 days of presentations and discussion, a compelling set of reoccurring themes emerged for follow-on attention.

BOX S-2
Workshop Common Themes

- Clarity on the basic principles of clinical data stewardship.
- Incentives for real-time use of clinical data in evidence development.
- Transparency to the patient when data are applied for research.
- Addressing the market failure for expanding EHRs.
- Personal records and portals that center patients in the learning process.
- Coordinated EHR user organization evidence development work.
- The business case for expanded data sharing in a distributed network.
- Assuring publicly funded data are used for the public benefit.
- Broader semantic strategies for data mining.
- Public engagement in evidence development strategies.

- ***Clarity on the basic principles of clinical data stewardship.*** The starting point for expanded access and use of clinical data for knowledge development is agreement on some of the fundamental notions to guide the activities for all individuals and organizations with responsibility for managing clinical data. Workshop participants repeatedly mentioned the need for consensus on approaches to such issues as data structure, standards, reporting requirements, quality assurance, timeliness, deidentification or security measures, and access and use procedures—all of which will determine the pace and nature of evidence development.

- ***Incentives for real-time use of clinical data in evidence development.*** Current barriers to the real-time use of clinical data for new knowledge discussed at the workshop ranged from regulatory and commercial issues to cost and quality issues. Participants suggested the need for a dedicated program of activities, incentives, and strategies to improve the methods and approaches, their testing and demonstration, the cooperative decision making on priorities and programs, and the collective approach to regulatory barriers.
- ***Transparency to the patient when data are applied for research.*** Patient acceptance is key to use of clinical data for knowledge development, and patient engagement and control are key to acceptance. In this respect, clarity to individual patients on the structure, risks, and benefits of access to data for knowledge development was noted by participants as particularly important. Patient confidence and system accountability may be enhanced through transparent notification and audit processes in which patients are informed of when and by whom their information has been accessed for knowledge development.
- ***Addressing the market failure for expanding electronic health records.*** Currently, market incentives are not enough to bring about the expansion of use of electronic health records necessary to make the point of care a locus for the development, sharing, and application of knowledge about what works best for individual patients. Shortfalls noted by participants included demand by providers or patients that is not sufficient to counter the expense to small organizations, competing platforms, and asynchronous reporting requirements that work against their utility for broad quality and outcome determinations, and that even the larger payers—apart from government—do not possess the critical mass necessary to drive broader scale applicability and complementarity. Deeper, more directed, and coordinated strategy involving Medicare leadership will likely be needed to foster such changes.
- ***Personal records and portals that center patients in the learning process.*** Patient demand could be instrumental in spreading the availability of electronic health records for improving patient care and knowledge development. Such demand will depend on much greater patient access to, comfort with, and regular use of programs that allow either the maintenance of personal electronic health records or access through a dedicated portal to their provider-maintained electronic medical record. As noted during the workshop, many consumer-oriented products under development give patients and consumers more active roles in managing per-

sonal clinical information. These may help to demonstrate value in the speed and ease of personal access to the information, better accommodate patient preference in care, and foster a partnership spirit conducive to the broader electronic health records (EHRs) application.

- ***Coordinated EHR user organization evidence development work.*** The development of a vehicle to enhance collaboration among larger EHR users of different vendors was raised during the workshop as a means to accelerate the emergence of more standardized agreements and approaches to integrating and sharing data across multiple platforms, common query strategies, virtual data warehousing rules and strategies, relational standards, and engagement of ways to reduce misperceptions on regulatory compliance issues.
- ***The business case for expanded data sharing in a distributed network.*** Demonstrating the net benefits of data sharing could promote its use. Benefits suggested by participants included cost savings or avoidance from facilitated feedback to providers on quality and outcomes; quick, continuous improvement information; and improved management, coordination, and assessment of patient care.
- ***Assuring publicly funded data for the public benefit.*** Federal and state funds that support medical care and support insights into medical care through clinical research grant funding are the source of substantial clinical data, yet many participants observed that these resources are not yet effectively applied to the generation of new knowledge for the common good.
- ***Broader semantic strategies for data mining.*** Platform incompatibilities for clinical data substantially limit the spread of electronic health records and their use for knowledge development. Yet discussion identified strategies using alternative semantic approaches for mining clinical data for health insights, which may warrant dedicated cooperative efforts to develop and apply them.
- ***Public engagement in evidence development strategies.*** Generating a base of support for and shared emphasis on developing a healthcare ecosystem in which all stakeholders play a contributory role was noted by many participants as important for progress. Ultimately, the public will determine the broad acceptance and applicability of clinical data for knowledge development, underscoring the importance of keeping the public closely involved and informed on all relevant activities to use clinical data to generate new knowledge.

PRESENTATION AND DISCUSSION SUMMARIES

Presentations at the Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good included perspectives from healthcare sectors and beyond on the current state of clinical data and data systems, the implications of healthcare data as a public good, and potential opportunities for improving their collection and use. These workshop presentations served as the genesis of the papers that constitute the chapters that follow, and are summarized below.

Clinical Data as the Basic Staple of Health Learning

Clinical data consist of information ranging from determinants of health and measures of health and health status to documentation of care delivery. These data are captured for a variety of purposes and stored in numerous databases across the healthcare system. Advances in health information technology (HIT) and analytics raise the potential for these data to be used to fill substantial knowledge gaps in health care, but complicating the needed aggregation and use of these data are technical, cultural, and legal barriers. Although efforts are underway to address technical barriers and privacy concerns, many have suggested that a shift is needed to a data-sharing culture in which clinical data are considered a basic staple of health learning. Chapter 1 provides a brief overview of these issues and includes summaries of the context setting remarks of the workshop's two keynote speakers. David Brailer's presentation profiled current clinical data collection and use, and offered his perspective on future applications that would improve care delivery, research, and health outcomes; and Carol Diamond offered her perspective on what might be possible if data were treated as a public good and identified several policy and technical issues important to achieving this vision.

Clinical Data as the Basic Staple of the Learning Health System

Brailer, Chair of Health Evolution Partners, was appointed in 2004 as the first national coordinator for health information technology. He served in that position until 2006. He described the potential of clinical data as a key building block for a learning health system by profiling the utility of clinical data currently available, as well as what might be possible if all data sources could be readily and reliably drawn upon for new insights into healthcare effectiveness. A significant gap exists between the potential utility of clinical data and how data are treated in the current healthcare market—where clinical information is proprietary and used for strategic

benefit. A question important for progress is whether clinical data are a public or private good.

Brailer noted that significant progress has been made in the past few years to broaden adoption of HIT. Advances in HIT certification and standardization efforts have produced portable health information and enabled exchange of significant volumes of information. Moreover, many hospitals have made progress in implementing electronic records, as have many physicians, especially those in large group practices. Emphasizing the important role of engaging the public, Brailer discussed opportunities for HIT companies such as Microsoft and Google to interact with the public and to raise the potential for data access and use for health improvement.

Yet, misaligned incentives, based on current systems of reimbursement, and an outdated privacy paradigm currently hinder progress. Brailer suggested that the development of a framework for privacy that recognizes the dynamic, portable, and compounding nature of health information assets is a necessary first step to facilitating greater data sharing and use. A second major challenge is ensuring that information developed via data sharing efforts are truly useful to clinical care—and in this respect, there is a major tension between adoption and interoperability. Interoperability—or the capacity to share, integrate, and apply health information from disparate sources—has been the principal priority of the nation's health information agenda. However, as the push for adoption gains momentum, there is potential for moving health information tools into broad use. This may not be best suited to also support the interoperability needed for a learning health system.

Many aspects of the data and data systems essential for the development and assembly of coherent, representative, timely, and valid information that can inform decisions at patient or population levels are understood. However, leadership is needed to help maintain a focus on developing and using data to make health care smarter in the face of competing near-term priorities—especially given the many challenges faced by providers, payers, and consumers in term of access and cost. Progress in efforts to make clinical data structured, intelligent, useful, assembled, and applied in a way that makes care better requires a sharper focus on data stewardship. Under one scenario, health information could become a true public good that is not proprietary; alternatively, clinical information could become a private good, used differentially for comparative advantage.

Nothing in federal or state statutes, regulations, or other guidance confers control of health information to any data originator (e.g., provider, hospital, manufacturers); yet, in practice, clinical information today is largely a private good, controlled by data producers. Brailer observed that the *Health Information Portability and Accountability Act* (HIPAA) enables de facto provider control over health information as patients can-

not direct that their information be sent to a third party, nor are providers obligated to make data available in a timely or convenient format. Other regulations also create barriers to portable, available, and acceptable health information. Brailer described the potential implications of proprietary use versus transparency of health information for health system stakeholders, and noted that it should be feasible to create a system in which providers can gain advantage from being performance driven, yet not gain advantage from exclusive use of health information as a private good.

Ultimately, advancing the notion of clinical data as a public good is essential to a healthcare system that learns. Such efforts also offer the potential to extend the benefits of the information revolution—already experienced by many other industries—to health care by providing the power of choice to consumers. Brailer contended there is a limited window of opportunity available to achieve this end through technical and policy advances that make data more useful and valuable while also ensuring equitable access and maintaining a competitive marketplace. Clarification of data stewardship is needed to promote a shared understanding and transparency with respect to data control, ownership, and access.

Vision for the Future: Creating a Public Good for the Public's Health

Carol Diamond, managing director of the Markle Foundation's Health Program, delivered the keynote address on the second day of the workshop. She outlined a vision for clinical data positioned as a public good and provided guidance on the technical and policy issues required to build the public trust necessary to achieve this vision. The work of the public–private collaborative, Connecting for Health (CFH), was discussed to illustrate key opportunities to develop a health information sharing environment that seeks to improve the quality and cost effectiveness of health care. The work of CFH to improve how information is used to address research, public health, and quality measurement was emphasized.

Achieving population health goals requires analysis, decision support, and feedback loops embedded throughout the system. However, as revealed by the significant challenges of collecting, cleaning, and analyzing health data for existing data reporting demands, progress will require a new approach to collecting, accessing, and using health information. To guide system development, Diamond suggested the need to consider three central requirements for responsible information policies: fulfilling seven core privacy principles (openness and transparency, purpose specification, collection and use limitation, individual participation and control, data integrity and quality, and security safeguards and controls); ensuring sound network design; and enabling accountability and oversight. As the needs of information users constantly evolve, Diamond also raised the importance

of developing a flexible information technology architecture that is adaptable to different users, data sources, and research methods. A vision for a 21st-century approach to information sharing for public health was illustrated through nine "First Principles for Population Health": (1) designed for decisions; (2) designed for many; (3) shaped by public policy goals and values; (4) boldly led, broadly implemented; (5) possible, responsive, and effective; (6) distributed, but queriable; (7) trusted through safeguards and transparency; (8) layers of protection; and (9) accountability and enforcement of good networking citizenship.

To convey the potential impact of a 21st-century vision for data, Diamond offered three scenarios for how decision making by providers, consumers, and policy makers might be enhanced by broader access to information grounded in reliable evidence. Realizing this vision will require moving to a new paradigm for health information in which, instead of collecting data in centralized databases for research, questions are brought to the data. Such an approach would emphasize the specific information needs of decision makers; a networked approach that supports efficient research analyses and allows data to remain distributed; and greater involvement of consumers as participants and producers of information.

U.S. Healthcare Data Today: Current State of Play

The first set of workshop sessions provided an overview of existing healthcare data—the sources, types, accessibility, and uses in the United States. In an exploration of example initiatives in the current healthcare marketplace that collect and use these data, presentations considered factors motivating the work and profiled elements of the system from different perspectives. Issues considered included the accessibility of data for new clinical insights, the extent of current uses of clinical data, and barriers to the advancement of next-generation data applications. The manuscripts in Chapter 2 reflect opportunities present within the current healthcare data profile to assess and manage clinical outcomes, as well as to glean new healthcare insights through the use of data from public and private sources.

Current Healthcare Data Profile

When discussing elements associated with evidence-based medicine or when defining the data or the taxonomies regarding health and health care, the healthcare community does not always consider all of the potential effects on health. As evidence-based medicine is more fully adopted, it will be important to evaluate all facets of evidence development and application necessary to transform health care. Simon P. Cohn, chair of the National

Committee on Vital and Health Statistics (NCVHS) and associate executive director of the Permanente Federation at Kaiser Permanente, provided an overview of the current state of healthcare data. At the broadest levels, available data inputs include biomedical and genetic factors, individual health status and health behaviors, and socioeconomic and environmental factors, along with information about health resources, healthcare use, healthcare financing and expenditures, and healthcare outcomes. In aggregate, the scope and variety of healthcare data have the possibility to significantly shape the future of research and care delivery.

Clinical data today tend to be distributed widely across healthcare systems, patients, manufacturers, and researchers. Data are fragmented rather than integrated. To achieve the goal of using data to draw on evidence for patient and provider decisions and to promote increased comparability, interoperability, and standardization of data, improved terminologies and classifications of available and future data are required. Better analytical tools, networking, and data sharing are needed to leverage access to administrative and clinical data. In addition, better means of providing meaningful data to clinicians on demand for increased integration into practice for clinical decision support will accelerate adoption. Overall, Cohn suggested that a national strategy, supported by adequate funding, is needed to address these problems and fill gaps in data use.

Recommendations made through the NCVHS are relevant to these goals. In the area of health data stewardship, for example, the NCVHS suggested that covered entities be more specific about what data will be used, how, and by whom. Recognizing that transparency is very important to consumers, the NCVHS recommends that individuals should be able to request and be given information about the specific uses and users of their data. The NCVHS also suggests that data stewardship principles should be extended to include personal health data held by non-covered entities in personal health records (PHRs) and similar instruments.

Data Used as Indicators for Assessing, Managing, and Improving Health Care

Massachusetts Health Quality Partners (MHQP) is a multistakeholder coalition that measures and reports on physician performance using health plan claims data. Speaking to the benefits and challenges of using large aggregated databases for performance measurement, MHQP Executive Director Barbra Rabson shared some of the organization's experiences using data to assess, manage, and improve health care.

MHQP aggregates healthcare data to enhance transparency in measuring and reporting on physician performance. The organization develops reports for both doctors and consumers on provider performance at the

physician network, medical group, practice site, and individual physician levels. On MHQP's website, for example, consumers can compare primary care physicians and medical groups in the state on preventive care service and chronic disease management measures. The organization also provides information for consumers on conditions, measurement, and suggestions for what both patients and clinicians might do to improve care and outcomes. Rabson observed that to date, MHQP data reporting has had greater impact on physician behavior than on consumer behavior. Massachusetts physicians have improved over the past 4 years on eight of nine measures. The public release of the data has influenced physician organizations' investments in information systems, and MHQP continues to develop strategies to engage consumers with the measures to support quality and incentives for individual physicians. Although consumers access MHQP's website, the overall impact on consumer behavior is unclear. Information gathered in consumer focus groups indicates that perhaps consumers do not always value the information made available—for example, one woman preferred to know whether a physician would be likely to deliver treatment in a patient-centered, respectful way rather than how well physicians provided breast cancer screening. Rabson noted that this suggested the need for new types of measures and data sources that provide more meaningful information to consumers.

Based on the experience of MHQP, Rabson cited some of the challenges associated with creating quality measures from electronic clinical information, including the difficult trade-offs and tensions between offering physicians flexibility to enter data and standardization of data for easier data capture; the lack of standards in data definitions and terms; the lack of standardization across vendors; and the absence of required elements from EHR data. Encrypted patient identifiers, mechanisms for facilitating patient privacy, make it difficult to provide patient-specific feedback to physicians. Ideally, clinical claims and personal data would be integrated for quality improvement; Rabson cited important work at the national level by the American Health Information Community toward this goal to define how health information technology can effectively support quality improvement. At the local level, MHQP and the Massachusetts eHealth Collaborative have been designated as the Massachusetts Chartered Value Exchange with a goal to integrate quality and HIT. Additionally, MHQP is one of six organizations selected to be part of the Centers for Medicare & Medicaid Services-funded Better Quality Information project, which involves the aggregation of claims and other clinical data from commercial payers and Medicare. MHQP is also a lead partner in a project to implement and measure the impact of EHRs in three Massachusetts communities, and to use EHRs as a data source for clinical quality measurement.

Data Primarily Collected for New Insights

Clinical researchers and epidemiologists attribute success in understanding and discovering advances in health care to the ability to collect, sort, and analyze increasingly vast amounts of numerical data. Currently, clinical and public health scientists have at least three major types of data available to them:

- Data based on clinical care that come from electronic health records, clinic-based administrative datasets, and government payer datasets;
- Large-scale registries generated and maintained by government entities, professional societies, and the private sector; and
- Clinical trials, both publicly and privately funded.

Despite the wealth of data available to researchers and policy makers, a number of major limitations hinder researchers' use of data, observed Michael S. Lauer, director of the Division of Prevention and Population Sciences at the National Heart, Lung, and Blood Institute. Although soon to change, currently relatively few American clinicians use computers to document care, and even when they do, much of the imported data are unstructured narrative text that is challenging to analyze. Most data generated today are based on nonrandomized observations drawing from the care delivery experience. Although some examples demonstrate that it is possible to incorporate rigorous and prospective data collection into routine clinical care, most clinical data are not collected at the point of care in a manner that is easily retrievable later. Access to data varies. Some datasets are widely available, while others are only available to personnel working at specific clinical sites or for specific sponsors.

As Lauer noted, the Roundtable's goal of integrating evidence-based medicine into the routine clinical practice depends on the use of data as a staple for developing scientifically sound guidelines. If the Roundtable's goal is to be realized, Lauer suggested, clinical data must be recognized as a staple that should be widely available and integrated across sites and practices. As a caution, even if a "data paradise" could be achieved with universally obtained and available clinical data, policy leaders should use care in placing too much reliability on these largely observational datasets for generating evidence-based recommendations. Even though modern statistical techniques and collection of more data elements may reduce biases, observational analyses of treatments must be recognized as inherently biased because of failure to take into account selection biases and unmeasured confounders. Lauer also emphasized the importance of well-designed experiments for building a scientific base to support evidence-based medicine.

Health Product Marketing Data

Significant amounts of data are available and used by public and private organizations to better understand public attitudes and consumer trends in the healthcare marketplace. William D. Marder from Thomson Healthcare Administrative identified three major types of data used by public and private entities to market healthcare products and services: health survey data, information about general consumption patterns, and administrative data generated by the healthcare delivery system. Much of the information about patient/consumer attitudes comes from health survey data (private and public) combined with general consumption pattern and market segmentation data. Administrative data are often used in retrospective database studies to examine the cost effectiveness of interventions in the general population. Marder described how these data comprise an information base that organizations often use to develop effective communications with the public and the business models that support collection of the data.

A number of healthcare entities engage in marketing (e.g., hospitals, pharmaceutical companies, device suppliers, government agencies, physicians) that relies heavily on information that helps target marketing to specific consumers. One example is the Thomson PULSE Survey that models healthcare use as a function of household and neighborhood characteristics. Surveying 100,000 random households per year, PULSE results are identifiable by respondents' Census tracts and linkable to other Census tract data, including socioeconomic characteristics of a particular area and lifestyle modeling done by general marketing firms. These models can be used to drive healthcare marketing and planning decisions. Marder illustrated how these resources might be used to find the best groups for clinical trial participation.

Resources that provide retail store sales, billing services, health plans, and employer-based data are sometimes incomplete, but can be useful for marketing data analyses. Claims data can be applied tactically to identify the effect of marketing campaigns and measure sales-force effectiveness. At a more strategic level, claims data can offer insights for evaluating unmet medical needs, understanding the cost of acquiring a drug in a broader context, pricing new products, gaining a favorable formulary position, and convincing prescribers about the value of a drug.

However, data collection is often expensive. For surveys such as PULSE, revenue streams to offset data collection costs come from use of the data in marketing and planning tools sold to providers and suppliers. Revenue also covers licensing of general marketing information. As for the funding of administrative data, the costs of retail and product-switch data are largely covered by pharma. Health plan and employer data are largely covered by the operations of payer organizations, with additional support from consul-

tants serving many organizations, including pharma, government, benefits consultants, and reinsurance companies.

Licensing data in a for-profit setting has certain benefits. Licensing helps customers achieve their goals by making data easy to use and to be sorted based on their interests. For license holders, the process of licensing offers the capability to recoup some costs of developing the data. A key consideration is how to best manage the intersection and interaction of private data assets and academic research. Despite the challenge of balancing the costs and benefits of making data available to researchers, Marder suggested it is essential that channels be maintained to make data available at no charge to academic audiences and to ensure access to data for replication of results.

In sum, marketing data can be seen as a synergy of inputs and interests from a variety of entities. The public sector provides raw material and models of data collection, at minimal cost. The private sector builds databases with clear commercial value that fill needs suggested by, but not covered by, public sources. As electronic medical record systems become more common, a blend of databases can be envisioned to draw on both public and private data sources—the mix will depend on government willingness to fund aggregation.

Changing the Terms: Data System Transformation in Progress

Building on workshop discussion that described the current landscape of clinical data, several presentations explored the evolution of the national data utility by highlighting efforts to coordinate clinical data—through large linked sets, aggregated data, and registries—and to make medical care data more readily available and usable. Speakers described incentives and drivers that push this evolution—including integration dynamics and disincentives, including shortfalls, limitations, and challenges of various approaches to organizing and aggregating data.

Emerging Large-Scale Linked Data Systems and Tools: The Example of caBIG

The complexity of cancer research is reflected by the many, widely differing diseases categorized as cancer. Understanding the molecular mechanisms behind these diseases is an endeavor that must involve many individuals, laboratories, and institutions across an array of specialties and subspecialties and on an international scale. The Cancer Biomedical Informatics Grid (caBIG™) aims to help provide the resources needed for such research. It was developed in response to demand at the National Cancer Institute (NCI) for a more highly coordinated approach to informatics

resource development and management. As described by Peter Covitz of the NCI, caBIG is a voluntary network of infrastructure, tools, and ideas that enables the collection, analysis, and sharing of data and knowledge along the entire research pathway—from laboratory bench to patient bedside. It was designed to speed research discoveries and improve patient outcomes by linking researchers, physicians, and patients throughout the cancer community. The program was designed intentionally to identify governance structures, organizations, and technologies that can move cancer research forward and, ultimately, have a much bigger impact on patient health then we have been able to see thus far. In its first year, for example, caBIG defined high-level interoperability and compatibility requirements for information models, common data elements, vocabularies, and programming interfaces.

The caBIG vision is to connect the cancer research community through a sharable, interoperable structure; to employ and extend standard rules in a common language; to more easily share information, and to build or adapt tools for collection, analysis, integration, and dissemination. caBIG is seen as an essential resource to fulfill the NCI's goal of eliminating suffering and death due to cancer. Moreover, Covitz suggested, the experience of designing a governance structure for caBIG as well as the nuances of the initiative's internal architecture can be instructive for healthcare data as a whole. In short, caBIG is a possible model or prototype for the broader challenge of creating an interoperable health information network across the nation.

Networked Data Sharing and Standardized Reporting Initiatives

Research questions today are more complex and data more complicated. At the same time, low-frequency events of interest demand larger pools of data, and greater geographic and demographic diversity is needed. Translational research uses institutional entities as the unit of analysis so researchers can compare outcome differences and patterns of practice. Data from single entities are insufficient, and thus the need for sharing data across research entities and collaborators continues to grow. Solutions must be tailored that are fast, inexpensive, sustainable, safe, and high quality with understood meanings. As described by Pierre-André La Chance, chief information officer at the Kaiser Permanente Center for Health Research, the Center and its research collaborators address these criteria in their data sharing efforts.

The chosen approach involves constructing research-friendly, secure, locally controlled data warehouses, as well as secure networks of local interoperable data warehouses. With controlled access to data warehouses with such characteristics, researchers can use available internal data quickly,

cheaply, and expertly. Warehoused data are of sufficiently high quality to have credibility for decisions that affect both treatment and policy. The system is able to show data quickly for more than 10 million members per month. Developers are working to create sharable versions of data that include enrollment, demographics, a tumor registry, pharmacy, vital signs, procedures, diagnosis, and laboratory values. Also in development is a biolibrary that will allow people from multiple institutions to access Kaiser Permanente tumor registries and histology data and to electronic inventories of slides that expedite the identification of appropriate participants for research studies. Developers have also focused on a specific aspect of data warehousing, creating counters or specific data marts to share deidentified data quickly; this is especially valuable for preparatory research purposes. These models are seen as valuable tactical tools that are available now to advance clinical data sharing.

Large Health Database Aggregation

Steven Waldren, director of the Center for Health Information Technology at the American Academy of Family Physicians (AAFP), provided an overview of AAFP's work to provide valued services and lower the costs for the technology. The holy grail of HIT is the ability to drive rapid improvement in the quality and safety of healthcare delivery. Yet, current financing of health care rewards high-cost, high-volume care, not low-cost, high-quality care. This disconnect creates potential conflict of interests between those who need the technology and those who will financially benefit from the technology.

As a central tool in data aggregation, the EHR can drive and support quality improvement, public reporting, health services research, clinical research, healthcare value analysis, biosurveillance, population management, and public health. In practice, however, the AAFP has found that physicians are adopting the EHR and other technologies not in the interest of data aggregation, but primarily for business support, suggesting that a paradigm shift must occur to achieve data aggregation at any level other than administrative data. Also important is the need to clarify the value proposition for those who collect clinical data. Data codification, structure, standardization, and input into systems present barriers to data aggregation; confidentiality and data privacy concerns also impede progress. These issues must be addressed to foster greater willingness to share data.

In support of data aggregation, the AAFP has worked to establish and promote HIT standards focused on clinical data, such as the American Society for Testing and Materials Continuity of Care Record standard, as well as to map individual data to a common data structure. The AAFP continues to advocate for payment reform to incentivize quality of care,

not volume of care, and works with the Ambulatory Care Quality Alliance to articulate the concept of a National Health Data Stewardship Entity. In addressing privacy and confidentiality issues, Waldren discussed AAFP's work to clarify members' misconceptions about the *Health Insurance Portability and Accountability Act* (HIPAA), which can be an unnecessary barrier to the sharing of data. Future uses for aggregated data that are of highest priority to AAFP members include quality improvement and clinical research efforts.

Registries and Care with Evidence Development

The challenges faced in data collection and knowledge dispersion to the point of care includes standardization of the language of medicine, individual confidentiality, and cooperation among professional societies. In addition, the methodology associated with information point-of-care decisions must be improved. Initiatives are needed to advance observational data adjustment techniques and to ensure that analysis of the data is unbiased because sharing of the data requires establishing trust and a common understanding between patients and providers on data issues. Finally, this is an expensive process, and allocating the expense of this to Medicare Part A or getting a fundamental base payment will be essential.

As described by Peter Smith, professor and chief of thoracic surgery at Duke University, one promising model that is impacting cardiac surgery outcomes is the Society of Thoracic Surgeons' (STS') Adult Cardiac Surgery Database (ACSD). The largest of three distinct databases that comprise the STS National Database, the ASCD is a voluntary clinical registry developed for the purpose of continuous quality improvement in cardiac surgery. It contains more than 3 million surgical procedure records from 850 participant groups, representing approximately 80 percent of adult cardiac surgical procedures performed nationally. Data are harvested quarterly, risk adjustment algorithms are updated, and each site is then provided with its raw and risk-adjusted outcomes compared to similar groups and national benchmarks. The publicly available, individual-patient STS risk calculator, based on the most recent risk adjustment algorithms, is a tool to rapidly disseminate knowledge to the bedside.

The ACSD is a clinical database that has been studied extensively and been shown to be more accurate than administrative databases. It has been selectively audited and endorsed for public use in several states. ASCD data have been linked to administrative data to demonstrate cost effectiveness of continuous improvement, and have been used to improve the accuracy of the Medicare Physician Fee Schedule nationally. In addition, over the past 7 years, STS/Agency for Healthcare Research and Quality grant programs have demonstrated that the use of a clinical data

repository and feedback can rapidly change physician behavior on a national scale. Smith noted that ultimately, the experience and success of the ASCD can be exported to inform the development of shared data in other medical specialties.

Healthcare Data: Public Good or Private Property?

Despite the potential for accelerated and expanded research offered by existing data systems and efforts to enhance their linkage and use, broader access to critical data hinge on whether healthcare data constitute a public good. As reviewed in Chapter 4, one workshop session explored this question from several perspectives. Examining the clinical data utility from a conceptual standpoint as well as from the perspectives of the marketplace and the legal system, presenters considered how the structure of the medical care data marketplace can affect research priorities, gaps, and possibilities. Questions of whether important distinctions should be made within the spectrum of data types or sources, and how a case might be made for improved access and sharing of medical data, were addressed. Several options were suggested for how to think about basic concepts related to shared data and guide their use through policy and legislation.

Characteristics of a Public Good and How They Are Applied to Healthcare Data

As understanding theoretical principles can help guide the development of practical policy and action, David Blumenthal, director of the Institute for Health Policy at Massachusetts General Hospital/Partners Health System, reviewed the classic definition of a public good and discussed how this definition applies to health information under varying circumstances. Pure public goods cannot be traded efficiently in the marketplace, he suggested, because they are both nonrival, meaning that using this good does not preclude others from using it, and nonexcludable, meaning that, even if a good is wholly owned and paid for, its use and benefit by others cannot be prevented. As an example, basic research is widely accepted in the United States—across the spectrum of ideological opinions about markets—as a public good. Basic research is considered both nonrival and nonexcludable; its support is an an appropriate and necessary role of government. In addition, Blumenthal described quasi-public goods, which he considered more relevant to discussion of healthcare data. Such goods may be relevant to the public and nonrival, but not nonexcludable or vice versa. Particularly emphasized, however, were quasi-public goods for which production or consumption generates or might generate effects (positive or negative externalities) on third parties not involved in the private purchase or sale of such

goods. Applied biomedical research has aspects of both a public good and a quasi-public good. It is excludable and rival within limits. For example, knowledge underlying a particular drug or device can be appropriated up to a point, but important information can also be kept secret and lead to benefit in the marketplace. Keeping knowledge private causes a potential loss of efficiency in the advancement of other knowledge, but Blumenthal notes that this loss is tolerated to incentivize innovation driven by opportunity for economic gain. Patent law seeks to mitigate this loss of efficiency enabling scientific progress based on protected information.

The principal questions relevant to workshop discussion concerned what to do with privately maintained databases, which are arguably quasi-public goods because they have private costs and value that given parties will not likely construct nor share out of altruism, but for which large externalities exist (i.e., if available, these data could generate significant social benefits). To realize these benefits, an approach is needed that does not eliminate losing the incentive to assemble such databases. Also relevant is another kind of informational public good. For example, data found at the National Institutes of Health (NIH) or developed through the National Health Interview Survey or National Census represent situations in which the taxpayer has paid for the information to be collected. Efforts are needed to ensure these data are efficiently made available and used.

Rationales for making data publicly available, even when not meeting the definition of a public good, apply in situations in which government has supported—through financial or other means—the development or enrichment of data or in which making the data publicly available has significant benefits not captured in traditional market transactions. Blumenthal highlighted two solutions to contending with this quasi-public good or public good nature of information. The first is to increase the appropriateness or excludability of information. Traditionally we have used patents and copyrights to accomplish this by granting a period of exclusivity on the condition of revealing the science and practice that led to the patent. The second is to have the government produce the good in question. The NIH and the National Science Foundation are examples of this approach. Blumenthal concluded that the question of data is complex, and that nuanced information uses will arise that require public guidance, perhaps on a case by case basis.

Characteristics of the Marketplace for Medical Care Data

Various sources of medical and prescription drug data are available to support safety surveillance and generation of evidence for healthcare decision makers. Repositories are constructed as potential mechanisms for research and commercial application. Data linking drug information with

medical claims data provide an opportunity to view treatments, whether by procedure or pharmaceuticals, and capture elements of the other healthcare use patterns of those patients.

Claims aggregators commonly create deidentified research databases to license to third parties, including the federal government. These databases are also licensed to academic researchers if they can afford to pay (and if they cannot, they are often given access in a spirit of good will). The largest market for these commercially licensed databases is the pharmaceutical sector, which uses them for a variety of purposes. William H. Crown, president of i3 Innovus, detailed the planning involved in the construction of large, complex datasets. Documenting many potential sources of data that can be compiled for such a purpose, Crown outlined potential barriers to data aggregation, such as adjusting and standardizing data pooled from multiple sources, protecting patient privacy, and monitoring cautions required when using data for a purpose other than the one for which the data were originally collected.

In terms of the trade-offs between a pooled mega-database and pulling data from different data aggregators, Crown indicated a growing need for a mega-database that could house data from multiple health plans, government providers and payers, as well as other sources to promote standardization and create a public good available for research, for cost-effectiveness studies, for real-world drug safety, and for guiding compliance of physician practice.

Legal Issues Related to Data Access, Pooling, and Use

The legal system enters the public good debate because it reflects and so perpetuates the current excludability state of clinical data with property and intellectual property models. Furthermore, market exchanges or shifts to public good nonexcludability face legal barriers (e.g., privacy, confidentiality, and security) that are designed to reduce or eliminate negative externalities—effects that negatively impact individuals not directly involved in the collection or use of data—suffered by data subjects. Nicolas P. Terry, Chester A. Myers Professor of Law and codirector of the Center for Health Law Studies at Saint Louis University School of Law, offered some observations on aspects of the legal system relevant to the debate of whether clinical data should be treated as a public good. These include the perceived mandate to create or support structures that treat clinical data as a private good; the design of data protection laws to eliminate or reduce potential negative externalities of data sharing on the data subjects; and the uncertainty inherent in the legal system—an indeterminacy increased by the legion of "legacy laws," such as records laws predating electronic clinical data collection and the potential for data mining of records to improve outcomes and effectiveness. Three major clusters of legal rules that

create barriers to clinical data evolving into a public good were reviewed: property or inalienability rules (ownership of medical records, and IP and trade secret protections); federal–state vectors (state restrictions on data collection, processing, or security, and state initiatives on HIT and Health Information Exchange policy); and changing data protection models (the HIPAA privacy model, and personal health records and consumer-directed health care).

Ongoing work toward solutions to challenges related to IP and data protection models were reviewed, including the notion of balancing proprietary rights in information property with public duties such as obligations of accuracy and confidentiality, and the need to to facilitate scientific, technical and educational uses of information. Terry argued that a more rigorous data protection model will be required as a predicate for greater access to patient data, noting that, as stated by NCVHS, "erosion of trust in the healthcare system may occur when there is divergence between what individuals reasonably expect health data to be used for and when uses are made for other purposes without their knowledge and permission." Examples of efforts to confront what he described as "a tension between data protection and public utility" include the National Center for Health Statistics' stewardship framework report and the European data directive.

Healthcare Data as a Public Good: Privacy and Security

In addition to proprietary issues, concerns related to privacy and security restrict the use of healthcare data. Maintaining confidential and secure data records is of paramount importance to ensuring public trust in the healthcare system, and is an important factor in discussions about sharing of health data. As presented in Chapter 5 and summarized below, four speakers considered key legal and social challenges to privacy and security issues from a variety of perspectives—including insights gained from prevailing public opinion, implications of HIPAA as a means of ensuring privacy, experiences of organizations outside health care, and the privacy and security practices of healthcare delivery organizations.

Public Views

Privacy is all-pervasive in terms of the future of HIT. Public beliefs about privacy issues link directly to the trust level that individuals have in the entire healthcare establishment, and factors significantly in the move to electronic health records, personal health records, and interoperability exchanges. Alan Westin, professor emeritus of public law and government at Columbia University and principal of the Privacy Consulting Group, presented results of a 2007 national Harris/Westin survey that measured

public attitudes toward the current state of health information privacy and security protection, health provider handling of patient information, health research activities, and trust in health researchers.

Westin's study indicates that 83 percent trusted their own healthcare providers to protect the privacy and confidentiality of personal medical records and health information. Sixty-nine percent believed researchers can be trusted to protect the privacy and confidentiality of medical records and health information on research participants. Fifty-eight percent said they do not believe there is adequate protection today for their health information when asked whether privacy of personal medical records and health information is protected enough by federal and state laws and organizational practices.

The results of Westin's survey confirm that the public has strong privacy concerns regarding the handling and protection of their personal health information, especially concerning uses of data not directly used for providing care. According to Westin, privacy is a matter of balance and judgment, and it is contextual. The results also suggest that a new code of privacy confidentiality and security written into legislation might support new health information technology and the adoption of EHR systems. In addition to encouraging models of voluntary patient control privacy policies offered through repositories of personal health records such as Microsoft's HealthVault and Google Health, Westin suggested the need for an independent health privacy audit of the verification process. New, easy-to-use technologies for implementing patient notice and choice could revolutionize the role of individuals in the process of how their personal information is used. Conducting additional field research into privacy in the EHR programs and sponsoring a national educational campaign to promote privacy-compliant, evidence-based health research might advance the public perception of data.

HIPAA Implications and Issues

As the healthcare system and HIT systems evolve, experience suggests that modifications are needed in the HIPAA Privacy Rule to strike the proper balance between protecting patient privacy and making data available for research necessary to improve healthcare quality and lower costs. Early changes to HIPAA allowed the disclosure of limited datasets and lightened administrative burdens on healthcare providers and plans that made data available for research purposes. Marcy Wilder, a partner at the law firm of Hogan and Hartson, LLP, and former deputy general counsel at the Department of Health and Human Services (HHS), served as the lead attorney in the development of HIPAA. She asserted that identifying the most significant barriers that remain, including those related to future

unspecified research and data deidentification, and clearly defining policy alternatives will be essential to promoting the research enterprise.

HIPAA rules can be confusing and require administrative recordkeeping that challenges many covered entities—in particular, smaller hospitals. There are also liability concerns on the part of the covered entities. For these reasons, researchers who seek clinical data for evidence development and application find that such data are hard to obtain. Because HIPAA is so often used as a smokescreen to preclude the sharing of data, a more difficult challenge in policy discussions will be separating out and defining for the regulators and legislators what the real problems are regarding data sharing. Wilder suggested that HIPAA as it stands today may be somewhat outdated. Advisory committees, Congress, and agencies within HHS itself have recognized that the research provisions need improvement to encourage the use of data for research and innovation. Discussions in this regard will prompt important conversations with both regulators and legislators.

Wilder identified several areas for consideration. Under HIPAA, individuals are not permitted to give their consent for the use of data for future unspecified research, which prompts an important policy discussion. In light of developments in HIT and other technology, another issue is deidentification and safe harbor standards. Another topic for discussion concerns liability burdens distributed across covered and noncovered entities.

Examples from Other Sectors

Greater openness about data can be seen in the collaborative research of the Human Genome Project, the public registration of clinical trials, and the growth of new models of disclosure/publication of research results in open-access journals and digital repositories. As the basis for the creation of new datasets underlying evidence-based medicine, greater openness is transforming the relationship between doctors and patients, increasing market incentives for improved health care, and providing new means for detecting emerging diseases. The challenge, as with privacy and clinical records, is to determine what level of openness is most appropriate for the particular purpose to be accomplished. To provide perspective on these issues, Elliot E. Maxwell, a consultant, Fellow in the communications program at Johns Hopkins University, and Distinguished Research Fellow at Pennsylvania State University, presented an overview of the adjudication of information in the context of the report, *Harnessing Openness to Transform American Health Care*, from the Committee for Economic Development.

Among other recommendations, the report suggests that the Food and Drug Administration (FDA) should review existing requirements on patient consent to participate in clinical trials and make changes as appropriate. The report also suggests that the FDA should require electronic filing for

all drug and device approvals and should set standards for, and require the filing of, underlying clinical data, upon approval, in a form that allows subsequent machine aggregation, search, and manipulation.

Regarding EHRs, the report recommends that individuals and groups providing and funding health care should institute appropriate incentives for the adoption of information and communication technologies (including EHRs) to reduce health care's burdensome administrative costs. Maxwell suggested that federal research agencies should increase their support for the development of the large databases necessary for progress toward evidence-based medicine, including development of the necessary data standards. They should also evaluate and amend HIPAA to require that those parties who hold a patient's medical records provide the patient with the opportunity to receive copies of those records in digital form pursuant to HIPAA.

Institutional and Technical Approaches to Ensuring Privacy and Security of Clinical Data

Healthcare providers view the protection and security of patient health information as essential to maintaining the trust and confidence of their patients and as an important element of patient satisfaction. At the same time, healthcare providers are rich sources of data, which have the potential to enhance the quality of clinical care and may result in better clinical outcomes, improved efficiencies, cost savings, and other medical advances. Alexander Eremia, associate general counsel and corporate privacy officer at MedStar Health, Inc., discussed the implications of these tensions from the standpoint of a major healthcare provider organization. He reflected on the institutional challenges inherent in balancing patient privacy interests with providing access for research purposes. In particular, Eremia indicated that providers must address perceived and actual privacy or security hurdles, patient trust considerations, potential legal consequences, and actual costs associated with retrieval of data; all of these pose barriers to releasing data for research purposes.

Healthcare providers may find HIPAA privacy and security requirements confusing, and health information data custodians and researchers may have limited awareness of HIPAA's data access and disclosure requirements. Furthermore, even when access and disclosure are permitted under HIPAA, the willingness to make certain disclosures of identifiable information may be impeded by physician concerns related to violating the trust of their patients, minimum necessary standards, accounting for disclosure obligations, and even concern about losing patients to physician/researchers. In addition, it is often costly for healthcare providers to divert resources and personnel away from clinical care activities to attend to system and records

access activities. As a result, healthcare providers are often more motivated to protect patient privacy, respect physician–patient relationships, minimize the administrative impact on data retrieval, and minimize legal risks and customer complaints. According to Eremia, without adequate financial or strategic incentives, regulatory amendment, and greater appreciation of the public benefits of research, access to identifiable data for research will remain a challenge.

Creating the Next-Generation Data Utility: Building Blocks and Action Agenda

The development of new data utility builds on the considerable past progress in health care. As summarized in Chapter 6, both theoretical perspectives and specific ideas for practice were presented at the workshop. Reviewed first are workshop presentations that identified lessons learned on important components or building blocks for a next-generation data utility. The chapter concludes with a summary of comments on emerging, practical opportunities to align policy developments with improved data access and evidence development offered by a discussion panel of key policy makers.

Building Blocks for the Next-Generation Public Agenda

Important strategic priorities for the development of an architecture for a next-generation data utility emerged from workshop discussion of collaborative models that offered insights into collaborative clinical data system management, and suggested a framework for expectations, purposes, incentives, priorities, structures, roles and responsibilities, and principles for data entry, access, linkage, and use. Similarly, presentation of efforts to aggregate clinical data from multiple institutions raised considerable technical, organizational, and operational challenges that need to be addressed. Finally, economic incentives and legal issues were considered as important levers to realize the full potential of health data.

Building on collaborative models. The Institute to Transform and Advance Children's Healthcare at Children's Hospital of Philadelphia (CHOP) is spearheading a novel effort to harness clinical and business information to improve children's health, make their health care more efficient, and transform the delivery system. The Institute has developed a data system that links the full spectrum of information about a child's health needs, from genomics to clinical to environmental data, in order to build out a vision of personalized pediatrics. Christopher Forrest, professor of pediatrics and senior vice president and chief transformation officer at CHOP, described

the hospital's approach to data. He discusses issues related to collaborative relationships needed to realize a vision of personalized pediatrics, including forming linkages with multiple pediatric institutions, giving patients and families access to their data, obtaining information from them, and creating provider–payer collaborations.

CHOP's model is predicated on giving care at the right time by the right person in the right setting, minimizing waste, and shifting services from specialty care to primary care. In evolving into a data-driven organization, CHOP developed a concept of personalized pediatrics, which relies on collaborations with other pediatric institutions, public institutions, payers, patients, and families. The concept of personalized pediatrics focuses on outcomes, changes in health, and reductions in costs, both financial and nonfinancial. Apart from issues related to establishing and then sustaining strong collaborations with CHOP's partners, other challenges include communications and changing cultural assumptions. Changing the culture of providers to collect data in a high-quality way is dramatically difficult: providers can be added to an EHR, but getting them to change what they do with the EHR requires education and time. Communication across the board is important, according to Forrest, especially in regard to engaging families in a dialogue about how they can partner in personalized pediatrics. CHOP's model of care is family centered and designed in partnership with families. Forrest also suggested that none of the programs will work without the support and participation of families.

Technical and operational challenges. Efforts to aggregate clinical data from multiple institutions for the purposes of gaining insights on clinical effectiveness or drug/device safety face many technical, operational, and organizational challenges. Drawing on experiences from previous pilot projects and other work in this area, Brian J. Kelly, the executive director of the Health & Sciences Division at Accenture, provided an on-the-ground, real-life implementation perspective on the challenges with aggregating data from multiple sources for secondary use. He also discussed the impact of current privacy regulations, based on work to prototype the Nationwide Health Information Network, in which researchers aimed to aggregate data from 15 completely separate organizations in four states.

Among the challenges to optimizing the use of data, both in patient care and for secondary uses, is getting the data into equivalent standards and terms, and finding ways to draw data into one repository from multiple systems. Systems for such data are in place and to some extent entrenched, and changing those systems will be incremental. Kelly drew from experiences in the area to suggest that a sophisticated approach to information governance is needed. Another consideration focuses on ownership of the data—by patients or the entity that enters the data into a database. Approaches to

addressing technological and architectural challenges are needed to best support envisioned goals for the data. Because states can place restrictions on data sharing in addition to HIPAA rules, the standard notification of privacy practices is changed to say data can be used if they are deidentified for secondary use in clinical research, but there will be continued trouble aggregating data among various institutions. To share data among delivery organizations, there could be a different approach to notification for privacy purposes, which Kelly indicated is one of the biggest policy areas that must be addressed.

There is a growing need for advocacy for using data as a public utility. Many organizations have started marketing campaigns to educate patients and their families on the importance of participation in clinical trials and related research endeavors. Kelly pointed out that we need to do the same thing to educate people on how important it is to be able to use data for secondary purposes. Such efforts would have to contend with security and privacy issues, but those factors can be addressed.

Economic incentives and legal issues. If we wish to change behavior, then we must directly address incentives, argues Eugene Steuerle, senior fellow at the Urban Institute. Steuerle suggested that existing incentive structures discourage information sharing, giving great weight to possible errors in protecting privacy relative to errors deriving from failing to take advantage of ways to improve public and often individual health. In addition, incentives internal to the bureaucracy also discourage optimal use of information, even functions such as merging already existing datasets. Because government now controls nearly three fifths of the health budget, including tax subsidies, it bears substantial responsibility to improve these incentives. Some incentive changes are possible now, through reimbursement and payment systems. Others require examining the reward structure internal to the bureaucracy. In the end, however, the primary incentive needs to come from consumer demand, operating either directly on providers and insurers or on the voters' elected representatives.

In many cases the benefits of clinical data are shared by all, but in fact the benefits to the individual come from clinical data treated as a public good. Accordingly, data-sharing solutions should examine and change the incentive structure and manage the tensions between privacy and confidentiality of data used to improve well-being. Steuerle also highlighted the failure to improve the public good when and noted evidence to the contrary. We lack data sharing for individual care; we lack data sharing for an early warning system (e.g., through the Centers for Disease Control and Prevention [CDC] or other organizations); and we lack data sharing for basically solving problems and finding cures or better treatments for various health problems. In the end, however, Steuerle encour-

aged engaging the public to support data initiatives. Another incentive problem is the lack of bureaucratic incentives to share datasets or allow datasets under an agency's purview to be shared. Notwithstanding good will among many public servants, there are strong disincentives in the bureaucracy to share data; consideration is needed on how to introduce incentives into the bureaucracy to reward people for enabling the sharing of data.

For solutions to some of these issues, Steuerle outlined several opportunities through which the government can leverage its position, for example, higher reimbursement of drugs prescribed electronically instead of through traditional methods. It could differentially pay for lab tests put into electronic form for sharing with patients or the CDC. It could pay for electronic filing of information on diagnoses and treatment. Government could also provide more incentives for participation in clinical trials. Steuerle also suggested that people working in the public sector need incentives to encourage data sharing.

The Action Agenda

Also summarized in Chapter 6 are discussions of a stakeholder panel charged with moving the conversation about data utility to an action agenda, by offering practical ideas on strategies or incentives that advance the development of an improved data utility, and what strategies or incentives might be necessary to make that happen. As session chair David Blumenthal observed, the environment for clinical data is much more distributive than ever, a phenomenon that overrides traditional instincts of policy makers to develop solutions by identifying roles and responsibilities for local, state, and federal governments. In a distributed environment, such an approach is too narrowly framed. For example, the conversation that engages consumers directly, and focuses on the personal health record, is a very different policy environment from one that could be addressed through a centralized authority. At the same time, the federal government is a big stakeholder and player in the collection of health-related data. However, the environment surrounding data differ from one part of the government to the other—the NIH, for example, has the capacity to focus on promoting sharing of data and has a broad mandate for data collection sharing, whereas Medicare operates in a much more restrictive environment. With these observations as context, panelists offered comments on decisions and actions that could best enable access to and use of clinical data as a means of advancing learning and improving the value delivered in health care.

Government-sponsored clinical and claims data. Steve E. Phurrough, director of coverage and analysis at the Centers for Medicare & Medicaid

Services, provided an inventory of the data that Medicare collects, what it does with the data it collects, and what some of its challenges are in data utility. Medicare currently collects data in each of the four parts of the program: A, B, C, and D. Collected data are used as the basis of paying claims. Different data collection programs look at how different payment systems may affect outcomes versus clinical issues. Data are collected to help improve quality of health care, for payment purposes, and to develop pay-for-performance qualitative information. Another set of data collection programs is in Medicare demonstration projects, which look at a variety of issues and, generally, examine how different payment systems may affect outcomes versus clinical issues. Data are also collected in the interest of evidence development.

Given the limits of its authority, Medicare has had to be somewhat innovative. One example is linking some clinical data to collections to coverage of particular technologies. One carrot Medicare has developed is that it has required the delivery of clinical data beyond the typical claims data as a provision for payment for certain services; a few years ago, the system required, for example, additional clinical information for the insertion of implantable defibrillators. Such an approach has the potential to provide significant amounts of information if, in fact, we can learn how to meet the challenge of what we can do with data that have been collected, and merge those data with other sources of data so that data collection can inform clinical practice.

Government-sponsored research data. The molecular biology revolution was founded on the commonality of DNA and the genetic code among living things. Discoveries at the molecular level provide unprecedented insight into the mechanisms of human disease. This understanding has developed into an expectation of wide data sharing in molecular biology and molecular genetics. Now that powerful genomewide molecular methods are being applied to populations of individuals, the necessity of broad data sharing is being brought to clinical and large cohort studies. This has prompted considerable discussion at the NIH that have resulted in the NIH Genome Wide Association Study Policy for data sharing, and a new database at the NIH's National Center for Biotechnology Information (NCBI) called the Database of Genotypes and Phenotypes (dbGaP).

James M. Ostell, chief of the NCBI Information Engineering Branch, heads the group that provides resources such as PubMed and GenBank, an annotated collection of all publicly available DNA sequences. He observed that in the course of collecting and distributing terabytes of data, the branch has wrestled with questions concerning which data are worth centralizing versus which should be kept distributed. Although technical and policy requirements sometimes dictate answers to those questions, nature some-

times directs information engineers to pursue certain tactics. For example, the commonality of molecular data might drive the desire to have all related information in one data pool, so that a researcher could search all the data comprehensively, perhaps not even with a specific goal in mind. This could lead to the kind of serendipitous connection that is fundamental to the nature of discovery. At the same time, however, there must be a balance toward collecting only those pieces of data that make sense in a universal way.

The NIH has required researchers to pool data collected under NIH grants so that other investigators might benefit from those data. NIH created dbGaP to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype. Such studies include genome-wide association studies, medical sequencing, and molecular diagnostic assays, as well as association between genotype and nonclinical traits. The advent of high-throughput, cost-effective methods for genotyping and sequencing has provided powerful tools that allow for the generation of the massive amounts of genotypic data required to make these analyses possible. dbGaP incorporates phenotype data collected in different studies into a single common pool so the data can be available to all researchers. Dozens of studies are now in the database, and by the end of 2008, the database was expected to hold data on more than 100,000 individuals and tens of thousands of measured attributes.

Hundreds of researchers have already begun using the resource. There is also a movement on the part of the major scientific and medical journals to require deposition accession numbers when they publish the types of studies alluded to above, the same as required for DNA sequence data. The publications recognize the importance of other people being able to confirm or deny a paper's conclusions, which requires investigators to review the data that informed the paper. To further encourage secondary use of data, other accession numbers are used when people take data out of a database, reanalyze the data, and then publish their analysis.

Professional organization-sponsored data. Guidelines and performance measures in cardiology developed by the American College of Cardiology (ACC), often in association with the American Heart Association, typically are adopted worldwide. ACC Chief Executive Officer Jack Lewin described ongoing efforts to ensure that ACC guidelines, performance measures, and technology appropriateness criteria are adopted in clinical care, where they can benefit individual patients. Although most guidelines are currently available on paper, the vision is to have clinical decision support integrated into EHRs.

The ACC's National Cardiovascular Data Registry (NCDR) was designed to improve the quality of cardiovascular patient care by providing

information, knowledge, and tools; benchmarks for quality improvement; updated programs for quality assurance; platforms for outcomes research; and solutions for postmarket surveillance. The NCDR strives to standardize data and to provide data that are relevant, credible, timely, and actionable, and to represent real-life outcomes that help providers improve care and that help participants meet consumer, payer, and regulator demands for quality care. The NCDR's flagship registry, the national CathPCI Registry, is considered the gold standard for measuring quality in the catheterization laboratory. Other NCDR registries collect data on acute coronary syndrome, percutaneous coronary interventions, implantable cardioverter defibrillators, and carotid artery revascularizations. The ACC is currently working to standardize registry data to be able to measure gaps in performance and adherence to guidelines, with an ultimate goal of being able to teach how to fill those gaps and thus create a cycle of continuous quality improvement.

Mandates from Medicare and states have pushed hospitals to use the ACC registries, but there is room for wider adoption. The ACC is working to alleviate barriers such as the need for standardization, the expense of collecting needed data, and the lack of clinical decision support processes built into EHRs. The ACC would also like to see a national patient identifier that would enable the tracking of an individual's overall health continuum while preserving patient privacy; such an identifier would bolster longitudinal studies. The ACC believes wider adoption of data sharing via registries is within reach, should be encouraged, and would ultimately result in better health care overall, but that strategies need to be developed and implemented that foster systems of care versus development of data collection mechanisms specific to a single hospital. Toward the development of business strategies needed to develop the clinical decision support capacity, standardization, and interoperability, the ACC wants to collaborate with other medical specialties, EHR vendors, the government, insurers, employers, and other interested parties. Going forward, the ACC supports investment in rigorous measurement programs, advocating for government endorsements of a limited number of data collection programs, allowing professional societies to help providers meet mandated reporting requirements, and implementing systematic change designed to engage physicians and track meaningful measures.

Product development and testing data. The pharmaceutical industry collects and shares a great deal of clinical data. Because the industry is heavily regulated, the data it collects are voluminous and made available publicly under strict regulations that, it is hoped, ensure their accuracy and the accuracy of their interpretations. Eve Slater, senior vice president for worldwide policy at Pfizer, noted that the pharmaceutical industry is interested

in ensuring the widespread availability of data to support research at the point of patient care and care at the point of research. In the pursuit of that goal, the industry is interested in pursuing the alignment of data quality, accessibility, integrity, and comprehensiveness. An influx of regulations and an acknowledged need for transparency are prompting the appearance of product development and testing data in the public domain. Nonetheless, attention is needed to ensure data standards, integrity, and appropriate, individualized interpretation.

Although significant amounts of product development data are required by law to be in the public domain, roadblocks prevent the effective sharing of clinical data. In the area of clinical trials posted on www.clinicaltrials.gov, for example, shared information can be incomplete, duplicative, and hard to search, and nomenclature is not always standardized. The information also needs to be translated into language that patients can understand. The lack of an acceptable format for providing data summaries for the public is linked to concerns about disseminating data in the absence of independent scientific oversight; once data are in the public domain, controlling quality assurance and the accuracy with which the information is translated to patients become difficult. Policies to address some of these issues lag behind the actual availability of data.

These issues argue in support of the data-sharing and standardization principles that the IOM has articulated. The Clinical Data Interchange Standards Consortium (CDISC) and other organizations are currently focused on the issues of standardizing electronic data.

Regulatory policies to promote sharing. Although large repositories now exist for controlled clinical trial data, including primary data, Janet Woodcock, deputy commissioner and chief medical officer at the FDA, observed that much of that information unfortunately resides on paper in various archives, not in an electronic form that would readily enable sharing. The FDA's Critical Path Initiative is an aggressive attempt to be able to combine research data from the various clinical trials in different ways and to extend learning beyond a particular research program. The FDA has been working with the CDISC to try to standardize as many data elements as possible.

Several years ago, the FDA established the ECG Warehouse, an annotated electrocardiogram (ECG) waveform data storage and review system, for which a standard was established for a digital ECG. The FDA asked companies engaged in cardiac safety trials to use that standard. Today the ECG Warehouse holds more than 500,000 digital ECGs along with the clinical data, and the FDA is collaborating with the academic community to analyze those data to learn new knowledge that would not have been accessible before the development of a standardized dataset.

The FDA is constructing quantitative disease models from clinical trials

data, building electronic models that incorporate the natural history of the disease, performance of all the different biomarkers about the disease over time, and results from interventions. Given multiple interventions, the approach allows researchers to model quantitatively. The FDA expects more of these models to evolve in the future.

Within the Critical Path Initiative, the FDA worked with various pharmaceutical companies to pool all their animal data for different drug-induced toxicities, before the drugs are given to people. This groundbreaking consortium worked to cross-validate all the relevant biomarkers in each other's laboratories. The first dataset, on drug-induced kidney toxicity in animals, has been submitted to the FDA and is under review. Similar approaches could be undertaken with humans; pooling those data from various sources could lead to new knowledge.

The FDA also plans to build a distributed network for pharmacovigilance. The Sentinel Network seeks to integrate, collect, analyze, and disseminate medical product (e.g., human drugs, biologics, and medical devices) safety information to healthcare practitioners and patients at the point of care. Required under the 2007 *Food and Drug Administration Amendments Act* (FDAAA), the Sentinel Network is currently the focus of discussions by many stakeholders about how best to proceed. One approach is to build a secure distributed network in which data stay with the data owners, but are accessible to others.

Legislative change to allow sharing. The Center for Medical Consumers, a nonprofit advocacy organization, was founded in 1976 to provide access to accurate, science-based information so that consumers could participate more meaningfully in medical decisions that often have profound effects on their health. Arthur Levin, the center's cofounder and director, believes government has a role to play in regulating the healthcare sector; key questions in this arena concern what government can and cannot do, and what it should and should not do.

Legislatively, most of the action concerning data sharing is currently in the states. Levin noted that we may face a scenario similar to that with managed care legislation, where in the absence of federal legislation, states moved ahead on their own, for better or worse. Currently states are moving ahead rapidly with HIT and health information exchange. Issues of privacy and confidentiality are very much in the forefront and driving state legislation. In terms of legislation covering data sharing, we need to make sure that whatever policy is developed moves things in an agreed-upon direction that does not create new obstacles and barriers. A first step will be to develop a much better understanding of what barriers exist in the states and federal government to aggregating data for research, quality improvement, and similar goals.

Another issue is that data sharing is, in essence, a social contract between individuals and researchers who want to use their data. Patients are told there will be some payoff from sharing data, but perhaps patients do not hear enough about how that is supposed to happen. Where does the payoff come? How does the other side of that contract deliver? What are the deliverables? Is there a time line for those deliverables? Is there accountability for those deliverables? As part of the social contract, there should be a burden on collecting data, a requirement that the collector do something specific with the data being collected. Privacy and confidentiality rules and remedies can be legislated; however, trust must be built. All who believe that data represent a public good—and that data sharing is a public responsibility to advance the public interest in improving healthcare quality, safety, and efficacy—also understand that such a message may not resonate so readily with the public. The public has not yet been brought up to that level, and more is needed to engage consumers in this enterprise.

Engaging the Public

The final session of the workshop examined the public's role in improving the clinical data utility, considering how the public currently views the use of clinical care data for research, what types of information the public is interested in deriving from such research, and how that interest might influence public response to future developments in the use of health information. The session further considered what technical, communication, and demonstration-of-value advances might help address the concerns of healthcare consumers. As summarized in Chapter 7, participants provided an overview of public knowledge, issues, concerns, and discussion of strategies on public understanding, engagement, and support for the changes necessary to create the next-generation public data utility. Also discussed were the design and implementation of tools that would be enhanced by wider availability of clinical data—such as those that help improve patient access and use of information from, about, and by those who are dealing with similar circumstances. Finally, the nature and potential use of personal health records, safeguards for data access and entry, and possible influence on public perceptions about privacy and data use were considered.

Generating Public Interest in a Public Good

In many respects, the greatest challenge associated with establishing a medical care data system to serve the public interest lies in the fact that such data largely reside in the private sector, where commercial interests and other factors inhibit sharing. This paradigm has benefited discrete entities, but it has failed to serve the public health interests of the broader

U.S. population or to promote awareness of how such information can be used to improve clinical decision making at the individual level. Though the public should have considerable interest in this information, the limitations of the data system as currently structured severely inhibit demonstration of the value proposition for consumers, both individually and collectively. Alison Rein, senior manager at AcademyHealth, identified key issues to be addressed to develop public awareness and perception of medical care data use for public good applications. She provided an overview of what little is known about this domain from the public's perspective; discussed some assumptions and attitudes that may impede progress in this direction; and highlighted examples from which we might learn and share strategies for generating public interest.

Rein discussed the public's limited understanding of how their clinical data move within or outside our fragmented system and the consequences for discussions about data access and data protection and security. Although lessons might be learned from other industries' transition to electronic systems for data management, the public expectation of trust and privacy between providers and patients, as well as the potential for irrevocable harm inherent to health care, enhance the challenge. Progress will require public education, outreach, and the demonstration of value in the use of health data.

Generating interest in electronic access to personal health information might help overcome market obstacles related to sequestering data for proprietary interests. However, Rein suggested that until greater regulation is put in place to compel providers and healthcare institutions to share data appropriately, use of clinical data for the public good will remain constrained. Efforts should also be made to align public and research interests toward pursuing common goals and helping the public develop a deeper appreciation for research as a public good. Public demonstration of the value of data sharing might help in this regard—showing, for example, the potential impact of clinical data on personal lifestyle, the bottom line, or other endpoints of interest to the public. Possible approaches to demonstrating the value of research as a public good included expanded reporting of limited, but meaningful, clinical health data to public health entities; the enhancement and expansion of clinical data registries; and the development of a nationwide health tracking network that could yield information of value to researchers, the public health community, providers, policy makers, and consumers.

Implications of "Patients Like Me" Databases

The longstanding tension between an individual's desire for personalized information and the population's interest in healthcare research is

exacerbated by scientific advances such as molecular profiling, information sharing on the web, and modern data management tools. Both the public and private sectors are struggling to navigate this logistically challenging landscape to gain medical insights and occasionally to monetize these insights. Patient-focused clinical trial information services created in the past decade provide a unique view of how patients feel about healthcare research at both the individual and the population level. Courtney Hudson, chief executive officer and founder of EmergingMed, provided an overview of EmergingMed, a company that helps cancer patients gain access to clinical trials and search for treatment options. Hudson discussed how this service addresses the intersection of an individual's need for information, access, and transparency with the U.S. healthcare system's desire for population-based research and data sharing in light of modern data management and data-sharing capabilities.

Patients in this country support mining clinical databases for the good of public health and for learning, and they believe overwhelmingly that it already happens. Patients seek information to inform treatment decisions, and Hudson indicated it would be unconscionable to not provide as much information as we have available in the public domain to possibly help each patient. As ways to use and aggregate public datasets are developed, it would be extremely difficult ethically to justify any decision to withhold information from patients. Similarly, Hudson highlighted the concept of promoting evidence-based medicine and garnering public approval and cooperation in terms of the potential benefit to the public, rather than the public understanding of research. Transparency and trust were also emphasized. The more transparent the system, the more likely patients' trust is gained. Regarding the informed consent process, a basic ethical concern is that the clinical trials system as it stands today has a narrow definition of informed consent. Hudson encouraged workshop participants to consider ways to provide context, full disclosure, or transparency to patients or to inform them about the larger process. A key distinction in considering the patient's point of view might be to view clinical data utilities in terms of patient-driven solutions versus system-driven solutions.

Implications of Personal Health Records

Dramatic increases in medical information and increases in consumer access to information via the Internet, are making health care one of the most significant hot spots for technology innovation today. Currently the practice of medicine suffers from an information management problem. Control will eventually shift, moving the current top-down doctor-patient relationship to one that is characterized by mutual control. For physicians, the issue is about aggregating data within and across provider organiza-

tions, and for consumers it is about aggregating health data across all of their sources. Ultimately, these views will connect to enable informed health decisions and better clinical outcomes. Today, we have more personal health data than ever; however, the data are dispersed over a variety of facilities, providers, and even our own monitoring devices and home computers.

As described by Jim Karkanias, partner and senior director of applied research and technology at Microsoft Corporation, Microsoft is working to address gaps in the healthcare data management system, both from an enterprise and a consumer standpoint, to enable a more connected, informed, and collaborative healthcare ecosystem. Microsoft HealthVault, a consumer health platform with specialized health search capabilities, delivers a platform that puts users in control of their information so they can access, store, and recall it on demand. Karkanias indicated that such a level of access and control contributes to the ability to make good decisions. The platform is built on the premise that the consumer is at the center of health care, so patients are the logical aggregators of this information.

HealthVault seeks to help patients to proactively manage their own health care—substituting, for example, costly visits to a doctor's office with daily in-home monitoring to allow for proactive measures to be taken as they can be detected. Chronic conditions and more serious illnesses could be handled proactively. With appropriate privacy consents, a caregiver could have a full view of a patient's underlying data; others could be granted access to different parts of that same data—an approach useful, for example, to adult children caring for their parents from afar.

CLINICAL DATA AS THE BASIC STAPLE OF HEALTH LEARNING: IDEAS FOR ACTION

The availability of timely and reliable evidence to guide healthcare decisions depends substantially on the quality and accessibility of the data used to produce the evidence. Important information about the results of different diagnostic and treatment interventions is collected in multiple forms by many institutions for different reasons and audiences—providers, patients, insurers, manufacturers, health researchers, and public agencies. Medical care data represent a vital resource for improving insight and action for more effective treatment. With the increasing potential of technical capacity for aggregation and sharing of data while ensuring confidentiality, the prospects are at hand for powerful and unprecedented tools to determine the circumstances under which medical interventions work best, and for whom. However, these data are usually held in a proprietary manner instead of being considered a public good that can be pooled and mined for new research and, ultimately, better patient care and outcomes. There are a number of challenges to the use of such

data—coding discrepancies, platform incompatibilities, patient protection tools—yet practical approaches are and can be developed to contend with these issues. The most significant challenge may be the barriers and restrictions to data access inherent in treating clinical outcome data as a proprietary commodity.

Chapter 8 summarizes the themes emerging from workshop discussion and opportunities for follow-up action by the Roundtable. Key issues discussed include clarifying basic principles of data stewardship; creating next-generation data utilities and models; creating next-generation data policy; and engaging the public. Potential opportunities for follow-up attention by the members of the IOM Roundtable on Value & Science-Driven Health Care include those noted below—Roundtable Innovation Collaboratives already engaged in related follow-on work are indicated in parentheses.

1. *Principles:* Foster the development, review, and implementation of basic principles for data stewardship.
2. *Use of electronic health records for knowledge development:* Convene an affinity group of EHR users and vendors to consider approaches to cooperative work on knowledge development, including issues related to standards and rules for governed data query and application (EHR Innovation Collaborative).
3. *Collaborative data mining:* Organize exploratory efforts to investigate cutting-edge data-mining techniques for generating evidence on care practices and research (EHR Innovation Collaborative).
4. *Incentives:* Convene an employer–payer workgroup to explore the use of economic incentives to reward providers/groups working to improve knowledge generation and application in the care process.
5. *Privacy and security:* At the conclusion of a current IOM study on HIPAA and privacy protection regulations, convene a series of meetings to explore and clarify definitions as well as reduce the tendency toward unnecessarily restrictive interpretations, in particular as they relate to data sharing and secondary uses.
6. *Transparency and access to federal data:* Explore the marketplace for data, opportunities to enhance data sharing, governance/stewardship issues, and ways to make federally sponsored clinical data widely available for secondary analysis. This includes not only data from federally supported research but also Medicare-related data, including from Part D (pharmaceutical) use.
7. *Public involvement in the evidence process:* Engage the public through communication efforts aimed at increasing public understanding and involvement in evidence-based medicine (Evidence Communication Innovation Collaborative).

REFERENCES

Arrow, K., J. Bertko, S. Brownlee, L. P. Casalino, J. Cooper, F. J. Crosson, A. Enthoven, E. Falcone, R. C. Feldman, V. R. Fuchs, A. M. Garber, M. R. Gold, D. Goldman, G. K. Hadfield, M. A. Hall, R. I. Horwitz, M. Hooven, P. D. Jacobson, T. S. Jost, L. J. Kotlikoff, J. Levin, S. Levine, R. Levy, K. Linscott, H. S. Luft, R. Mashal, D. McFadden, D. Mechanic, D. Meltzer, J. P. Newhouse, R. G. Noll, J. B. Pietzsch, P. Pizzo, R. D. Reischauer, S. Rosenbaum, W. Sage, L. D. Schaeffer, E. Sheen, B. M. Silber, J. Skinner, St. M. Shortell, S. O. Thier, S. Tunis, L. Wulsin, Jr., P. Yock, G. Bin Nun, S. Bryan, O. Luxenburg, and W. P. M. M. van de Ven. 2009. Toward a 21st-century health care system: Recommendations for health care reform. *Annals of Internal Medicine* 150(7):493–495.

Blumenthal, D., E. G. Campbell, M. Gokhale, R. Yucle, B. Clarridge, S. Hilgartner, and N. A. Holtzman. 2006. Data witholding in genetics and the other life sciences: Prevalence and predictors. *Academic Medicine* 82(2):137–145.

Detmer, D. E. 2003. Building the national health information infrastructure for personal health, health care services, public health, and research. *BMC Medical Informatics and Decision Making* 3(1):1–40.

Editorial. 2005. Let data speak to data. *Nature* 438:531.

Health care spending in the United States and OECD countries. 2007. http://www.kff.org/insurance/snapshot/chcm010307oth.cfm (accessed July 14, 2008).

Hrynaszkiewicz, I., and D. Altman. 2009. Towards agreement on best practice for publishing raw clinical trial data. *Trials* 10(17).

IOM (Institute of Medicine). 2009. *Beyond the HIPAA Privacy Rule: Enhancing privacy, improving health through research*. Washington, DC: The National Academies Press.

Ness, R.B. 2007. Influence of the HIPAA Privacy Rule on health research. *Journal of the American Medical Association* 298(18):2196–2198.

NRC (National Research Council). 2009. *Computational technology for effective health care: Immediate steps and strategic directions*. Washington, DC: The National Academies Press.

Piwowar, H. A., M. J. Becich, H. Bilofsky, and R. S. Crowley. 2008. Towards a data sharing culture: Recommendations for leadership from Academic Health Centers. *PLoS Medicine* 5(9) e183.

Safran, C., M. Bloomrosen, W. E. Hammond, S. Labkoff, S. Markel-Fox, P. C. Tang, D. E. Detmer, with input from the expert panel. 2007. Toward a national framework for the secondary use of health data: An American Medical Informatics Association white paper. *Journal of the American Medical Informatics Association* 14(1):1–9. http://www.healthlawyers.org/Members/PracticeGroups/HIT/Toolkits/Documents/5_Health_Data_AMIA_Summary.pdf (accessed August 18, 2008).

1

Clinical Data as the Basic Staple of the Learning Health System

INTRODUCTION

A modern evidence and value-driven healthcare system must have the capacity to learn and adapt—to track performance in real-time and generate and apply information for future improvements in safety, quality, and value of care received. As information technologies supporting clinical documentation continue to advance, the volume of clinical data generated in the natural course of care rapidly grows. Understanding, accessing, managing, and interpreting the widening variety of healthcare data available requires coordination of resources, efforts, and incentives to ensure that researchers, clinicians, and patients have access to the right data, in the right context, at the right time (Detmer, 2003; Kawamoto and Ginsburg, 2009; NRC, 2009). Integrated datasets and other approaches to link data and broaden or share findings only extend the potential to use these data to learn what works in health care. Fostering broader access to and appropriate use of these data will be key to progress—and will require both cross-sector discussions to better characterize the technical, organizational, and legal barriers that currently limit the use of existing and emerging data resources and cooperative action to address these challenges (Arrow et al., 2009; Piwowar et al., 2008). These and other issues were the focus of discussion at the Institute of Medicine (IOM) Roundtable on Value & Science-Driven Health Care's February 2008 workshop, Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good (Box 1-1).

BOX 1-1
Issues Motivating Discussion

1. Discovering what works best in medical care—including for whom and under what circumstances—requires that clinical data be carefully nurtured as a resource for continuous learning.
2. Transformational opportunities are presented by evolving large and potentially interoperable clinical and administrative datasets.
3. Clinical data are recorded and held in multiple activities and many institutions, including medical records, administrative and claims records, and research studies.
4. Public policy and public awareness lag behind the technical, organizational, and legal capacity for reliable safeguarding of individual privacy and data security in mining clinical data for new knowledge.
5. A significant challenge to progress resides in the barriers and restrictions that derive from the treatment of medical care data as a proprietary commodity by the organizations involved.
6. Even clinical research and medical care data developed with public funds are often not available for broader analysis and insights.
7. Broader access and use of healthcare data for new insights require not only fostering data system reliability and interoperability but also addressing the matter of individual data ownership and the extent to which data central to progress in health and health care should constitute a public good.

The Roundtable and Clinical Data

The IOM's Roundtable on Value & Science-Driven Health Care provides a trusted venue for key stakeholders—patients, health providers, payers, employers, manufacturers, health information technology, researchers, and policy makers—to work cooperatively on innovative approaches to generating and applying evidence to drive improvements in the effectiveness and efficiency of medical care in the United States. Participants seek the development of a learning health system that is designed to generate and apply the best evidence for the collaborative healthcare choices of each patient and provider; to drive the process of discovery as a natural outgrowth of patient care; and to ensure innovation, quality, safety, and value in health care. They have set a goal that, by the year 2020, 90 percent of clinical decisions will be supported by accurate, timely, and up-to-date clinical information, and will reflect the best available evidence.

Central to fulfilling the Roundtable's goal is a change in how the healthcare system is structured to capture and apply the results of clinical experience. This publication, *Clinical Data as the Basic Staple of Health Learning*, summarizes the workshop's examination of the current national profile of healthcare data sources; the tools and datasets employed to transform data linkage and application; the notion of clinical data as a public good; and the legal and social elements of data privacy and security. Through invited presentations, workshop participants explored these issues and discussed possible next steps in the creation and maintenance of the next generation of data utility.

Overview and State of Play

Composed of information ranging from determinants of health (e.g., biomedical, demographic, and genetic factors; health behaviors; socioeconomic factors and environmental factors) and measures of health and health status (e.g., laboratory data, physical exam findings, imaging studies, diagnoses, treatments prescribed, responses to interventions applied) to documentation of care delivery, healthcare data in the United States are distributed widely across the healthcare system. These data are captured as part of the delivery of clinical care, administration and claims processes, and research. Stored across the country—in personal and electronic health records, paper charts, claims receipts, and research registries in office practices, hospitals, academic medical facilities, insurance companies, and research labs—these data may represent discrete test results or information from handwritten notes about the interaction between a healthcare provider and a patient. They are collected and maintained by organizations supporting these activities in numerous databases (NRC, 2009).

Healthcare data and databases are used for many purposes. Patients, providers, payers, researchers, and government registries collect health information with the goal to assess and improve care provision and treatment, advance discovery and research, direct reimbursement, develop the evidence base for medical practice, and inform public health and health reform policy development. Progress in health information technology and analytic tools have dramatically expanded our capacity to capture and use these data. However, few sources, taken individually, provide comprehensive, longitudinal views about individual patients or have data in sufficient numbers to adequately power studies of safety and effectiveness. Instead, data that collectively could provide a picture of individual and population health, advance our understanding of what works in practice, and improve health outcomes are fragmented across a complex system of collection and storage. Potential exists for these data to be used to fill substantial knowledge gaps in health care, including research on best practices, reducing costs,

increasing quality, and on effectiveness of medical interventions in clinical practice (Hrynaszkiewicz and Altman, 2009; IOM, 2009; Kawamoto and Ginsburg, 2009; Safran, 2007).

As detailed in the chapters that follow and briefly summarized here, complicating the use of these data are technical, cultural, and legal barriers. Terminology used for data collection in different organizations and sectors are not standardized and numerous systems of electronic data collection lack interoperability, making synthesis and comprehensive analysis of pooled data a tremendous challenge (NRC, 2009). While such technical barriers will require collaboration by the stakeholders involved in issues relating to clinical data, issues around data ownership (including societal concerns about privacy or the treatment of data as a proprietary commodity) pose significant challenges to realizing the full potential of clinical data as the basic staple of a learning health system (Piwowar et al., 2008).

The advent of electronic health information technology as a means for collecting, housing, and analyzing clinical data has prompted concern about who has access to what data and for what purpose. With the goal of ensuring the protection of an individual's privacy while still permitting information exchange necessary for providing appropriate clinical care and research, the Department of Health and Human Services (HHS) developed a set of federal standards for protecting the privacy of personal health information under the 1996 *Health Insurance Portability and Accountability Act* (HIPAA) (IOM, 2009). Certain provisions raised concerns among healthcare institutions, research entities, and providers about compliance and among patients about privacy and security of these data. The HIPAA Privacy Rule in particular has been the focus of much discussion about data sharing for both clinical and research endeavors, and some have suggested variable interpretations of this rule have hampered important health research (IOM, 2009; Ness, 2007).

In February 2009, the IOM released a report, *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research,* authored by an IOM consensus committee charged in part with proposing "recommendations to facilitate the efficient and effective conduct of important health research while maintaining or strengthening the privacy protection of personally identifiable health information" (IOM, 2009). The report characterizes the tension between individual privacy concerns and potential societal benefits reaped from sharing of clinical data as follows: "The primary justification for protecting personal privacy is to protect the interests of individuals. In contrast, the primary justification for collecting personally identifiable health information for health research is to benefit society. But it is important to stress that privacy also has value at the societal level because it permits complex activities, including research and public health activities, to be carried out in ways that protect individuals'

dignity" (Ness, 2009). The report notes several examples of important findings derived from medical research databases (Box 1-2), and suggests that the opportunities will only expand with health information technology advancement. The committee's recommendations aimed at promoting both enhanced privacy and research are presented in Appendix D.

While technical issues such as interoperability and standards as well as privacy concerns have hampered efforts to share and utilize clinical data, many have observed that the needed shift to a data sharing culture—among scientists, clinical researchers, and health organizations—might pose a greater challenge (Altman, 2009; Blumenthal et al., 2006; *Nature*, 2005; Piwowar et al., 2008). Critical to promoting such a culture are clarification on roles and responsibilities with respect to clinical data; viewing the development of incentives, guidance, and appropriate requirements as critical to promote such a culture; and leadership from all sectors in health care.

The aims of the workshop and this publication are to provide an overview of these issues; to survey some of the current, potentially transformative research and clinical data initiatives under way; to discuss the notion of public and private goods; to consider implications of privacy, security, and proprietary concerns; and to suggest some possible opportunities to encourage a data sharing culture and the engagement of the public in advancing progress to the next generation of clinical data resources.

Perspectives on Clinical Data and Health Learning

To build a foundation for the presentations that would follow, each of the two days of the workshop began with a keynote address designed to take a broad look at relevant issues. The first day's keynote speaker was David Brailer, chairman of Health Evolution Partners. As the nation's first national coordinator for health information technology, Brailer led federal and private-sector efforts to improve healthcare quality, accountability, and efficiency through health information technology (HIT) and create a strong foundation for the adoption of digitalized medicine in the United States. His keynote presentation profiled current collection and use of clinical data and reflected on how these data might be used in the near future in terms of care delivery, research, and health outcomes. The second day's keynote speaker, Carol Diamond, managing director of health programs from the Markle Foundation, presented a vision for future health care in which clinical data are treated as a public good as a way to illustrate current technical and policy challenges. Her remarks explored three key questions: What might be achieved if clinical data could be positioned as a public good? How would such a system work, and what are the technical and policy issues to engage in fostering its evolution? Do we want to define integrated data as a public good?

BOX 1-2
Examples of Important Findings from Medical Database Research (adapted from IOM, 2009)

Herceptin and breast cancer: Data were collected from a cohort of more than 9,000 breast cancer patients whose tumor specimens were consecutively received at the University of San Antonio (1974–1992) from across the United States. Results showed that amplification of the HER-2 oncogene was a significant predictor of both overall survival and time to relapse in patients with breast cancer. This information subsequently led to the development of Herceptin (trastuzumab), a targeted therapy that is effective for many women with HER-2–positive breast cancer.

Folic acid and birth defects: Medical records research led to the discovery that supplementing folic acid during pregnancy can prevent neural tube birth defects (NTDs). Studies in the 1970s found that vitamin (folate) deficiency and use of anticonvulsive drugs that deplete folate were associated with higher rates of NTDs, and studies in the 1980s found that use of folate supplements was associated with decreased rates. Population-based surveillance systems showed that the number of NTDs decreased 31 percent after mandatory fortification of cereal grain products.

Effects of intrauterine DES exposure: Starting in the 1940s, diethylstilbestrol (DES) was used by millions of pregnant women to prevent miscarriages and other disorders in pregnancy. In the 1970s, retrospective studies of medical records began to show that infants exposed to DES during the first trimester of pregnancy had an increased risk as adults of breast, vaginal, and cervical cancer as well as reproductive anomalies. In November 1971, the Food and Drug Administration (FDA) sent a *FDA Drug Bulletin* to all U.S. physicians advising them to stop prescribing DES to pregnant women.

Patient safety: Health services research estimated that tens of thousands of Americans die each year from medical errors in the hospital. A 1998 study led by David Bates (Brigham & Women's Hospital) found that computerized order entry of prescriptions at Brigham & Women's Hospital reduced medical error rates by 55 percent; rates of serious errors fell by 86 percent. In response to this groundbreaking work, hospitals around the country are installing their own computerized physician order entry systems.

Mortality risks of antipsychotic drugs in the elderly: In 2005 the FDA issued a public health advisory stating that the atypical (second generation) antipsychotic medications increase mortality among elderly patients. This decision was based on the results of 17 placebo-controlled trials with such drugs that enrolled a total of 5,106 elderly patients with dementia who had behavioral disorders. Risk of death with older, conventional agents was not known. Results from two subsequent retrospective reviews of 27,000 and 37,000 medical records of elderly patients indicated that conventional antipsychotic medications are at least as likely as atypical agents to increase

the risk of death among those patients. As a result, the FDA now requires that the prescribing information for all antipsychotic drugs contain the same information about risks found in the *Warnings* section.

Child safety: Using the Partners for Child Passenger Safety (PCPS)—an ongoing child-focused, real-time, crash surveillance system established with the State Farm Insurance Companies in 1997—Flaura Winston (Children's Hospital of Pennsylvania) found that only 25 percent of children between 3 and 7 years of age were appropriately restrained in crashes; children in seat belts alone were at a 3.5-fold increased risk of serious injury. Winston's analysis of PCPS data led to the rapid adoption of belt-positioning boosters as the appropriate form of restraint for children once they have outgrown car seats. Appropriate restraint by children in this age group has doubled, and child fatality from crashes is at its lowest level ever.

Obesity: Eric Finkelstein (RTI International) used data from the late 1990s to find that obesity is responsible for up to $92.6 billion in medical expenditures each year; approximately half of obesity-related healthcare costs are borne by Medicare and Medicaid. A 2002 study by Roland Sturm (RAND) found that the effects of obesity on a number of chronic conditions were larger than those of smoking or problem drinking. Since then, obesity has been escalated to the top of the list of health care priorities, and policy makers have appropriated funds for federal agencies to fund health services research that encourages people to understand the effects of diet and exercise on their health.

Rural health: Stephen Mick (Virginia Commonwealth University) and colleagues examined rural hospital performance in the late 1980s and early 1990s. They found that activity typical of urban hospitals is beyond the capacity of most rural facilities and recommended that a new federal approach would be required to preserve rural acute-care services. This work helped form the intellectual basis for Medicare's highly successful Critical Access Hospital program, which was designed to improve rural healthcare access and reduce closures of hospitals that provide essential community services.

Workforce and health outcomes: In 1997, Jack Needleman (University of California–Los Angeles) and Peter Buerhaus (Vanderbilt University) analyzed more than 6 million patient discharge records from 799 hospitals in 11 states. They found that patients in hospitals with fewer registered nurses stay hospitalized longer and are more likely to suffer complications, such as urinary tract infections and upper gastrointestinal bleeding. This research established a causal link between the nursing shortage and outcomes, and helped move the nursing shortage into the public's eye.

SOURCES: Bates et al. (1998); FDA (1971, 2005, 2008); Finkelstein et al. (2003); Gill et al. (2007); Herbst et al. (1971); IOM (2000b); Mick et al. (1994); Needleman et al. (2002); Pitkin (2007); Schneeweiss et al. (2007); Slamon et al. (1987); Thorpe et al. (2004); Veurink et al. (2005); Winston et al. (2000).

CLINICAL DATA AS THE BASIC STAPLE OF HEALTH LEARNING

David Brailer, M.D., Ph.D.
Chairman, Health Evolution Partners

The idea that clinical data are a basic staple of the learning system is perhaps one of the least appreciated and most important aspects of American health care. It is about more than clinical data per se because the need for data is obvious. At its core, it is about whether clinical data are a public or private good. A significant gap exists between our desire to use clinical information to improve health care and the reality that we see in today's healthcare market, where clinical information is proprietary and used for strategic benefit.

To view this challenge in its broader context, one must take a step back. The United States is well into the big step of health information technology adoption. In 2004, President Bush declared the goal that most Americans would have access to electronic health records (EHRs) in 10 years, and to achieve this goal, he created the Office of the National Coordinator for Health Information Technology. This step struck a chord of resonance with Americans who view change in the healthcare industry as necessary, if not inevitable. The concept of health information technology is not new. It follows at least 30 years of work that preceded our progress today. The work followed the publishing of the seminal papers on how information given to a clinician at the right time and at that teachable moment can have a profound effect on care.

This goal is a significant challenge to the status quo of health care, and we have made some good progress toward meeting it. Our nation is far along in efforts to develop standards for interoperable HIT and to certify HIT products that meet minimum standards. In health care, standards development and certification are the equivalents of cellular handsets and wireless connectivity for telecommunications. They are the basic building blocks for producing portable health information. Many hospitals are far along with putting electronic records in place. Many physicians are still struggling, but we are still seeing signs of incremental progress in adoption among physician groups, particularly large ones. Most importantly, the public has become aware of health information as an asset or a good. Consumer awareness is a critical foundation that is necessary for ensuring that HIT serves as an enabler for portable clinical information. Today, several insurers are differentiating their services in national markets based on their solutions for health information access. Companies such as Microsoft, Google, and others are beginning to offer health information access as part of their industry's vertical solutions. The United States has made enormous progress because the public has begun to appreciate how important infor-

mation is in their care experience. As the American public begins to want and demand more, the direction of change will sharpen and accelerate.

Yet numerous open issues remain. We are still facing enormous difficult incentives. Misaligned incentives have evolved from our perilously obsolete reimbursement system. We still have a privacy paradigm from a paper age. We do not have a framework for privacy that recognizes that health information is no longer a static good, but instead is a portable, moving, compounding, and growing asset. We do not have even the consciousness to understand what we do about this privacy challenge because most are still focused on what did and did not work under the old paper-based privacy statutes.

In spite of these obstacles, person-based portability of health information that truly moves along with the patient across healthcare settings is closer than it was. We have seen some great examples at the national level through the American Health Information Community, the advisory council of former Department of Health and Human Services Secretary Michael Leavitt, and at the state level with Regional Health Information Organizations. However, interoperable health information is still a novelty and the exception rather than the rule. New information-sharing efforts have encountered challenges in taking the health information assets that we have produced, whether they are at a regional level or in a big healthcare system, and truly used them as a health improvement asset. The true test of health information interoperability is whether the information is truly useful to clinical care.

Using the prescription process as an example, there are two large needs for improving prescription use, which include (1) the indication of why a patient is taking that drug, and (2) a termination order recognizing that the drug was stopped or another drug was substituted. Just those two simple examples of why the drug was given and when and why it was stopped have impaired most analyses of prescribing patterns, and we cannot think of how to put this information into the workflow of a doctor in a way that does not cause disruption and backlash. We can tell that story again with respect to laboratory information, and referrals to specialists, or within genomics.

Although we face numerous challenges, the fundamental tension that must be navigated is between adoption and interoperability. Interoperability—the capacity to share, integrate, and apply health information from disparate sources—was the principal priority of the nation's health information agenda from 2004 forward, but the adoption agenda to push health information tools into point-of-service use is now beginning to overtake the interoperability agenda. Those two goals are in conflict because we lack all of the components necessary for EHRs and other information tools to be able to share information in a way that achieves our goals for the learning health system. The Office of the National Coordinator chose to put interoperability first to

take advantage of that lack of a legacy. This has been viewed as a one-time opportunity that required purposeful restraint that would push adoption, but not make it a relentless drive during the early part of the President's 10-year agenda.

Health information is a key vehicle for changing the healthcare system, but how do we create the data or evidence? How do we actually get assembled, coherent, representative, timely, and valid health information that can inform decisions at a patient level or at a broad population level or even at a very large population level?

The data and evidence are coming together. We know how to make this work. We do not lack knowledge in how to create a compiled, intelligent, useful, analytically sound, and interpretable set of clinical information. Today, we mostly know how to do this in laboratory experiments in very controlled circumstances, but many industries outside of health care have demonstrated the ability to take the artifacts of production—such as the information spun off from cars being made or financial services being offered, or some combination of those—and turn that information into data that can help manage workflow, manage processes, and identify opportunities for improvement or opportunities for failure. Establishing this "intelligence" does not happen as a single event because it is a generational shift, and this current generation of basic point-of-service transactions will probably be inadequate as infrastructure to take us far in health care.

The learning system is where actions are accompanied by feedback that is linked to accountabilities, whether they are incentives or changes in action. This occurs through an integrated process. Information drawn from actual experience in care delivery must be able to shape the care delivery process. Specifically, the data that inform our policy or inform population care should not be separate from what is really going on in care.

The learning system requires data that are structured, meaningful, representative, and duplicable in a way that supports consistent interpretations and conclusions across many different episodes. Today, even diagnosis codes do not always mean the same thing because they are used differently for the billing process in different kinds of organizations, and yet we want to be able to have those data artifacts compiled and used comparatively.

Finally, we want to have the means of evaluating the system so that we can translate findings into accountabilities and responses. That simple ability to drive information through a process is clearly what is required for clinical information to be used effectively in a learning system.

Many healthcare organizations of varying sizes are looking at this agenda and seeking leadership. They are looking for clarity on how they should go about compiling, analyzing, structuring, and creating accountability, but this effort cannot be teased out from the broader issues that are shaping health care. Thinking about ways of making health care smarter is

increasingly difficult when there are fundamental problems that providers, payers, and consumers are facing in terms of access and cost. To evolve a learning system in the midst of these competing near-term problems, we must relentlessly continue to pursue things that make clinical data structured, intelligent, useful, assembled, and applied in a way that makes care better. When we as actors begin to address this today, we go back to the foundational question: Who is going to do that? Who will control it and to what end?

This is not an easy question and the outcome is unknown. Under one scenario, health information could become a true public good as something that is truly nonproprietary. Under another scenario, clinical information could become a private good as something that is used differentially, for comparative advantage that benefits some, but not all. The reality today is that clinical information is largely a private good. Whether it helps or harms health care is an unanswered question.

The old English common-law adage that possession is nine-tenths of the law was originally applied to real property, largely to land. It was a rule of logic, as most old English common-law was, that applied to disputes about ownership. Figuratively speaking, the rule of nine-tenths applies to health information. Nothing in federal or state statutes, regulation, or other guidance says to providers or to any other data originator—a lab, hospital, physician, or device manufacturer—that they control the health information they produce. Yet in aggregate, the confluence of rules and business practices largely give nine-tenths of the benefit to data producers to control health information.

For example, HIPAA creates de facto control over health information by providers. First, patients cannot direct that their information be sent to a third party. Although some providers do this as a courtesy to patients, others still do not. The law is very clear that no provider is required to send a patient's health information to anyone other than the patient. This results in a barrier to true portability where agents acting on behalf of people to compile and move their information are at a disadvantage compared to the data originator.

Second, providers are not obliged to make data available in formats or through modalities that are not convenient to the data producer. There is a very good, paper-era reason for this rule. Such a requirement might have imposed a tremendous cost of infrastructure conversion or information technology on providers during the paper paradigm when HIPAA was established. We are in a different world today, although we do not have the ability to actually get information in a raw, useful, assembled analyzable format.

Third, providers have a long period of time to comply with the data request—as long as 120 days in most states—which makes shared information useless to most patient care.

Because of these limitations imposed by law regarding time, format, and distribution, we live in a world today where providers clearly own and control health information. This adversely affects its portability and it makes it hard to address the kinds of goals we have for it as a public good or as a staple of a learning system. These are only three limitations that have been identified. There are other regulatory barriers to portable, available, and acceptable health information.

Why does this matter? In our healthcare delivery system today, strategic use of health information is anchoring quasi-geographic cartels. So, much of healthcare delivery is controlled by a single physician group, hospital or integrated system, lab, specialty group or alliance, specific imaging center, etc. This is not apparent if one looks at a Herfindahl index—a metric of market concentration. However, the Herfindahl index and other market concentration tools are limited to understanding the micromarkets that occur in neighborhoods of health care.

The American public is clear that the primary attribute they want with health care is geographic access. Many factors affect access and choice of providers, and health information is one of them. In many instances, consumers stay with their existing providers because they are concerned that their health information could be lost if they move around, thereby reducing their ability to shop, which is an ability to exercise the kind of choice that should exist. This lack of choice hinders this emerging consumer force that has shaped other industries, but has been slow to assemble in health care. Whether in retail or technology, only a small share of consumers seeking transparency and value has driven the characteristics of products that the majority of other consumers use. Health care will eventually be the same, but in the near term, that small share of consumers is unable to exercise meaningful choice because their health information is difficult for them to get or to use.

The antithesis of transparency is proprietary health information. As consumers seek health care that is transparent and accountable, they are striking at the heart of health information as a source of competitive advantage in the marketplace. Many health systems, whether for profit or nonprofit, are seeking ways to use health data to maintain high price points and to differentiate their products and services. Proprietary use of health information will necessarily become a key component of their success.

As an example, one for-profit healthcare company stated the following in its S-1 (Initial Public Offering) filing with the Securities and Exchange Commission: "We have developed proprietary methods of care that are protected by patent and that cannot be easily replicated because of our unique information technology capabilities and use of health information." This is one of many healthcare companies that is making substantial investments in health information technology to drive long-term profitability. This raises

the important question of how we ensure that health information is not used in a way that creates a quality gap where some have access to it and others do not.

Another example may find more resonance. If pay for performance becomes a key means for producing revenue for our healthcare system, the system will be required by bond underwriters and others to secure the means of the revenue production. A whole subindustry has evolved around this, which includes two of the higher profile healthcare initial public offerings of 2007. Both of these companies are focused on revenue maximization or, as they call it, revenue cycle management. One focuses on physicians and the other focuses on hospitals. If revenue production is shifted away from volume and toward performance, these organizations will be obliged to create means by which they secure their performance-based revenue. This places enormous pressure on healthcare systems to protect their intellectual property and trade secrets on how they deliver care and the ability to understand which approaches to treatment work and which do not.

The future of an industry in which health information is proprietary does not need to happen. Providers should be able to gain advantage from being performance driven, yet they should not gain advantage from health information that is exclusive to them as a private good. Health information as a public good can enable competitive differentiation while supporting the evolution of the learning system as well as true consumer choice, which happens only when information follows the consumer.

There is a limited window to achieve these interoperability goals that underlie the learning health system. This brief period can allow us to ensure that the unintended consequences of policies that result in the nine-tenths possession rules are not allowed to continue. This opportunity is not only for the development of information standards, but also for their requirement as a part of our health policy that the American public understands and demands for portable health information.

Therein lies the challenge for the use of clinical data as a staple of the learning health system. It must be more useful, specified, efficient, assembled, and valuable. It must be equitable and not something that can benefit just the well financed and the well organized. It must also continue to keep the healthcare marketplace competitive.

Ultimately, whether we have a learning health system depends on whether health information is a public good. The ultimate test of this will be if health information does in health care what the information revolution has done in every other industry, which is to push power away from the large institutional providers of services—financial service companies, airlines, the media, hospital, insurers, or labs—and out to consumers.

VISION FOR THE FUTURE: CREATING A PUBLIC GOOD FOR THE PUBLIC'S HEALTH

Carol C. Diamond, M.D., M.P.H.
Managing Director, Health Program, Markle Foundation

Introduction

What might be achieved if clinical data could be positioned as a public good? How would such a system work, and what are the technical and policy issues to engage in fostering its evolution? This paper examines key definitions, assumptions, and approaches currently driving health data and research approaches. It posits that we need a new 21st-century health information paradigm that serves the public good while creating and building trust. The paper hinges on several assumptions. First, we should be open to resetting some definitions and assumptions about research. Second, we should be ready to articulate new working principles based on new paradigms for how information is created, shared, and used. Third, we need an information policy framework that addresses public hopes and concerns.

Connecting for Health

Since 2002, the Markle Foundation has convened and operated Connecting for Health (CFH), a public–private collaborative that works to accelerate the development of a health information-sharing environment to improve the quality and cost effectiveness of health care. The initiative, supported and operated by Markle with additional support from The Robert Wood Johnson Foundation, brings together a diverse group of health, policy, and technology leaders, including consumer groups, clinicians, hospitals, government entities, privacy advocates, technologists, and businesses.

CFH works to create a networked environment where vital information is available when and where it is needed, in a private and secure manner, to improve healthcare quality and to reduce medical errors. CFH is founded on the principle that many participants in the healthcare system want and need access to information, including consumers, providers, and researchers who must help build the evidence base for the most effective high-quality approaches to health care.

Making critical data available to inform high-quality care depends on having a robust information policy framework in place that fosters public trust. An overarching framework for health information technology called the CFH Common Framework has guided this work (Figure 1-1).

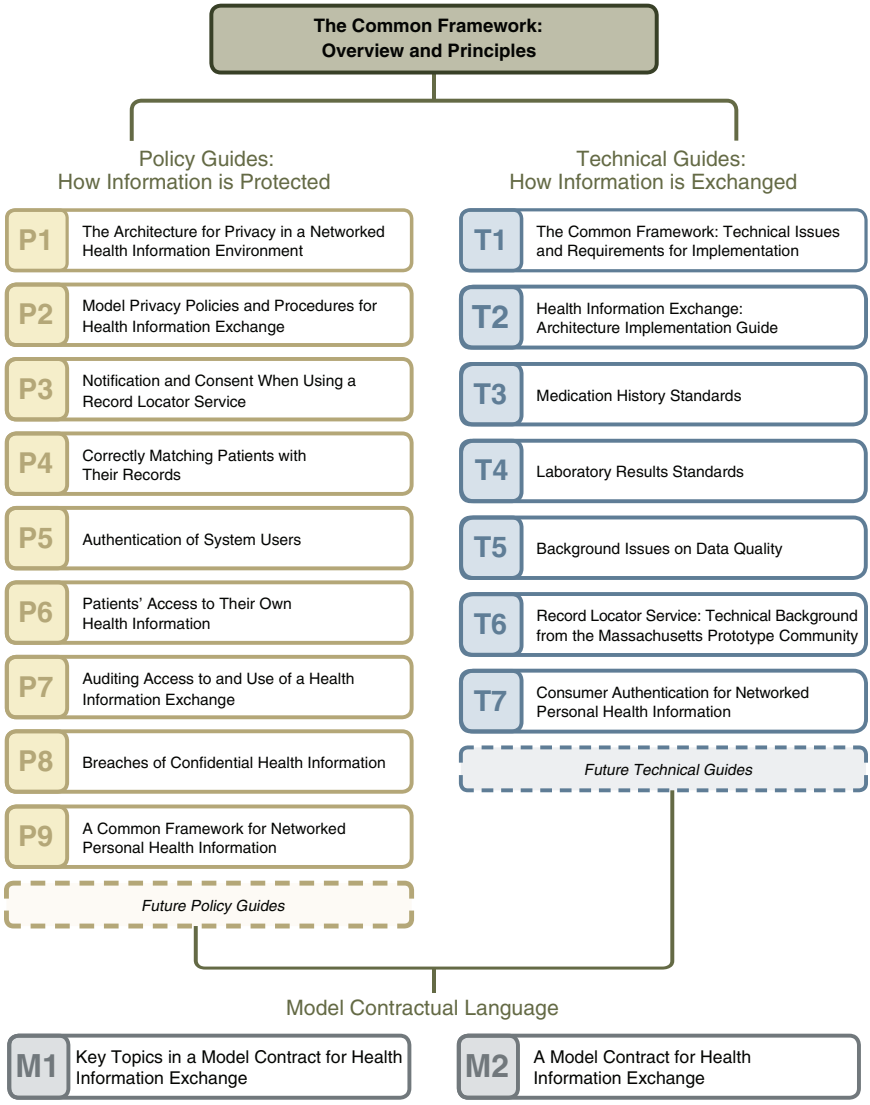


FIGURE 1-1 Connecting for Health: Common Framework overview and principles. SOURCE: Reprinted with permission from the Markle Foundation, 2009.

The CFH Common Framework outlines expectations for any health information technology effort in three areas:

1. Core privacy principles
2. Sound network design
3. Accountability and oversight mechanisms

These three key attributes of the framework are broadly applicable across HIT initiatives and business structures and can be used to help shape policies and technology choices for any HIT initiative, from regional health information exchange efforts to quality improvement. The framework is rooted in the assumption that key technical and policy decisions to create information sharing are inextricably linked, and must be jointly developed.

To apply and stress test the Common Framework, we worked with a diverse group of healthcare leaders and experts to develop detailed information policies and technical requirements that achieve the goals of the framework in three areas: (1) health information exchange (HIE) between individual healthcare providers or healthcare organizations (Markle Foundation, 2008a); (2) consumer access to their networked personal health information; and (3) public health and quality research using population-level information to support the nation's goals of improving clinical research, quality research, and public health and safety. This paper will focus on the last area, population health.

Population Health

CFH has defined improving population health as meeting three critical goals:

- Bolstering research capabilities and enabling clinical practice to fully participate in and use scientific evidence;
- Increasing the effectiveness of our public health system; and
- Empowering consumers and professionals with information about cost, quality, and outcomes.

The key objective is to improve how information is used to address research, public health, and quality measurement. Today, numerous and competing demands for data reporting are required of healthcare providers to satisfy the demands of researchers, those working for quality improvement efforts, and public health entities. However, the current healthcare environment is highly fragmented and poorly equipped to meet these often

redundant and idiosyncratic requests for data that occur daily. The result is that we often lack the robust information needed to measure and improve quality, conduct timely and effective research, and monitor threats to the public's health.

Although the demand for data is increasingly distributed and diverse, this demand has not been met with a network-based response. The difficulties in collecting, cleaning, and analyzing data harvested from multiple systems have remained consistently challenging. The result is long lag times in using the data once collected, inability to consistently yield valuable information, significant gaps in knowledge, and a chasm between applying the knowledge gained and consistently achieving better care.

To better understand this gap, leaders of the Connecting for Health Steering Group were interviewed on the current state of data aggregation and analysis for clinical research, public health, and quality measurement. The results indicate there is significant frustration with the current paradigm on the part of providers, as well as others responsible for population health. Although tremendous efforts have been devoted to amassing data, these expensive data collection efforts have not produced the anticipated and hoped-for benefits in terms of quality improvement or cost reduction. There is disappointment that over time, decision makers are still struggling with poor data to inform critical decision making.

However, common expectations for information *use* emerged as well. Experts believe a better model would be one where research becomes a normative part of health care, in which every intervention with a patient is a chance to learn something. The data must be inextricably linked with decision support and remeasurement, not merely serve as an episodic hiccup of a data dump. Simply put, information has to be fed back to somebody who can make a better decision based on that data.

Three Core Attributes and Population Health

Our work in elaborating on the Common Framework more fully as a complete set of policy and technical approaches is just beginning. To provide a robust approach, each element of the Common Framework must be considered.

The CFH Common Framework provides three central requirements that can guide the development of responsible information policies. First, population health approaches should meet "Core Privacy Principles" that are the foundation for creating the necessary information policies for a trusted information-sharing environment for research and public health: openness and transparency, purpose specification, collection and use limitation, individual participation and control, data integrity and quality, and security safeguards and controls. These seven principles draw extensively

on Fair Information Practices and Organisation for Economic Co-operation and Development principles that have been in use within the United States and internationally for decades.

It is important to consider that policies should be implemented before and with technology development. Post hoc policies are typically difficult to implement and often result in piecemeal fixes to policy problems—such as responding to a data breach with a laptop encryption fix—rather than proactively addressing the issue.

The second requirement of the CFH Common Framework is sound network design. Population health efforts should encourage information sharing or "interoperability" among decision makers, allow for flexibility across information systems or applications, and protect information through technology choices. Rather than working toward large, centralized networks, these efforts should take advantage of opportunities to decentralize information and architecture as described above.

The third component of the framework is accountability and oversight. Like the network itself, the accountability mechanisms for achieving this new paradigm for research will be distributed and shared among many groups. As a major funder of research and knowledge creation, government will have a clear leadership and accountability role in establishing specific requirements that achieve this objective and can serve as a catalyst in implementing a 21st-century approach. Researchers and research entities must also challenge themselves to develop, support, and innovate around new models that support the use of their findings by the people who can most benefit from them.

The Connecting for Health First Principles for Population Health

To enable rapid progress in achieving population health goals, there is a need to embed analysis, decision support, and feedback loops throughout the system. We cannot predict exactly who future information users will be or what questions they will bring. Because their needs will change over time, we have to start thinking more flexibly about the information and how to produce and use it. This is not a matter of returning to our old habits of creating centralized analytic functions. The challenge is to create alternative models that use modern information technology and take into account a wide variety of users, many and growing data sources, and a new approach to research and evidence creation.

CFH has developed nine "First Principles for Population Health" based on the Common Framework attributes of privacy protections, sound network design, and appropriate oversight and accountability.

1. **Designed for Decisions**
   A 21st-century health information environment will focus on improving the decision-making ability of the many actors in the health sector.
2. **Designed for Many**
   The 21st-century health information environment should empower a rich variety of users.
3. **Shaped by Public Policy Goals and Values**
   A 21st-century health information environment should achieve society's goals and values; examples include improving health, safety, and efficiency and reducing threats to public health.
4. **Boldly Led, Broadly Implemented**
   A 21st-century health information environment should be guided by bold leadership and strong user participation. The network's value expands dramatically with the number of needs it can meet and the number of participants it can satisfy.
5. **Possible, Responsive, and Effective**
   A 21st-century health information environment should grow through realistic steps.
6. **Distributed But Queriable**
   A 21st-century health information environment should be composed of a large network of distributed data sources.
7. **Trusted Through Safeguards and Transparency**
   A 21st-century health information environment should earn and keep the trust of the public through policies that provide safeguards and transparency.
8. **Layers of Protection**
   A 21st-century health information environment should protect patient confidentiality by emphasizing the easy movement of queries and responses, rather than of raw data.
9. **Accountability and Enforcement of Good Network Citizenship**
   A 21st-century health information environment should encourage and enforce good network citizenship by all participants.

As highlighted by these principles, a 21st-century approach needs to develop an information policy framework that broadly addresses public hopes and concerns. If we do not have an environment where people believe appropriate safeguards are in place to protect information, we will not realize our goals. Surveys indicate that consumers have serious concerns about the privacy of their health information (California HealthCare Foundation, 2005; FACCT Survey, 2003; Louis Harris & Associates, 1993). But we also know that if consumers believe safeguards are in place to protect their

information, they are willing to share personal information to help identify disease outbreaks or determine ways to improve the quality of health care (Markle Foundation, 2006).

A Vision for the 21st Century

A vision for 21st-century information sharing to improve population health will look at the problem from the perspective of the decision maker who needs to make better decisions. What would it look like if we achieved a future state where providers, consumers, payers, and policy makers all have access to information grounded in reliable evidence? In this regard, three CFH scenarios for the future were developed to illustrate the wide range of decisions that could be improved through better access to the right information at the right time (Markle Foundation, 2008b).

Scenario I: A Physician Practicing in a 21st-Century Health Information Environment

A physician in a small, four-doctor internal medicine practice in the suburbs is about to meet with a patient. The physician is trying to decide whether to put the patient on a new oral hypoglycemic. She runs a standardized network query to get information about whether this might be the right treatment for the patient. Later, she will benchmark herself against other physicians who might be caring for similar patients. She is also able to determine the most appropriate treatment for the patient's other presenting problem, a sputum infection. Although the literature indicates antibiotic A might be most appropriate, the latest information about a pneumonia outbreak in the local community suggests antibiotic B may lead to a better response. The scenario goes farther, imagining different financial models, new opportunities for collaboration, and a transformation of the basic care delivery model. This scenario is a way of imagining a future we want to achieve and is a starting place to outline the data "production function" that might get us there.

Scenario II: The Consumer Seeking Health Information in a 21st-Century Health Information Environment

From the consumer perspective, this scenario depicts the case of a mother who has questions regarding the care of her young son, who is asthmatic. She is able to use a readily available information network to examine data about physician quality, and can identify and select a doctor skilled and experienced in treating children like her son.

Scenario III: The Policy Maker Making Evidence-Based Decisions in a 21st-Century Health Information Environment

A third scenario is that of a policy maker who is faced with a decision about whether to reimburse a fictional new implantable renal device. This scenario addresses how a policy maker might approach such issues and how an information network could support decision making based on having access to evidence.

These scenarios provide an exciting glimpse of a future where evidence-based decision making is a matter of course. Yet the significant challenge that lies ahead is *how* to create the systems, analytic tools, and data sharing approaches that will support better decision making by consumers, providers, and policy makers. The IOM Roundtable on Value & Science-Driven Health Care has outlined a vision for a learning health system where clinical data are a staple resource. This is an important vision, but we may fail to achieve it if we are constrained by historical approaches for collecting and analyzing data.

It's Time for a New Paradigm

Nearly a decade later, the IOM's 2001 report *Crossing the Quality Chasm: A New Health System for the 21st Century* still provides an accurate description of the challenges at hand.

> Medical science and technology have advanced at an unprecedented rate during the past half-century. In tandem has come growing complexity of health care, which today is characterized by more to know, more to do, more to manage, more to watch, and more people involved than ever before. Faced with such rapid changes, the nation's healthcare delivery system has fallen far short in its ability to translate knowledge into practice and to apply new technology safely and appropriately. (IOM, 2001)

Progress is dependent on a bold new action agenda that is open to resetting some of our definitions and assumptions of health information and research approaches. All too often a great deal of time, money, and effort are spent collecting, cleaning, and analyzing data, only for them to be held in separate siloed repositories. This approach cannot efficiently meet the current needs of the many information sources and users. It is also a brittle approach in the sense that each new question or problem often requires another time-consuming round of data collection, cleaning, and formatting. Attempts to collect data for each population health initiative place a huge burden on data providers, who must field many requests for their data and report them repeatedly in many ways to different repositories. There is also the issue of privacy and security. As multiple or redundant large datasets

are created, privacy vulnerabilities can increase. Furthermore, this approach often lacks timely feedback loops and fails to inform better decisions at the point of care, which is the ultimate objective. Without timely feedback loops, the motivation to report or send data is low and can result in poor participation or compliance rates. Finally, the current approach does not contemplate a role or access to information by the consumer.

The way forward must start with an accepted set of working principles that are rooted in 21st-century paradigms. Businesses of earlier centuries thrived on command and control paradigms, but today's businesses depend on ideas and initiatives of the many. Other sectors such as banking or travel services or e-commerce are "networked." By tapping into information networks, consumers can pay bills, book flights, or pay a stranger on eBay. In this environment, success relies on distributing decision-making authority, incentives, and the rapid innovation of tools that create value to the participants. The U.S. healthcare system needs to be transformed in similar ways.

Today's environment is increasingly characterized by distributed needs for sharing and accessing actionable information for high-quality health care. The users and creators of clinical information—the "edges of the network"—are becoming increasingly sophisticated both in terms of having richer data and greater analytic capabilities. Because the information needed to conduct effective population health analyses is usually going to be distributed across many data sources in our highly fragmented healthcare system, leapfrogging the current paradigm will depend on finding ways to conduct these analyses effectively while allowing the data to remain distributed. In other words, rather than attempting to collect the data in centralized databases to address each research question, might it be more effective to push the question closer to the data, rather than always bringing the data to the question?

Several new models emerging within population health efforts take a distributed approach to how information is generated. One such example in public health that illustrates and provides important insights into the opportunities and challenges of this approach is the DiSTRIBuTE model developed and maintained by the International Society for Disease Surveillance (International Society for Disease Surveillance, 2008).

A longstanding goal of influenza surveillance has been to create a timely and accurate picture of flu-like illness trends regionally and nationally so that early detection and response to outbreaks can be managed. Traditionally, flu surveillance efforts have been based on a voluntary network of clinical providers who manually tally and report weekly counts of flu-like clinical visits during flu season. The considerable delay in reporting, high provider dropout, and lack of year-round data have been identified as major limitations with this system. More recently, national bioterrorism surveillance resources have been brought to bear on flu surveillance, and

have attempted to collect a broad range of raw patient data, from the clinical settings where data are generated, to derive whether or not somebody has the flu and thereby monitor flu trends. This recent approach, based on first collecting and then sending the required data fields to a centralized database, is cumbersome, and compliance with the reporting requirements on the part of clinical entities has been an ongoing challenge. In addition, the time for data collection, analysis, and communication is long making timely trend detection and response on the part of public health entities a difficult task.

Employing a different approach, the DiSTRIBuTE model considers those clinical delivery organizations that are already tracking flu-like rates by locally derived methods and asks whether meaningful information can be generated by electronically collecting only the summarized counts from each of these entities, regardless of how they were derived (Figure 1-2). This approach bypasses the need to collect full copies of detailed data fields at the individual patient level from each of these sources and limits the data request to the information that is truly the minimum required. The DiSTRIBuTE approach evaluates whether simply aggregating existing summary counts of locally determined flu cases can efficiently provide an accurate and timely trend analysis. Would it be possible to see and predict trends more quickly than manual reporting? More accurately? More comprehensively? So far the results are promising. Trend detection has been shown to be very effective as compared to other longstanding approaches, and timeliness of the information has improved. Also encouraging is that participation levels have increased over the life of the project. Within one year of its launch, the DiSTRIBuTE project has representative cities and states reporting from five of the nine national regions defined by the Centers for Disease Control and Prevention as well as Ontario, Canada. This provides data from more than 300,000 encounters per week—approximately equal to the entire national Sentinel Network output. Higher participation rates are likely due to two factors: the reduction of potential privacy or security concerns that local participants may have had about sharing individually identified data, and lowered barriers to entry for participation. The previously required effort to collect, assemble, and then report all the necessary data fields was replaced by simple requests of local participants to report weekly summary counts of flu cases they were already tracking.

Although this model is still being tested and requires further exploration regarding its potential applications to other research questions, it has demonstrated that a lot can be achieved quickly when new approaches are developed that focus on what information is really needed by a decision maker and how it will be used.

In a new paradigm, consumers should also be embraced as participants and producers of information. One example that demonstrates the chang-

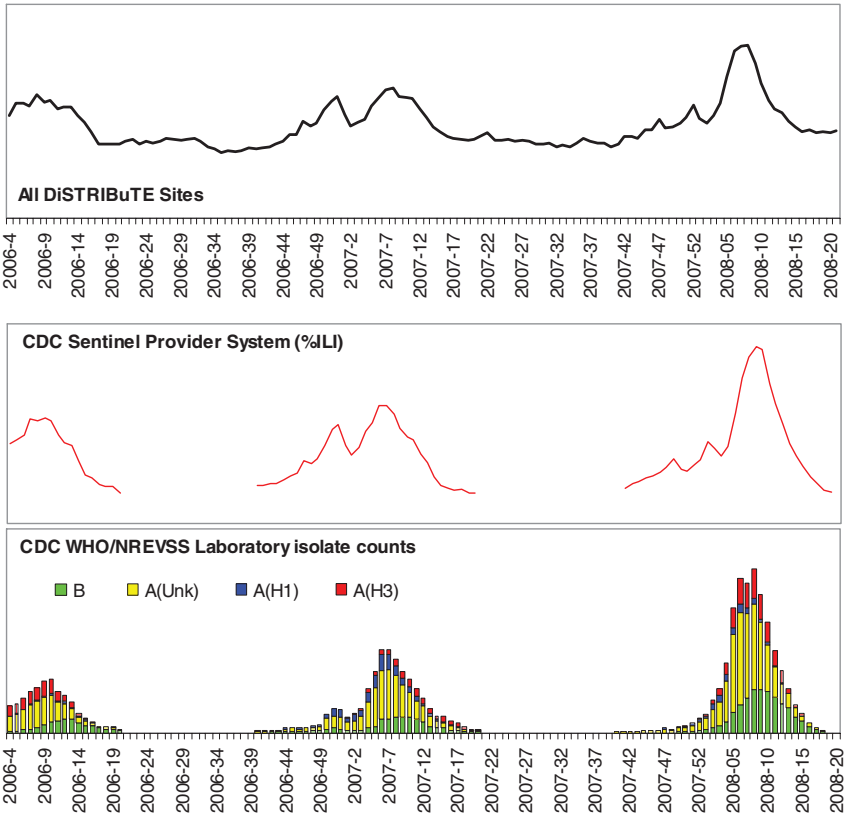


FIGURE 1-2 DiSTRIBuTE visualizations, week 2008–19 (ending Saturday, May 10, 2008). Time series depict respiratory, fever, and influenza-like syndrome emergency department visits by jurisdiction as percentage of total.
SOURCE: http://www.sydromic.org/projects/DiSTRIBuTE2008_02_09.doc (accessed August 31, 2010).

ing consumer role is a website called patientslikeme.com, an online community for patients with amyotrophic lateral sclerosis (ALS). On the site, patients share detailed information about themselves, their treatments, and their symptoms, building a warehouse of shared experiences and data. It is a highly sophisticated site built to accelerate the transfer of knowledge about what works and what does not from a patient perspective. It has been described as having information on the disease progression and history of more than 1,600 ALS patients—twice the number in the largest ALS trial in history. Remarkably, even before the trial results on lithium use for ALS were published, 50 patients worldwide had elected to start taking lithium

in collaboration with their doctors and were tracking their progression and blood levels on the site. This is more than twice the number of patients who were in the clinical trial itself. The site has data on historical forced vital capacity, the ALS Functional Rating Scale, and a standardized symptom battery. This example is compelling because it invites us to revisit our basic assumptions about the sources and uses of clinical data and about the nature and structure of the research process itself.

Conclusion

The future offers enormous possibilities. What if we create a climate of trust with a policy framework that truly enables information liquidity? What if we engage stakeholders in a constructive forward-looking process that prioritizes the creation of value for the participants? What if we embrace alternatives that involve and reward consumers for participating? What if we focus on the infrastructure requirements to push more questions to the data as opposed to trying to bring all the data to every question? What if we set our sights on a collective effort to address a small set of high-priority, public-good objectives using this new approach and enjoy some rapid learning?

Our "what if's" present many challenges—but those challenges exist now with our traditional approaches and are unlikely to go away. The goal of getting actionable data as quickly as possible to the people who need to make decisions every single day should drive the solutions. Improving health and health care depends on it.

REFERENCES

Arrow, K., J. Bertko, S. Brownlee, L. P. Casalino, J. Cooper, F. J. Crosson, A. Enthoven, E. Falcone, R. C. Feldman, V. R. Fuchs, A. M. Garber, M. R. Gold, D. Goldman, G. K. Hadfield, M. A. Hall, R. I. Horwitz, M. Hooven, P. D. Jacobson, T. S. Jost, L. J. Kotlikoff, J. Levin, S. Levine, R. Levy, K. Linscott, H. S. Luft, R. Mashal, D. McFadden, D. Mechanic, D. Meltzer, J. P. Newhouse, R. G. Noll, J. B. Pietzsch, P. Pizzo, R. D. Reischauer, S. Rosenbaum, W. Sage, L. D. Schaeffer, E. Sheen, B. M. Silber, J. Skinner, St. M. Shortell, S. O. Thier, S. Tunis, L. Wulsin, Jr., P. Yock, G. Bin Nun, S. Bryan, O. Luxenburg, and W. P. M. M. van de Ven. 2009. Toward a 21st-century health care system: Recommendations for health care reform. *Annals of Internal Medicine* 150(7):493–495.

California HealthCare Foundation. 2005. *National consumer health privacy survey*. http://www.chcf.org/topics/view.cfm?itemID=115694 (accessed August 18, 2008).

Blumenthal, D., E. G. Campbell, M. Gokhale, P. R. Yucel, B. Clarridge, S. Hilgartner, and M. N. A. Holtzman. 2006. Data withholding in genetics and the other life sciences: Prevalences and predictors. *Academic Medicine* 81(2):137–145.

Detmer, D. E. 2003. Building the national health information infrastructure for personal health, health care services, public health, and research. *BMC Medical Informatics and Decision Making* 3(1). http://www.biomedcentral.com/1472-6947/3/1 (accessed August 31, 2010).

FACCT (Foundation for Accountability) Survey. 2003. *Connecting for health: Consumer attitudes towards a personal health record*. http://www.connectingforhealth.org/resources/phwg_survey_6.5.03.pdf (accessed August 18, 2008).

Hrynaszkiewicz, I., and D. Altman. 2009. Towards agreement on best practice for publishing raw clinical trial data. *Trials* 10(17).

International Society for Disease Surveillance. 2008. *Home page*. http://syndromic.org (accessed August 18, 2008).

IOM (Institute of Medicine). 2001. *Crossing the quality chasm*. Washington, DC: National Academy Press.

———. 2009. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. Washington, DC: The National Academies Press.

Kawamoto, K., H. F. Willard, and G. S. Ginsburg. 2009. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Medical Informatics and Decision Making* 3(1).

Let data speak to data. 2005. *Nature* 438(7068).

Louis Harris & Associates. 1993. *Health information privacy survey, 1993*. New York: Louis Harris & Associates.

Markle Foundation. 2006. *National survey on electronic personal health records*. http://www.markle.org/downloadable_assets/research_doc_120706.pdf (accessed August 18, 2008).

———. 2008a. *Connecting for health: The common framework*. http://www.connectingforhealth.org/commonframework/ (accessed August 19, 2008).

———. 2008b. *Connectivity in the 21st century*. http://www.connectingforhealth.com/connectivity/ (accessed August 18, 2008).

Ness, R. B. 2007. Influence of the hipaa privacy rule on health research. *Journal of the American Medical Association* 298(18):2164–2170.

NRC (National Research Council). 2009. *Computational technology for effective health care: Immediate steps and strategic directions*. Washington, DC: The National Academies Press.

Piwowar, H. A., M. J. Becich, H. Bilofsky, and R. S. Crowley. 2008. Towards a Data Sharing Culture: Recommendations for Leadership from Academic Health Centers. *PLoS Medicine* 5(9):e183.

Safran, C., M. Bloomrosen, W. E. Hammond, S. Labkoff, S. Markel-Fox, P. C. Tang, D. E. Detmer, with input from the expert panel. 2007. Toward a national framework for the secondary use of health data: An american medical informatics association white paper. *Journal of the American Medical Informatics Association* 14(1):1–9. http://www.healthlawyers.org/Members/PracticeGroups/HIT/Toolkits/Documents/5_Health_Data_AMIA_Summary.pdf (accessed August 18, 2008).

2

U.S. Healthcare Data Today: Current State of Play

INTRODUCTION

Clinical data hold the potential to help transform the U.S. healthcare system. By providing greater insight to patients, providers, and policy makers into the appropriate application of interventions, and quality and costs of care, these data offer the opportunity to accelerate progress on the six dimensions of quality care—safe, effective, patient centered, timely, efficient, and equitable (Chaudhry, 2006; IOM, 2001, 2009; Safran et al., 2007). Understanding the scale of this potential and of the missed opportunities to improve health and health care due to gaps in data collection or barriers to their use requires an overview of existing healthcare data—the sources, types, accessibility, and uses. Through examples of healthcare data used to manage and drive improvements in care and for healthcare marketing, this chapter explores important aspects of healthcare data in the United States—examines what drives the collection of these data and the accessibility of these data for new clinical insights; reflects on how well these data are used and key barriers to wider use; and focuses attention on how clinical data from all sources—both public and private—could be made more widely useful to monitor clinical effectiveness.

As reviewed in this chapter, data are collected on socioeconomic, environmental, biomedical, and genetic factors; individual health status and health behaviors; biomedical and genetic factors, as well as on resource use, outcomes, financing, and expenditures. These data are stored in a variety of electronic health records (EHRs), personal medical records, disease registries, and other databases. However, the distribution of clinical

data across the healthcare system is highly fragmented, presenting significant opportunity for those offering services that coordinate and aggregate data resources. To generate and organize data for evidence-based decision support, it will be important to explore technologies to enhance interoperability, data standardization, and compatibility for future data utilities. Leveraging access to both administrative and clinical data may require additional investments in developing linkages across the variety of healthcare data and data warehouses. Furthermore, emerging opportunities to deliver data at the point of care for healthcare decisions may enhance the public's involvement in data-mining, data-sharing, and data-generating initiatives. Given the broad range of data sources and possible applications, a national strategy is needed to develop the requisite infrastructure and fill existing gaps in data collection and use.

Speaking from his experience at Kaiser Permanente and in his role as chair of the National Committee on Vital and Health Statistics (NCVHS), Simon Cohn offers an overview of current major activities in healthcare data collection and database capacity development, including those related to administrative and claims data, quality indicators, health status and outcomes data, clinical research data, industry-sponsored pre- and postmarket studies, regulatory studies, registries, and emerging datasets. To help frame the discussion, Cohn presents a taxonomy for health data, then reflects on key issues and barriers to address as we move to a learning health system. Cohn highlights the NCVHS recommendations for enhancing protections for secondary uses of data collected electronically as particularly informative for advancing the clinical data agenda. In the area of enhanced health data stewardship, NCVHS recommends that covered entities be more specific about what data will be used, how, and by whom; that notices of privacy practices need to be more meaningful; and that data stewardship needs to extend to personal health data held by noncovered entities in personal health records and similar instruments.

Massachusetts Health Quality Partners (MHQP) aggregates healthcare data to measure and report on physician performance in a more meaningful and transparent way—creating reports on performance at the physician network, medical group, practice site, and individual physician level, for both doctors and consumers. MHQP Executive Director Barbra Rabson shares aspects of this model, including its success in influencing investments in information systems to support quality and incentives for individual physicians and the challenges of engaging consumers. Overall, Rabson suggests, the MHQP experience and similar models hold promise for a world in which EHRs would be more fully and effectively integrated into medical practice, and clinical, claims, and personal data would be more fully integrated for quality improvement initiatives.

For decades, researchers and clinicians have taken advantage of sources

of rich clinical and population-based data to generate new insights, stimulate major research programs, and develop robust clinical guidelines. Michael Lauer, director of the Division of Prevention and Population Sciences at the National Heart, Lung, and Blood Institute (NHLBI), asserts that to achieve the goal of the IOM Roundtable, clinical data ultimately will need to be integrated across the research and care delivery continuum and be made available to patients, clinicians, and researchers. Examples from abroad and within U.S. health systems, such as the Health Maintenance Organization Research Network (HMORN) and the Department of Veterans Affairs, demonstrate that rigorous and prospective data collection can be incorporated into routine clinical care. Still, most clinical data are not collected at the point of care, and most are organized in isolated silos that are difficult to access. As data are increasingly integrated within the care continuum, Lauer cautions against using inherently biased observational data in lieu of well-designed experimental data for synthesizing evidence-based policy recommendations. Although confounders in observational data can be statistically controlled to reduce biases somewhat, an ongoing national need remains for enhancing, networking, and analyzing existing data.

Three major types of data are used by public and private entities to market healthcare products and services: health survey data, information about general consumption patterns, and administrative data generated by the healthcare delivery system. William Marder, senior vice president of the research and pharmaceutical units of Thomson Healthcare, reports on the use of data assets by providers and pharmaceutical companies, describing business models for the collection and analysis of these data.

CURRENT HEALTHCARE DATA PROFILE

Simon P. Cohn, M.D., M.P.H.
Chair, National Committee on Vital and Health Statistics
Associate Executive Director, The Permanente Federation,
Kaiser Permanente

This section aims to provide a brief overview of major current activities in healthcare data development and collection—including administrative and claims data, quality indicators, health status and outcomes data, clinical research data, industry-sponsored pre- and postmarket studies, regulatory studies, registries, and emerging datasets. The goal is to lay the groundwork and provide a context for addressing a variety of salient issues surrounding these data sources. Included are general comments about U.S. healthcare data today, with a view toward the future and a framework and taxonomy for health data, followed by reflections on key issues and barriers

that must be addressed before successfully moving forward. The conclusion contains an overview of a recent report from the NCVHS (or "the Committee") that was requested by the Department of Health and Human Services (HHS) to further investigate and consider "secondary uses" of electronically collected and transmitted healthcare data as we move into the world of the Nationwide Health Information Network (NHIN).

Background on the NCVHS

The NCVHS is a statutory public advisory committee to the HHS. It has a 59-year history of advising on national health information policy, including health data, standards, statistics, privacy, and issues related to developing the National Health Information Infrastructure (NHII). It has 18 members—16 appointed by the HHS Secretary and 2 by Congress. Members are leaders and experts in their field (e.g., public health, healthcare informatics, data standards, population health, privacy, and confidentiality). The NCVHS has a well-deserved reputation for open collaborative processes and the ability to deliver timely and thoughtful recommendations. These attributes allow it to work closely and effectively with HHS organizational entities such as the Office of the National Coordinator (ONC), with a particular focus on challenging and difficult crosscutting issues.

The NCVHS has a congressionally mandated role in relation to the *Health Insurance Portability and Accountability Act* (HIPAA), advising the HHS Secretary on HIPAA regulations and standards related to healthcare data, privacy and security, administrative and financial transactions, and healthcare identifiers. HIPAA code sets, including International Classification of Diseases (ICD), Current Procedural Terminology (CPT), and HIPAA Identifiers (including the National Provider Identifier), are key parts of the data infrastructure, and the NCVHS advises about the need for changes to those standards. Finally, the Committee monitors HIPAA implementation and advises Congress and the HHS Secretary with yearly status reports.

In 2000, as part of its charge under HIPAA, the NCVHS set forth a strategy, a framework, and selection criteria for interoperable clinical data standards.[1] This work provided the foundation for the selection of clinical message format standards and clinical terminology standards in 2002 and 2003, which became the core of Consolidated Health Informatics standards. Many of the standards were accepted by then-HHS Secretary Thompson and subsequently became key inputs to the Healthcare Information Technology Standards Panel (HITSP) process. Also, as part of the *Medicare Modernization Act* (MMA), the NCVHS was asked to investigate and advise the Centers for Medicare & Medicaid Services (CMS) and HHS

[1] See http://www.ncvhs.hhs.gov/hipaa000706.pdf.

on standards for e-prescribing—standards that have been accepted as part of federal rule making.

In 2001, after several years of investigation and hearings, the NCVHS published a strategic vision and strategy for building the NHII. The heart of the vision for the NHII is sharing information and knowledge appropriately so it is available to people when they need it to make the best possible health decisions. The NHIN is only one part of the larger vision: the NHII includes not only technologies, but more importantly, values, practices, relationships, laws, standards, systems, and applications to support all facets of individual health, health care, and public health (NCVHS, 2001). One important part of this report was an early recognition of the importance of HHS leadership, and a call for an office within the HHS reporting to the HHS Secretary, to coordinate and move this effort forward. Subsequently, the Office of the National Coordinator was created within the HHS.

Since the development of that office, the NCVHS has been tasked with working with the HHS and the ONC to deal with the more challenging cross-cutting issues—such as privacy and the implications for the NHIN. While not answering all questions, because it is unclear how the NHIN will develop and evolve, the NCVHS is beginning to pose the important questions and to start public discussions. The NCVHS has also recommended initial functional requirements for the NHIN and recently produced a report on enhanced protections for uses of health data in the context of NHIN. The NCVHS also investigates and makes recommendations to the HHS Secretary on healthcare quality measurement and data and on population health issues in general. Much of the following is based on the groundbreaking work of the NCVHS.

Health and Healthcare Data: Framework and Taxonomy

When thinking about evidence-based medicine and about the data or taxonomies needed to support that work, it is important to take a broad view of all possible factors that impact (or are impacted by) health and health care. Figure 2-1 was developed by the Committee in conjunction with the National Center for Health Statistics and the HHS Data Council and published in 2002 (NCVHS, 2002). This graphic provides a reminder of the many influences on the health of the nation. In the context of this discussion of more traditional health and healthcare data, as well as of issues and barriers, it is important to recognize how much information we do not routinely collect, or if we do, we do not normally integrate it into our vision of health and health improvement.

This NCVHS work was an important input to subsequent efforts to develop simpler, more approachable health and healthcare conceptual frameworks internationally. Figure 2-2, for example, shows a conceptual frame-

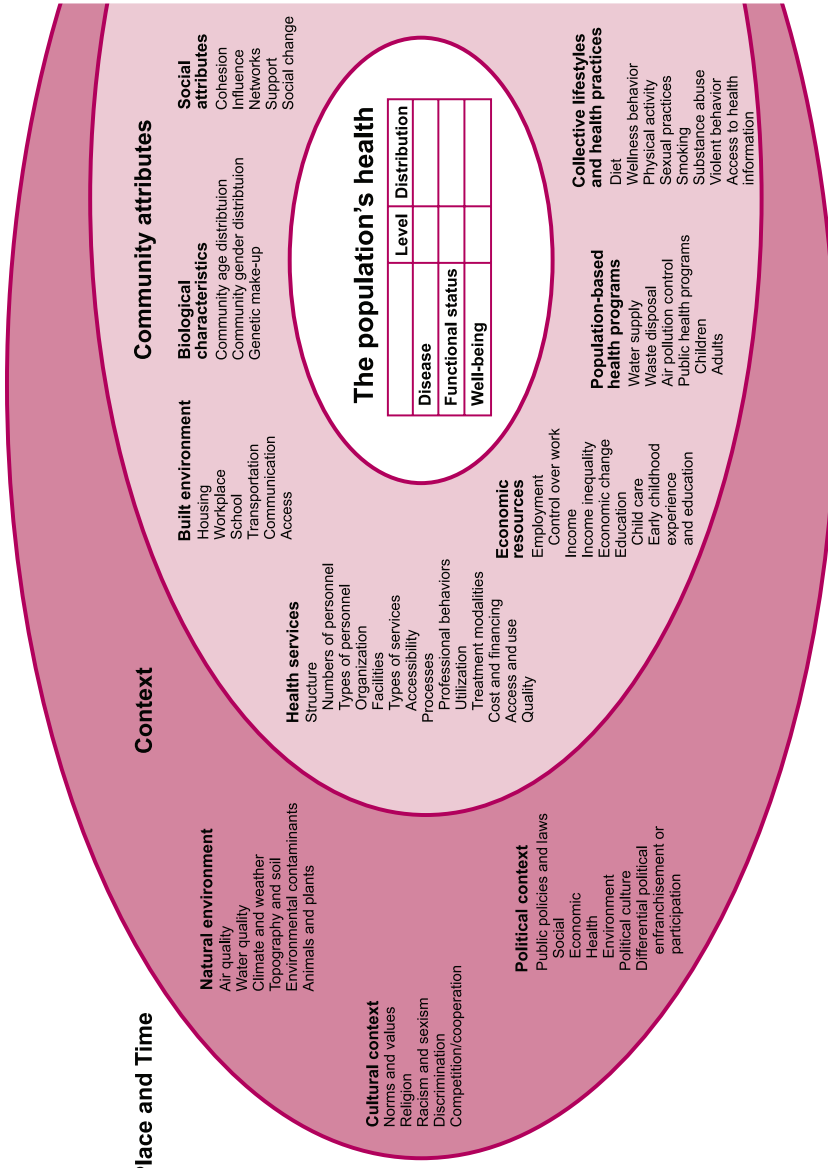


FIGURE 2-1 Influences on the population's health.

work, initially developed by the Australian Institute of Health and Welfare, for health system planning. It was subsequently published by the World Health Organization—which has used the diagram as a tool for healthcare terminology and classification planning (Madden et al., 2007).

This useful tool frames thinking about the data needed for a learning healthcare system as well as the development of sound health policy. In the center are the key concerns we need to monitor and focus on: health and well-being, including key aspects such as life expectancy, mortality, our own sense of well-being, state of functioning and disability, and, of course, illness, disease, and injury. Impacting these are health system interventions, including prevention and health promotion, and the major activity of the healthcare system—treatment, care, and rehabilitation. Determinants are important inputs into health and well-being such as biomedical and genetic factors, health behaviors, socioeconomic factors, and environmental factors. Impacting our ability to make interventions are resources and systems—human, economic, and others. This particular graphic begins to frame the discussion as we think about evidence-based medicine and data needs going forward.

The taxonomy represented in Box 2-1 provides more specifics. Used by

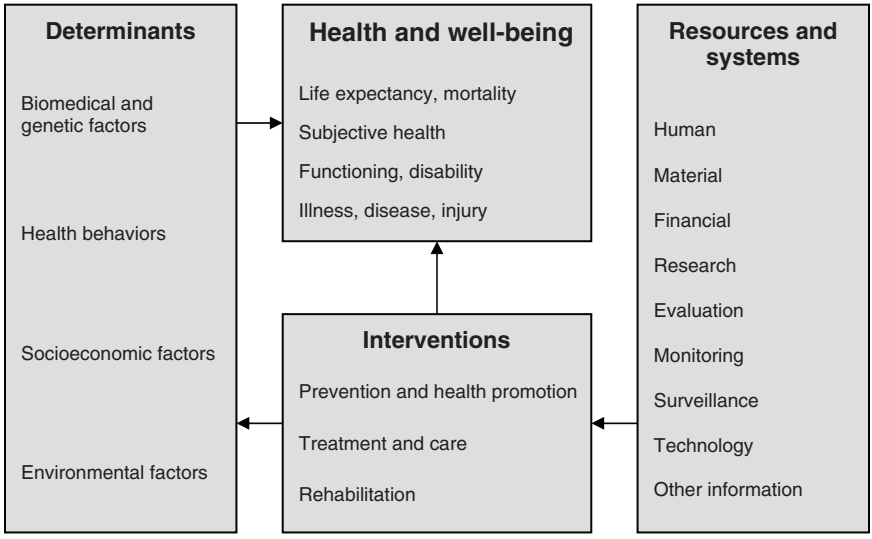


FIGURE 2-2 A conceptual framework for health.
SOURCE: Madden et al. (2007). Reprinted with permission from the World Health Organization © 2007.

the HHS Data Council for health data and health statistics planning, this taxonomy is focused on what we would traditionally describe as healthcare data and represents data that are central to a learning health system. One notable component of this taxonomy is its explicit recognition of the importance of longitudinal data. Unless we can understand the key factors that influence outcomes (including underlying health status and socioeconomic data) and connect them with the interventions and outcomes, it becomes difficult to have a learning health system.

BOX 2-1
Taxonomy Used by HHS Data Crucial for Health Data and Statistics Planning

- Demographics and Socioeconomic Data
  - — Age, sex, race, ethnicity, education, and related demographic/socioeconomic variables
- Health Status Data
  - — Individual health status, including morbidity, disability, diagnoses, problems, complaints, and signs and symptoms as well as behavioral and health risk factor data
- Health Resources Data
  - — Capacity and characteristics of the provider, plan, or health system
- Healthcare Utilization Data
  - — Nature and characteristics of the medical care visits, encounter, discharge, stay, or other use of healthcare services. Includes time, data, duration, tests, procedures, treatment, prescriptions, and other elements of the health encounter
- Healthcare Financing and Expenditure Data
  - — Costs, prices, charges, payments, insurance status, and source of payment
- Healthcare Outcomes
  - — Outcomes of prior or current prevention, treatment, counseling, or other interventions on future health status over time in a cyclical, longitudinal process
- Other Factors
  - — Genes and proteins, environmental exposures

SOURCE: Adapted from HHS Data Council (2007).

Key Infrastructure Barriers, Issues, and Suggested Next Steps

A number of infrastructure issues must be resolved to move to a true learning health system. The good news is that work is under way on some identified issues and others may be relatively inexpensive to resolve. Later chapters will address the political and competitive barriers and issues regarding a learning health system.

A barrier in the current healthcare data environment to implementation of the frameworks and taxonomy discussed is the wide distribution of data across the system and significant fragmentation of the data. Given the fact that the national healthcare enterprise consumes 16 percent of the gross national product and given the complexity of the human organism, it is not surprising that the system would be complex and the data systems complex. Currently, data are collected and held in many places—by the patient, provider(s), payers, and government repositories for public health and planning purposes, to name a few. Some of the data held are discrete and unique, and in other cases an extract or copy of data produced as a result of a healthcare interaction or event is stored. Few places, however, have comprehensive, longitudinal views about individuals. The inability to connect data that may include risk factors, medical history, and interventions in a comprehensive way is a fundamental flaw in moving forward. The hopeful news is that the vision of the NHIN is intended to help consolidate the data, but we are rife with fragmentation of health and healthcare data at this point.

In addition to the fragmentation of data, the data itself represented in the framework and taxonomy are heterogeneous. Some of the data—such as diagnosis (ICD), procedure (CPT), medication (National Drug Code or NDC), and other administrative data as required in HIPAA administrative and financial transactions—are usually of relatively high quality, coded, and computerized. Laboratory data are becoming increasingly standardized and codified; however, most other data are not available in a computerized form, or are generally in free text even if computerized. EHRs offer the opportunity for computerization and codification of additional key data elements; however, there is limited penetration of EHRs and thus "incomplete" computerization of data in health care.

Another issue of concern is variation in the timeliness of data. Timing ranges from clinical data (coded or not) being almost immediately available, at least for caregiving, to coded administrative data, which may take days or weeks to become available, to health statistics in government repositories used for planning purposes or research databases, which may lag by 1, 2, or more years.

Lest readers react in despair about the widely distributed nature of the data, uneven data quality, and time delays, previous testimony has

highlighted what we have learned from the current distributed environment. In particular, institutions such as Mayo, Kaiser Permanente (KP), the Veterans Administration (VA), and the Department of Defense (DoD) have longitudinal stores of relatively comprehensive, high-quality information on their patients. This is infrastructure that can be leveraged now to help identify evidence-based best practices. Certainly, the work of the ONC and HHS toward the vision and instantiation of the NHIN needs our support. Various initiatives that are also under way to help consolidate healthcare data for important purposes such as quality measurement deserve ongoing support and encouragement.

Considerations During the National Transition to EHR

To achieve the goal of having most decisions based on evidence as we move toward widespread EHR implementation, two focuses are needed. First, we need to be able to identify those evidence-based best practices, then we need ways to communicate those best practices to the *clinician* in a way that supports work overflow and high-quality clinical care.

The first focus, which is extensively discussed in this roundtable report, relies heavily on access to comparable and standardized data. Such data standardization and comparability, as we move towards fuller use of EHRs, requires uniform healthcare messaging standards (e.g., HL7 messages), an area that has received significant national attention, and robust healthcare terminologies and classifications, an equally important requirement that has, until recently, received much less attention As EHRs are being implemented, they are increasingly using clinical terminologies to codify their data, such as the Systemized Nomenclature of Medicine (SNOMED) and Logical Observation Identifiers Names and Codes (LOINC). Thus, for some time to come, we will need to consider strategies that can leverage claims, administrative health data, and the more specific, clinically rich information that is expected to come from EHRs.

In 2003–2004, the NCVHS looked at this transition issue and recommended a set of clinically rich terminologies to form a core for EHRs, calling for an aggressive mapping strategy between these and the HIPAA-mandated terminologies and classifications. The National Library of Medicine was asked to take the lead on this, but the mappings have been notoriously difficult (especially trying to map an archaic ICD classification to more modern clinical terminologies). Another problem is that both sides of the mapping have ongoing changes, so the mapping requires significant upkeep and runs the risk of being inaccurate.

We are encouraged by the current discussions about harmonization between SNOMED and ICD and plans to develop ICD-11—building off of an ICD-10 base, which includes plans to develop the clinical richness

of SNOMED with the classification discipline and international use of ICD. Linking administrative classifications and clinical terminologies could become an important tool and part of a transition strategy to help maximize the use of computerized data through both the transition to EHRs and the newer versions of ICD. Issues of concern remain, however. These include lack of adequate funding—the ICD-11 classification development work, for example, is currently funded mostly by the Japan Hospital Association. It is in our own national self-interest to get behind this as a way to ensure maintenance of the value of our data as we continue the transition to more current classifications and EHRs. A second issue is that U.S. representation needs to be further strengthened. We need to have a strong voice in how this goes forward because it will be an important piece of the infrastructure.

Other data terminology issues remain as we move forward with clinical interoperability and the implementation of standards and clinical terminologies to support MMA e-prescribing and the transition to EHRs and the NHIN. Clinically rich data, all standardized and interoperable, will provide a fertile environment for the learning health system, but many of these terminologies will be stretched to their limits. Unforeseen problems will need to be remedied. The bottom line is that federal terminology development and improvement initiatives are extremely underfunded. Furthermore, we will need adequate funding to fix problems and fill gaps as these standards and terminologies go into wider use.

Lack of quick action to fix problems will slow widespread adoption of EHRs and may undermine the NHIN. We are not talking about a huge amount of money: Funding in the range of $10 million per year may be sufficient to deal with both U.S. contributions and these real-world data issues.

The second critical issue is communicating evidence-based best practices to the clinicians in a way that supports workflow and high-quality clinician care—in other words, optimizing clinical decision support (CDS). Determination of best practices is critical, but the rate is limiting step may be getting that information to the busy care provider at the point of care in a way that is useful and actionable, and will impact decision making. These practices range from flu shot reminders to warnings about potential medication complications, and the number of evidence-based guidelines and recommendations continues to explode. There is no lack of evidence-based practices. (Most physicians have binders full of them written by their own organizations, by specialty societies, by accrediting organizations, by governmental organizations, etc.) For example, the Agency of Healthcare Research and Quality has 342 guidelines in its national clearinghouse on cardiovascular disease alone.

Unfortunately, although CDS exists in many healthcare organizations that have EHRs, it is generally proprietary and nonstandardized, and there

is no widespread agreement on how to share CDS information among organizations in an automated fashion. Furthermore, rules themselves are frequently not developed in a way that encourages computerization. Although CDS seems to work well with data entered within an EHR by an individual organization, the ability to merge and leverage data coming from elsewhere, especially administrative data, remains an issue (including trust issues). Work is being done in this arena, but significant efforts will be needed to address this important barrier to the vision of a learning healthcare system.

Enhanced Protections for Secondary Uses of Health Data

A transformation in health and health care is being enabled by health information technology (HIT): electronically available health data are no longer just claims data, but include more clinically rich data and can be linked more readily with other databases. This affords an opportunity to assess clinical outcomes over time, but also creates the risk of data being linked to databases that might jeopardize privacy, employment, or insurance eligibility. Sources and holders of electronic health information are expanding beyond HIPAA protections for personal health records (PHRs). Additionally, in areas such as personal EHRs, electronic solutions to protect and secure data continue to evolve, including the emergence of approaches to allow individual consent to follow data.

Against this backdrop, the NCVHS was asked last year by the ONC to look at issues and opportunities related to expanded uses of health data as we move from paper to electronic and from point-to-point data exchange to the vision of a nationwide health information network. The NCVHS was asked to develop an overall conceptual and policy framework to balance risk, benefits, obligations, and protections of various uses of health data. The Committee was also asked to develop recommendations for the HHS on possible next steps, including recommendations on data stewardship principles and approaches and other measures to enable optimal uses of health data while respecting individual privacy (NCVHS, 2007). The NCVHS was asked to pay particular attention to health data used for quality measurement, both for reporting and quality improvement—fundamental to a learning health system.

The committee held 8 days of hearings and heard from more than 60 people in person, with additional input received electronically. Two major themes emerged from the testimony. The first theme was the acknowledgment of the great benefit that can be achieved using electronic health data, including: improvements in care and care coordination; improved, more streamlined, and less burdensome quality measurement and reporting; automated monitoring for complications of drugs and devices; and

improved public health surveillance. Benefits from health data enabled by HIT/HIE (health information exchange) include timely access to information with relevant decision support, coordination of care across providers, automated and structured data collection for quality measurement and reporting, expedited accrual of cases for timely identification of complications from drugs and devices, and timely public health surveillance and responsiveness.

The second theme was a concern about the potential for harm, including the possible erosion of trust with potential compromise in health care when there is a divergence between expected and actual use of health data. Discrimination or confidentiality concerns may be amplified with increased ability to collect longitudinal data, coupled with sophisticated means to reidentify data.

To guide the development of recommendations and maintain consistency with other NCVHS work, the committee developed guiding principles for evaluating each recommendation. These principles include precepts that healthcare data protections should: maintain or strengthen an individual's health information privacy; enable improvements in the health of Americans and the healthcare delivery system of the nation; facilitate appropriate uses of electronic health information; increase the clarity and understanding of laws and regulations pertaining to privacy and the security of health information; and build on existing legislation and regulations whenever they are appropriate and do not result in undue administrative burden. The purpose was not to recreate HIPAA or create new regulatory or legislative burdens—in fact, many of the recommendations fall into best practices, guidance from the HHS, model forms and contracts, and similar approaches.

The recommendations fell into several categories. First, in the area of enhanced health data stewardship, it was recommended that covered entities strengthen the terms of their business associate contracts to be more specific about what data will be used, how, and by whom. Included is the recommendation that covered entities and their business associates confirm on a regular basis that practices are in compliance with the business associate contract. Second, recognizing the importance of guaranteeing transparency to the patient when data are applied for research, also a Roundtable goal, another recommendation was that the notice of privacy practices needs to be more meaningful, and individuals should be able to request and be supplied with additional information about what specific uses and users there are of their data—drawn from greater specificity in the business associate contracts. Third, data stewardship also needs to extend to personal health data held by noncovered entities in personal health records and similar instruments. The HHS and its offices have roles in this regard to monitor adherence to posted privacy policies, and this may be an area for legislative assistance (NCVHS, 2007).

The Committee paid special attention to uses of health data that are most immediately enhanced through HIT and HIE—quality measurement, reporting and improvement, and research—all fundamental to a learning health system. For example, the Committee reaffirmed that uses of health data for quality measurement, reporting, and improvement are within the scope of HIPAA Treatment, Payment, and Operations when conducted by covered entities. It was suggested that, as the industry begins the transition to HIE and an NHIN, there needs to be evaluation of new tools and technologies—which could include tools to help individuals manage their authorizations and new methods and techniques to deidentify health data. (This is an area for HHS and ONC leadership.) Finally, the NCVHS recognized that HIPAA has limits. Other protections beyond data stewardship may be needed, and certainly HIPAA protections only apply to covered entities. The NCVHS has long supported more inclusive federal privacy legislation to cover all organizations that have access to personal health data. At a minimum, expanding HIPAA coverage to new entities that are holding personal health information (PHI)—such as personal health record vendors, data banks, and similar entities—makes sense.

As next steps, the NCVHS plans to further investigate uses of deidentified data and how data stewardship might apply. The Committee is also monitoring work of both the Office for Health Research Protections and the Office of Civil Rights and may have further hearings related to the issues of overlap of quality and research.

In conclusion, the challenges are many. The good news is that significant initiatives are underway. However, many key issues identified continue to need national attention and focus to create a true learning health system.

DATA USED AS INDICATORS FOR ASSESSING, MANAGING, AND IMPROVING HEALTH CARE

Barbra G. Rabson, M.P.H.
Executive Director, Massachusetts Health Quality Partners

Background

Massachusetts Health Quality Partners is a multistakeholder coalition established in 1995 by a group of healthcare leaders who recognized the importance of having valid, comparable measures to drive quality improvement. The partnership includes physician, hospital, and health plan representatives, as well as representatives of government, consumer organizations, academic institutions, and employers. A guiding philosophy of

MHQP is that those being measured should be involved in the measurement process. Another is that through a collaborative process (e.g., aggregating data across health plans), we can improve care better together than any one plan or stakeholder group can do alone. This collaborative process philosophy falls in line with the Roundtable's theme of developing the business case for expanded data sharing in a distributed network.

MHQP reports trusted, reliable information to physicians to help them improve the quality of care they give their patients, and to consumers to help them take an active role in making informed decisions about their health care. This dual commitment to both physicians and consumers creates a healthy tension for MHQP. Ultimately, we believe consumers will have greater confidence in healthcare quality data if the data are trusted by their physicians.

This paper will discuss the work MHQP has done over the past several years to measure and report on physician performance, both privately and publicly, using health plan and Medicare claims data. It will focus on the benefits and challenges of using large aggregated databases for performance measurement, including MHQP's experience as one of six Better Quality Information (BQI) pilot sites involved in aggregating Medicare and commercial data. It will also reflect on recent efforts to capture electronic clinical data to measure and report on physician performance in partnership with the Massachusetts eHealth Collaborative. Finally, it will discuss the impact of MHQP's performance reporting to date and identify opportunities to create more meaningful quality measures from existing and future data sources.

MHQP's clinical measurement reporting has evolved since the organization began to report on the performance of Massachusetts physicians in 2004. We currently engage in four clinical reporting initiatives:

- Aggregation of Healthcare Effectiveness Data and Information Set (HEDIS) data across health plans;
- Aggregation of commercial health plan claims data and Medicare Fee-for-Service (FFS) claims data as a BQI pilot;
- Capture and aggregation of electronic clinical data in a quality data warehouse with the Massachusetts eHealth Collaborative; and
- Aggregation of health plan HEDIS data.

MHQP aggregates HEDIS data already calculated by health plans (numerators and denominators for individual physicians) across five of our member health plans: Blue Cross Blue Shield of Massachusetts, Fallon Community Health Plan, Harvard Pilgrim Health Care, Health New England, and Tufts Health Plan. Since 2004, MHQP has been issuing reports to primary care physicians in Massachusetts about how well

they perform on clinical HEDIS measures focused on the management of patients with chronic disease and the management of preventive care services. MHQP's *Statewide Comparative Clinical Quality Reports* address physicians' performance at multiple levels: individual physician, practice site, medical group, and network. For example, MHQP issues reports that compare (1) how the nine physician networks in Massachusetts perform compared to each other; (2) how the different medical groups within those networks perform; (3) how practice sites within the medical groups per-

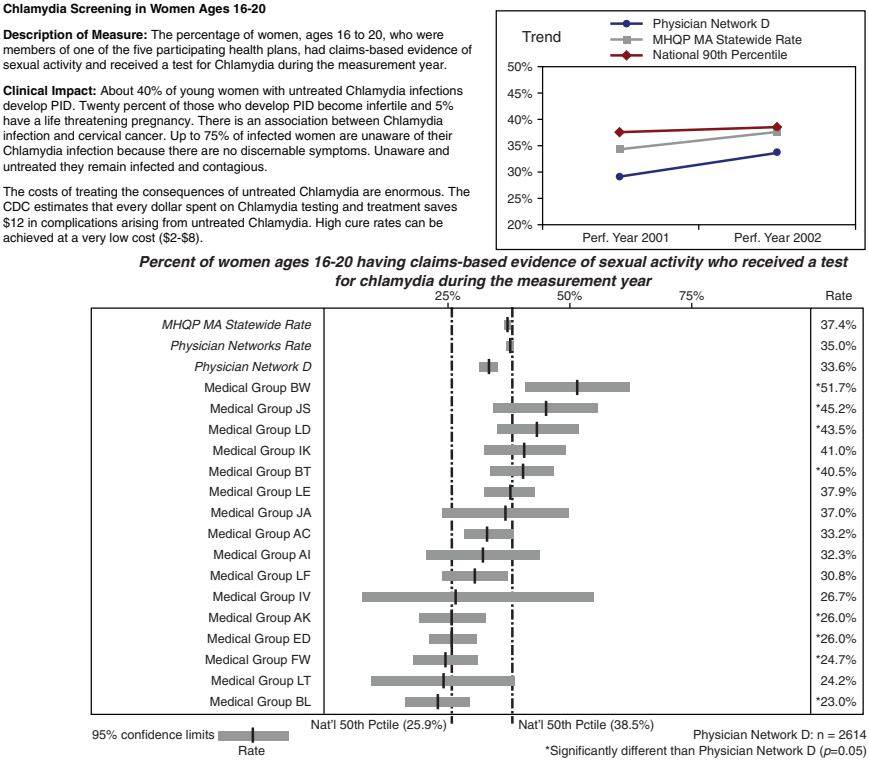


FIGURE 2-3 Snapshot of Massachusetts Health Quality Partners Statewide Comparative Clinical Quality Reports.
SOURCE: MHQP (2008). Reprinted with the permission of the Massachusetts Health Quality Partners.

form; and (4) how individual physicians within the practice site perform (Figure 2-3). This requires us to accurately map Massachusetts physicians to the sites and groups with whom they practice. To do this, MHQP has developed a process to gather and validate physician grouping information in collaboration with the physician offices.

MHQP first released physician quality data publicly on its website in 2005 and has publicly reported this information every year since, comparing the performance of 150 medical groups across MHQP. MHQP has a policy to issue private reports to physicians prior to the public release of the information.

On the MHQP website,[2] consumers can compare how well medical groups in Massachusetts provide preventive care services and manage chronic diseases in MHQP's *Quality Insights: Clinical Quality in Primary Care* report (Figure 2-4). The MHQP website also includes information about the conditions being measured, including information about what consumers can do to help manage their medical conditions and what they can expect their physicians to do to manage their care.

Aggregation of Commercial Health Plan Claims Data and Medicare FFS Claims

In 2006 MHQP was one of six organizations across the country to be designated as a Better Quality Information to Improve Care for Medicare Beneficiaries pilot. BQI is a CMS initiative to combine public and private information to measure and report on physician performance. One of the BQI project's major goals is to provide recommendations on the most effective methods to aggregate Medicare claims data with data from other payers in order to produce the most accurate, comprehensive measures of the quality of services being provided by physicians to Medicare beneficiaries. MHQP contracts with ViPS as a data aggregator to support MHQP's analysis of the claims data. The reports from MHQP's BQI project represent both primary care physicians and select specialists who participate in the Medicare FFS program and in four commercial health plan health maintenance organizations (HMOs) and preferred provider organizations (PPOs).

The key challenges in the BQI pilot have been as follows:

- Linking physician data across plans/payers: Requires the creation of a master physician directory because the National Provider Identifier is not broadly available.
- Attributing patient's care to appropriate physician(s) for non-managed–care patients: Requires the development of methodology

[2] See http://www.mhqp.org.


quality reports : clinical quality

QUALITY INSIGHTS: CLINICAL QUALITY IN PRIMARY CARE

Medical Groups Summary:
Diabetes Care For Adults

click on the measure name to learn more information about the measure

Medical Group	HbA1c Test	Cholesterol (LDL-C) Screening Test
Carney IPA *Go to Group's Website*	★★☆☆	★☆☆☆
Harvard Vanguard Medical Associates, Copley *Go to Group's Website*	★★★★	★★★★
Massachusetts General Hospital PHO, Partners Community HealthCare *Go to Group's Website*	★★★★	★★★☆

Click on a medical group to view results on all measures

FIGURE 2-4 Quality insights: Clinical quality in primary care report.
SOURCE: MHQP (2008). Reprinted with the permission of the Massachusetts Health Quality Partners.

to assign patient to physicians and to test whether these attribution methods reflect actual doctor/patient relationships.

- Validating data while maintaining privacy of PHI: Encrypted patient identifiers make it difficult to provide patient-specific feedback to physicians. This makes validation of the data very challenging.
- Reporting reliably at a level other than individual physician: Requires the mapping of physicians to the appropriate medical group. Tax ID numbers do not necessarily mirror practice affiliations.

Attributing Patient's Care to Appropriate Physician

Prior to working with FFS Medicare claims and commercial PPO claims, MHQP did not have to attribute patient care to a physician because MHQP used commercial managed-care data where patients are assigned to primary care physicians. For the BQI pilot, MHQP developed rules to attribute care of Medicare beneficiaries and the PPO population to a relevant physician based on the claims data. The BQI pilots developed and tested different attribution methodologies. There is an inherent trade-off between attributing as many patients as possible to physicians (e.g., there is a "one-touch rule" where a patient is attributed to all providers who have had any "Evaluation and Management" [E&M] visits with that patient) versus trying to assign care in a way that more accurately reflects actual accountabilities given the true relationships between clinicians and their patients. For example, is it reasonable to attribute a patient to a physician who saw a Medicare patient for a single visit and to hold the physician responsible for making sure that patient has had a mammogram, hemoglobin test, and so forth? It is important that we develop attribution methodology that allows for meaningful measurement of physicians.

The six BQI pilots are all in different markets and use different measurement models and data sources. Each BQI pilot health delivery market brings many unique characteristics. For example, the Indiana Health Information Exchange has access to a rich clinical data source because of the work the Regenstrief Institute has accomplished over the years, and the Wisconsin Collaborative for Health Care Quality does not use health plan claims data, but rather uses source data provided from the large physician groups. These models all provide a rich opportunity to learn a great deal about data aggregation.

Capture and Aggregation of Electronic Clinical Data

In 2006, MHQP was selected to work with the Massachusetts eHealth Collaborative (MAeHC) to build a Quality Data Warehouse (QDW), and to create quality metrics from electronic clinical data. The MAeHC is a multistakeholder organization funded by Blue Cross Blue Shield of Massachusetts with $50 million to encourage implementation of EHRs and HIE in Massachusetts. Three Massachusetts communities were selected by MAeHC to receive EHR systems in physician offices and a community-wide HIE. Working with MAeHC and our technology partner, Computer Sciences Corporation, MHQP is creating a QDW that holds data from the HIEs in the three pilot communities. The QDW is designed to collect clinical-quality data and report on quality measures for use by physicians, researchers, and others in the MAeHC communities. The QDW extracts

predefined, deidentified clinical data from the HIEs in the three pilot communities. MHQP is working to create quality metrics from these clinical data, then will provide performance feedback reports to the pilot communities at the physician, practice, and community levels.

Capturing clinical data and deriving quality measures from the EHRs and HIEs has been quite challenging. The technical specifications for measure creation have been based historically on data elements available in claims data, not electronic clinical data. Another challenge is that clinical information needs to be entered into the EHR in a standard format to easily capture the data for creating quality measures. There is an inherent trade-off between offering physicians flexibility in how they enter their information into the EHR (sometimes necessary to ease physicians into participation) versus the ability to capture useful information for quality metrics.

Lack of standards regarding data definitions and terms among EHR vendors, hospitals, labs, and radiology centers requires mapping data elements across sites, and can slow down the process of creating quality measures. A variety of codes are being used by different sites (e.g., NDC versus Multum codes), requiring time-intensive individualized crosswalks to be developed to bridge these coding systems. Finally, data from physician office EHRs only capture activity in the physician office. Many measures define their eligible population by an event that takes place in a hospital (e.g., a heart attack or heart surgery for patients with coronary artery disease [CAD]). To look at care across the continuum, a mechanism is needed to capture clinical information about patients from other locations of care, including hospitals, labs, and other entities.

Challenges with creating quality measures from electronic clinical information include:

- Difficult trade-off between offering physicians flexibility to enter data and standardization of data for easier data capture;
- Lack of standards in data definitions and terms;
- Lack of standardization across vendors;
- Measures not limited to physician office activity—for example, CAD measure requires hospital documentation (e.g., date) of an acute myocardial infarction (AMI); and
- Required elements (e.g., ICD-9, E&M, and NDC codes) to establish measure numerators and denominators are not always available in EHR data.

Impact of MHQP's Public Reporting

To date, MHQP's public reporting has had a greater impact on physician behavior than on consumer behavior. Knowing that your physician organization is going to be listed on a public website or appear on the front page of the *Boston Globe* in comparison with your competitors is a strong motivator for improving performance. MHQP has been told by physician organizations that our public reports have influenced physician organization investments in infrastructure to support quality. They have influenced decisions to accelerate implementation of electronic health record systems, and decisions about how standardized individual EHR systems should be within a physician organization given budget considerations. Physician organizations are also using MHQP's private reports within their organizations to focus improvement efforts and to reward individual physician performance. This means MHQP's reports have already become integral to the operations in some physician offices; they have become a tool for improvement and their impact will continue to grow as more physician organizations discover their utility and value.

Of the measures where MHQP has reported comparative physician performance publicly and privately, primary care physicians in Massachusetts have improved on eight of nine measures over the past 4 years. Although MHQP cannot claim that its reporting is responsible for the improvement, we do know that MHQP's reporting provides the yardstick that allows the tracking of physician performance over time.

On the consumer side, we know consumers go to MHQP's website, especially to find a new doctor, but we do not have a good sense about the overall impact the website is having on consumer behavior. From focus groups with consumers, we know they highly value MHQP's information about patient experience because it gives them information that resonates with them. Consumers would prefer all information about physician performance at the individual physician level.

MHQP has anecdotal information that indicates consumers do not always value the clinical information available. For example, in a recent focus group with consumers, one woman's reaction to MHQP's report about how well physicians provided breast cancer screening to women was to ask *why* she should care about results for a medical group that showed 95 percentage of the women who should be screened for breast cancer receive a mammogram. She noted that she gets *her* mammogram and wasn't concerned if other women get theirs. This consumer wanted data that would tell her that a physician would be more likely to cure breast cancer in a patient-centered and respectful way. MHQP wants to engage consumers, but clearly new types of measures and data sources need to be developed

to provide more meaningful information for consumers. We also believe a "quality framework" should be developed so consumers can increase their understanding and evaluation of quality measurement data.

Opportunities to Create More Meaningful Quality Measures

The best way for MHQP to create more meaningful quality measures is to be able to capture clinical outcome data from EHRs and other electronic data sources. Experts in the HIT world are beginning to pay attention to the need to capture healthcare quality information for measurement purposes, and on the national level, the health information community and the quality community are beginning to work with a broad-based set of stakeholders to define how HIT can effectively support quality improvement. Some progress has been made on this front, but we need to continue to push it forward. From MHQP's perspective, "nirvana" would be the integration of clinical and claims data, and, ultimately, the incorporation of personal health information as well.

In Massachusetts it is exciting that MHQP, in partnership with the MAeHC, has been officially designated as the Massachusetts Chartered Value Exchange with a goal of integrating quality and HIT. This will allow better access to electronic data sources that will enable reporting of outcome measures to help everybody improve care.

DATA PRIMARILY COLLECTED FOR NEW INSIGHTS

Michael S. Lauer, M.D., F.A.C.C., F.A.H.A.
Director, Division of Prevention and Population Sciences,
National Heart, Lung, and Blood Institute

Well over 100 years ago, Lord Kelvin identified numerical data as the cornerstone of successful science when he stated, "If you can not measure it, you can not improve it. . . . When you measure it and express it in numbers, you know something about it" (Kelvin, 1883). Modern clinical researchers and epidemiologists owe much of their success to their ability to collect, sort, and analyze increasingly vast amounts of numerical data. The Institute of Medicine's goal of making evidence-based medicine the norm depends on the use of data as a staple for developing scientifically sound guidelines. In fact, one of the Roundtables themes is to ensure that publicly funded data are used for the public benefit.

TABLE 2-1 Examples of Available Clinical Data Used for Generating New Insights

Type of Data	Examples
Clinical	Single-site electronic medical records
	Cleveland Clinic Stress Laboratory (Frolkis et al., 2003)
	Administrative data
	Premier's Perspective (Lindenauer et al., 2005)
	Medicare Claims (Krumholz et al., 2006)
	Health systems clinical data
	Veterans Administration
	HMO networks (*National cardiovascular data, 2007*)
Registries	County birth and death statistics
	"Eight Americas" project (Murray et al., 2006)
	State-mandated quality registries
	New York state revascularization (Hannan et al., 2005, 2008)
	Industry-supported disease registries
	National Registry of Myocardial Infarction (Cannon et al., 2000)
	Government-supported cohorts and surveys
	NHANES (Gregg et al., 2007)
	Framingham Heart Study (D'Agostino et al., 2001)
	Atherosclerosis Risk in Communities (Diez Roux et al., 2001; McPherson et al., 2007)
Clinical trial datasets	NHLBI-funded studies available for public use

NOTE: HMO = health maintenance organization, NHANES = National Health and Nutritional Examination Survey, NHLBI = National Heart, Lung, and Blood Institute.

Sources of Data

At least three major types of data are available to clinical and public health scientists (Table 2-1). Data based on clinical care come from electronic health records, clinic-based administrative datasets, and government payer datasets. Large-scale registries are generated and maintained by counties, state health authorities, professional societies, pharmaceutical and device companies, and the federal government. Clinical trials, whether publicly or privately funded, can function as rich sources of observational data, useful for exploring questions that go beyond their original hypotheses. Common features of all these types of data include an electronic format, predefined fields, and for most, large numbers that enable robust analyses.

Site-Based Electronic Health Records

The purest type of electronic clinical data is that which is obtained prospectively at the point of care and which is based on clearly defined objective quantitative variables. For example, in the early 1990s, physicians and exercise physiologists at the Cleveland Clinic assembled a computerized database within the exercise stress laboratory (Cole et al., 1999). For

all patients referred to the laboratory, providers directly entered into a computer server data on demographics, test indications, medical history, standard cardiovascular risk factors, medications, resting electrocardiogram findings, and exercise-test findings. This database was used initially to generate rapid, legible, and easily retrievable clinical reports. Clinical researchers later realized that the database could be combined with other sources of data, such as death registries or databases of other commonly obtained diagnostic tests, to study a variety of hypotheses. Published reports from these data demonstrated the prognostic value of simple measures such as functional capacity (Snader et al., 1997), chronotropic response (Lauer et al., 1999), heart rate recovery (Cole et al., 1999), and exercise-related ventricular ectopy (Frolkis et al., 2003). When researchers from the Cleveland Clinic collaborated with researchers at Kaiser Colorado, they were able to use their database to develop and validate a prognostic model for patients with suspected coronary disease and a normal resting electrocardiogram (Lauer et al., 2007).

Hospitals and insurance systems maintain administrative data for billing and quality monitoring purposes. Some investigators have employed administrative data to generate clinical insights, such as the potential value of beta-blockers for preventing perioperative deaths in high-risk patients (Lindenauer et al., 2005). Despite concern that administrative databases inherently yield biased estimates, some investigators have found that predictions based on administrative data closely approximate those based on rigorously obtained clinical data (Krumholz et al., 2006).

Registries

Large-scale registries are supported by counties, states, industry, professional societies, and the federal government. Counties have long maintained data on birth and death rates. Combining county and Census data, University of Washington researchers have shown marked inequalities of health according to demographics (Murray et al., 2006). A particularly alarming report focused on worsening life expectancy in some regions of the United States (Ezzati et al., 2008).

One of the best known state registries comes from New York, where data are routinely collected on all patients undergoing revascularization. These data have been used to produce "scorecards" specific to hospitals and providers (Topol and Califf, 1994). Patients or referring physicians can use these data to make better informed decisions. The data have also been used for observational comparative effectiveness studies of commonly available treatments. For example, Hannan and colleagues recently published analyses demonstrating probable superiority of coronary artery bypass grafting

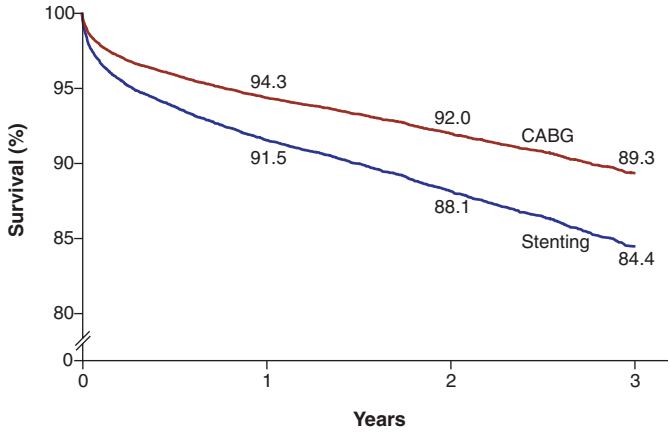
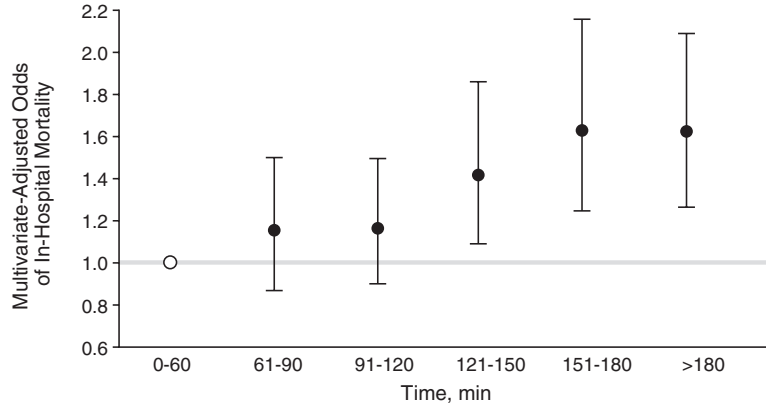


FIGURE 2-5 Three-vessel disease with disease of proximal LAD artery.
NOTE: Example of observations derived from the New York State revascularization registry. Patients who underwent coronary artery bypass grafting (CABG) had better outcomes than those who had stenting. Values are percentages at 1, 2, and 3 years; they were adjusted for the ejection fraction; the presence or absence of diabetes, congestive heart failure, chronic obstructive pulmonary disease, carotid-artery disease, aortoiliac disease, shock, renal failure, femoral or popliteal disease, and stroke; age; and sex.
SOURCE: Hannan et al. (2005).

over percutaneous coronary intervention among patients with severe multivessel coronary artery disease (Figure 2-5) (Hannan et al., 2005, 2008).

Pharmaceutical and device companies have funded multicenter registries that collate data on common clinical problems. For example, the National Registries of Myocardial Infarction (NRMI) have recorded baseline characteristics and short-term outcomes of literally millions of patients with acute coronary syndromes admitted to hospitals. One valuable report demonstrated the strong association between rapidity of percutaneous revascularization, commonly known as "door-to-balloon time," and mortality (Cannon et al., 2000) (Figure 2-6); these and similar observations were the basis for major national efforts to research and improve care (Nallamothu et al., 2007). In the past few years, the NRMI and similar registries have been taken over by a major specialty society, the American College of Cardiology.

The federal government has long supported population-based cohorts that were instrumental in discovering public health risks that are now common knowledge, including the dangers of smoking, diabetes, hypercholesterolemia, and hypertension (Executive summary, 2001). The best



No. of Patients	2230	5734	6616	4461	2627	5412
Multivariate OR	1.0	1.14	1.15	1.41	1.62	1.61
95% CI		0.87-1.48	0.89-1.49	1.08-1.84	1.23-2.14	1.25-2.08
P Value		0.35	0.29	0.01	<0.001	<0.001

FIGURE 2-6 Door-to-balloon time.
NOTE: Example of observations derived from an industry-supported myocardial infarction registry. Patients who had a longer door-to-balloon time had a higher risk-adjusted hospital mortality. The graph depicts multivariate-adjusted relationship between door-to-balloon time and mortality (χ^2 trend = 99.5; $P < 0.001$). Error bars indicate 95% confidence intervals (CIs); OR, odds ratio; open circle, the reference value. Door time refers to time of arrival at hospital and balloon time, time of first balloon inflation of the primary angioplasty procedure.
SOURCE: Adapted figure from Cannon et al. (2000).

known may be the Framingham Heart Study (D'Agostino et al., 2001), though a number of other cohort studies have yielded important findings on racial (Hozawa et al., 2007), socioeconomic (Diez Roux et al., 2001), genetic (McPherson et al., 2007), and subclinical (Detrano et al., 2008) aspects of common cardiovascular diseases. The Framingham Heart Study cohort is the basis for one of the most commonly accepted means of global risk assessment of patients at risk for coronary heart disease (Executive summary, 2001).

Other federal agencies support a number of registries and surveys that are commonly used by research epidemiologists. The Centers for Disease Control and Prevention (CDC), in conjunction with other federal agencies such as the National Institutes of Health (NIH), supports the National Health and Nutrition Examination Surveys (NHANES), which attempt to generate nationally representative estimates of risk and disease distributions. Combining NHANES data with federal national death registries, researchers have been able to show, for example, that women with diabetes

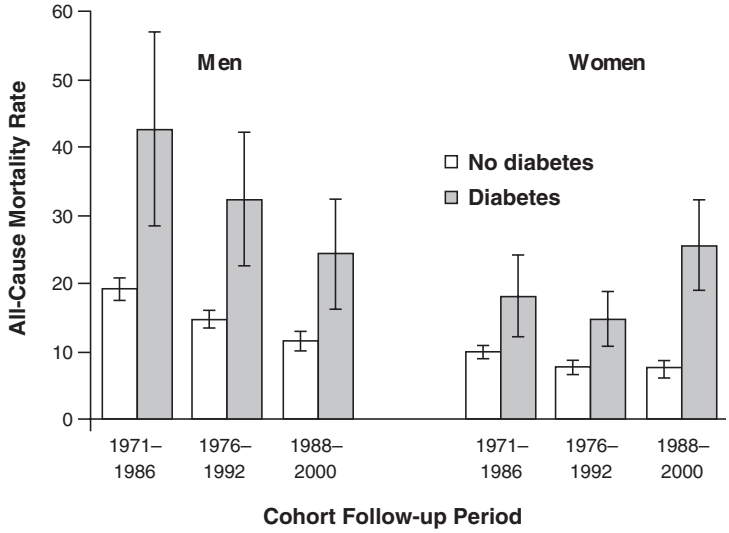


FIGURE 2-7 Mortality rate trends for men and women with diabetes.
NOTE: Example of observations derived from the CDC's NHANES surveys. Over a 30-year period, mortality rates decreased for men with diabetes, but actually increased among women. Age-adjusted all-cause mortality rates among the U.S. population aged 35 to 74 years with and without diabetes, by cohort and sex.
SOURCE: Gregg et al. (2007). Reprinted with permission of the *Annals of Internal Medicine*.

have seen worsening survival over the past 30 years, while men's outcomes have improved (Gregg et al., 2007) (Figure 2-7).

The CDC also supports the Behavioral Risk Factor Surveillance System (BRFSS), a large telephone survey that tracks health-related behaviors and self-reported risk factors. A recent report based on the BRFSS demonstrated marked geographical variability in blood pressure control and an association between this and variability in cardiovascular outcomes (Ezzati et al., 2008). Medicare supports the Medicare Provider Analysis and Review File (MEDPAR) that has been used, for example, to generate robust prediction models for outcomes of patients hospitalized for acute myocardial infarction or decompensated heart failure (Krumholz et al., 2006).

Limitations of Available Clinical and Population-Based Data

Despite the wealth of data available to researchers and policy makers, a number of major limitations must be realized. Relatively few American clinicians use computers to document care (Jha et al., 2006), and even

when they do, much of the imported data are free text that are inherently difficult to analyze. Most data are based on nonrandomized observations. With a few notable exceptions, most clinical data are not integrated across clinical practices and sites. Access to data varies, with some datasets widely available (like NHANES), whereas others are only available to personnel working at specific clinical sites or for specific sponsors.

Observational Data

Nearly all data derived from electronic health records and public or private registries are observational, that is, not based on randomized experiments. Although some have found that in general observational and randomized observations correlate well (Benson and Hartz, 2000; Concato et al., 2000), modern medical history is replete with examples of major discrepancies between observational findings and results of randomized trials (Pocock and Elbourne, 2000). Examples include hormone replacement therapy for prevention of chronic disease in postmenopausal women (Rossouw et al., 2002) and vitamin E for prevention of coronary disease events (Lee et al., 2005). A major problem with nearly all observational data is an inherent inability to correct for unmeasured confounders. Analysts have attempted to use modern statistical methods, such as propensity score or instrumental variable corrections (Stukel et al., 2007), but they have met variable levels of acceptance (D'Agostino and D'Agostino, 2007). Randomized trials have been criticized for being expensive and difficult to generalize, yet they remain the only method by which unmeasured sources of confounding and bias can be reliably considered.

Observational data still have value (Radford and Foody, 2001). In some cases, hypotheses are based on exposures that cannot be randomized based on natural or socioeconomic factors. Examples include biomarker levels, smoking, and small-particulate-matter air pollution (Miller et al., 2007). Observational analyses based on such exposures can be used to stimulate development of new treatments, but even so, randomized trials are eventually needed for evidence on which robust guidelines are based. For example, extensive epidemiological evidence has linked low-density lipoprotein (LDL) and high-density lipoprotein (HDL) levels to cardiovascular risk (Kannel, 1995). Some drugs that reduce LDL levels, such as statins, have been clearly shown to improve outcomes (Baigent et al., 2005), whereas others, such as torcetrapib (Barter et al., 2007), have not. Observational data are also useful for confirming results of randomized trials in groups of patients who were excluded from trials (Radford and Foody, 2001) and also for identifying rare safety signals (Graham et al., 2004) that even large trials are not powered to detect.

Data Integration

In the United States, most clinical data are not integrated across sites and practices. In contrast, other countries, such as the United Kingdom and Finland, have well-integrated databases that make it possible to follow patients easily regardless of where care is obtained. These integrated databases facilitated discoveries such as the high risks of angina in women (Hemingway et al., 2006) and the association of psoriasis with coronary disease events (Gelfand et al., 2006).

There are some notable American exceptions in which data have been successfully integrated. These include Medicare, the Department of Veterans Affairs, and HMO networks. Analyses of Medicare data have been used to define the potential benefits of aggressive management of patients with myocardial infarction (Stukel et al., 2007). Recently, integrated HMORNs have supported research programs focused on cancer and cardiovascular care (National cardiovascular data, 2008).

Data Access

In 1989, Claude L'Enfant, then NHLBI director, sent a memo to division directors, calling attention to a policy for widespread data release of Institute-supported, multicenter clinical trials and epidemiological studies (Figure 2-8). The Institute's policy has been to see data as a valuable resource paid for by taxpayers, and hence a resource that should be made available to the general scientific community, allowing for appropriate research subject protections. A number of researchers have successfully taken advantage of publicly available data, demonstrating, for example, the dangers of digoxin use in women with heart failure (Rathore et al., 2002), the epidemiology of valvular heart disease (Nkomo et al., 2006), and the public health threats posed by obesity in adults (Peeters et al., 2003).

The genomic revolution has led to a new level of data sharing, whereby highly detailed genotypic and phenotypic data are made available to qualified researchers. In October 2007, the NHLBI launched the Framingham SNP (single nucleotide polymorphism) Health Association Resource. Genotype data on more than 550,000 SNPs have been combined with data on hundreds of phenotypes among nearly 10,000 Framingham Heart Study subjects and stored in the NIH database on Genotype and Phenotype (dbGaP). Sharing genetic data on such a high level has been termed by some as a "bold experiment," in that it allows for a wealth of discovery, but it also raises questions about what levels of informed consent and privacy and confidentiality protection are appropriate (Caulfield et al., 2008; Psaty et al., 2007). For nongenetic data registries and cohort studies, standard models have been applied to require or waive written informed consent

December 8, 1989

NOTE TO DIVISION DIRECTORS:

Now that the Institute's Policy on Release of Data is established (copy distributed at the December 7 Executive Staff meeting), I would appreciate your sending it to your contractors (and even grantees for information) as appropriate. It would be unfortunate if down the line, someone would object, saying "But you never told me!"

Please let me know if you hear any comments.

Claude

Claude Lenfant, M.D.

cc: Dr. Frommer
Dr. Packard

I. General Policy and Applicability

It is the policy of the National Heart, Lung, and Blood Institute (NHLBI) to make available detailed data from collaborative clinical trials, epidemiological studies, and other large-scale studies conducted under contract, interagency agreement, or direct operations (but not under grant or cooperative agreement), with adequate protection of the confidentiality and privacy of research subjects.

FIGURE 2-8 Memorandum dated December 8, 1989, from Claude Lenfant, Director of the National Heart, Lung, and Blood Institute (NHLBI), to division directors regarding public release of data generated by large institute-supported studies.

requirements according to the Common Rule, as outlined in 45 C.F.R. 46 (NIH public access, 2008). Plans to share widely complex genetic data have raised new concerns about the level of consent needed, as reflected in the NIH's recently released Genome-Wide Association Studies policy (National Heart, 2007). In one program, the Personal Genome Project, an "open consent" model is being proposed by which adults volunteer to give DNA samples along with health information with the understanding that their data will be widely available and that there are no guarantees of anonymity, confidentiality, and privacy (Lunshof et al., 2008).

Many clinical data are produced as part of industry-supported clinical trials. These data are typically not made available to the public or even the general scientific community. Failure to share data exists on several levels.

Results of many trials are never published, leading to a biased impression about the efficacy or effectiveness of some treatments, such as antidepressants (Turner et al., 2008). In other cases, trial data that are not published can be obtained in incomplete format by researchers with affiliations with the Food and Drug Administration (FDA); these have been used, for example, to generate suspicions of the safety of commonly used drugs, such as rofecoxib (Vioxx) (Mukherjee et al., 2001) and rosiglatizone (Avandia) (Nissen and Wolski, 2007). On a more fundamental level, data may be published in aggregate form, yet access to raw data may be limited or delayed to academic researchers, as recently occurred in a multicenter trial of the cholesterol-lowering drug ezetimibe (Berenson, 2007).

Some sectors have taken steps to maximize access to clinical trial data, at least for academic researchers. A coalition of journal editors have required authors to attest to having access to all data (Davidoff et al., 2001), to having had clinical trials registered on a public forum (e.g., www.clinicaltrials.gov) (Laine et al., 2007), and, for some journals, to having obtained independent statistical analyses (DeAngelis and Fontanarosa, 2008). Recent federal legislation requires publicly funded research publications to be posted on a government website (National cardiovascular data, 2007) and requires results for many clinical trials, whether publicly funded or publicly noted, to be made publicly available.

Summary and Closing Thoughts

For many decades, researchers and clinicians have taken advantage of many sources of rich clinical and population-based data to generate new insights, stimulate major research programs, and develop robust clinical guidelines. The story of the cholesterol hypothesis is an excellent example of the power and limitations of clinical and population-based data. Epidemiological cohort studies established and described the strong link between blood cholesterol levels and cardiovascular risk (Kannel, 1995). These observational findings led to a reasonable, but unproven (Moore, 1989), hypothesis that lowering cholesterol could improve health. Drugs were developed that could reduce cholesterol levels, with some (Baigent et al., 2005), but not all (Barter et al., 2007), eventually shown in randomized trials to yield substantial improvements in patient outcomes. Postmarketing surveillance studies demonstrated the safety of statins; however, one exception, cerivastatin, was found to have an unacceptably high risk of a rare side effect, rhabdomyolysis, leading to withdrawal of that drug from the market (Graham et al., 2004). The cholesterol story illustrates the value of observational data for generating hypotheses and detecting safety signals, while also illustrating the critical role of randomized trials to generate robust evidence in support of specific therapies.

If the Institute of Medicine's evidence-based medicine goal is to be realized, clinical data must be recognized as a staple that should be widely available and integrated. Examples from abroad and from some U.S. health systems, such as HMORN and the VA, demonstrate that it is possible to incorporate rigorous and prospective data collection into routine clinical care. Still, most clinical data are not collected at the point of care in an easily retrievable manner and most are organized in isolated silos that are difficult for many analysts to access.

Even if a "data paradise" could be achieved with universally obtained and available clinical data, there is concern that policy leaders may place too much reliability on these largely observational datasets for generating evidence-based recommendations. Observational analyses of treatments must be recognized as inherently biased because of failure to take into account selection biases and unmeasured confounders. Modern statistical techniques and collection of more data elements may reduce these biases, but even with large numbers of observations, biases are still biases. I do accept the notion that a national priority for growing, sharing, and analyzing vast quantities of observational, clinical, and population data is an essential element toward reaching a vision of routinely practiced evidence-based medicine. This will only be true, though, if accompanied by a healthy dose of skepticism and recognition that, just as in Lord Kelvin's day, well-designed experiments are also critical for building a scientific evidence base.

HEALTH PRODUCT MARKETING DATA

William D. Marder, Ph.D.
Senior Vice President and General Manager, Thomson Healthcare

Three major types of data are used by public and private entities to market healthcare products and services: health survey data, information about general consumption patterns, and administrative data generated by the healthcare delivery system. Private health survey data are patterned after government-sponsored surveys such as the National Health Interview Survey (NHIS) or the Hospital Consumer Assessment of Healthcare Providers and Systems Survey (HCAHPS). Analyses drawn from general consumption patterns and market segmentation data keyed to census tract data can guide modeling behavior and messaging strategies. Much of the information about patient/consumer attitudes comes from this source. The administrative data include retail store sales data, patient eligibility and medical claims data, and a growing availability of short- and long-term disability claims data as well as health risk appraisal data. This paper describes

use of these data assets by providers and pharmaceutical companies and the business models that support collection of the data. The administrative data assets are often used in retrospective database studies to examine the cost effectiveness of interventions in the general population (outside the context of clinical trials, where both providers and patients are strongly encouraged to be on their best behavior). The interaction of private data assets and academic research will be discussed, including how access to data can be provided for replication of results.

The fundamentals of marketing are often described as a mix of "four Ps": Product, Price, Place/Positioning, and Promotion. In health care, many entities conduct marketing efforts that blend these factors to strategic advantage. Healthcare entities that engage in marketing include physicians, hospitals, pharmaceutical companies, device suppliers, and government agencies. The range of marketing activities in which these entities engage can be vast and varied. Examples might include planning for a new ambulatory surgery center, gaining acceptance for a new antidepressant, introducing a generic version of an established drug, raising mammography rates, or increasing enrollment in Medicaid or the State Children's Health Insurance Program.

Historically, health surveys have relied on government-collected data, long considered the reliable gold standard. Such data are not particularly helpful as marketing data, however, in that they tend to be fairly old and not easily linkable to general marketing tools. Those circumstances create an opportunity for the private sector to develop marketing data that are more current and linkable to marketing tools.

The marketing of health products is a thriving industry. Marketing to the public draws on lessons learned and information gained in the work of specific, targeted marketing such as the examples just cited, and also relies on information from additional sources, such as census data on population characteristics in small areas and customer buying habits. Increasingly, data compiled in support of the marketing of health products are being linked to health behaviors.

Within the private sector, many such marketing surveys exist. One example is the Thomson PULSE Survey, a questionnaire modeled on the NHIS. Based on a random telephone survey of 100,000 households per year, with replicates of 10,000 per month, 10 months per year, the survey offers results that are available 3 weeks after the close of each month in the field, identifiable by the census tract of the respondent. The survey is linkable to other census tract data, including socioeconomic characteristics of a particular census area and lifestyle modeling done by general marketing firms. In addition to this type of survey, there are provider-funded customer satisfaction surveys modeled on or incorporating HCAHPS. The Thomson PULSE Survey models healthcare use as a function of household and neigh-

borhood characteristics. Such models can then be integrated into software products that can help drive marketing and planning decisions of entities such as hospitals, government agencies, contract research organizations, and pharmaceutical companies.

As an example, assume that we want to find the best groups for clinical trial participation in "anytown," "anystate," ranked by clinical trial participation. The popular PRIZM (Potential Rating Index for ZIP Markets) system provides a standardized set of characteristics, known as clusters, for each U.S. ZIP Code. PRIZM is the nation's leading marketing segmentation system. (See www.claritas.com for more details, which are on a PRIZM poster available from Medstat.) Medstat licenses the system from Claritas and puts its unique health information into the system, especially disease prevalence information. Claritas assigns each block group to a PRIZM lifestyle segmentation cluster based on numerous demographic and socioeconomic variables, including age, income, population density, education, occupation, homeownership, and household composition. Media data come from Simmons Media Research Bureau, which conducted a separate survey of 50,000 households by PRIZM cluster. Answers to the PULSE Survey also help create the clusters. The objective of the clusters is to separate the population into groups that have strong differences in purchasing and health behaviors.

Using such an approach, we can, for example, pinpoint a target demographic group of blue-collar or farm couples, aged 35–54, who are high school graduates and owners of single-family dwelling units (our sample turns out to include a notable number of mobile homes). In terms of income, our group ranks at 45 out of 66 clusters. Mining the available data, we can determine that our group might be more likely than others to do crafts and needle work, go freshwater fishing, read *Flower & Garden* magazine, listen to country music, and own a Chevrolet Silverado. We can also differentiate that this given group has 7,069 patients who participated or seriously considered participating in a clinical trial, in contrast to a similar but slightly different group that has just 75 patients who were inclined to take part in a clinical trial.

Marketing data analyses draw on a rich abundance of administrative data that offer both advantages and shortfalls. Retail store sales data, for example, can include information on pharmaceutical use; available quickly, such data can sometimes identify the prescribing physician. There are billing service or product-switch data. Although these data are quickly accessible, they can sometimes be incomplete; such data can provide information on medical and pharmaceutical claims. There are health plan data, which have information on eligibility and claims for covered services, but may miss carved-out services. Finally, there are employer-based data, which can include eligibility and claims for covered services, sometimes include health

risk assessment data, and offer information on short- and long-term disability and worker's compensation claims.

Claims data offer significant marketing uses. Such data can be applied tactically—retail and product-switch data can be used, for example, to identify the effect of marketing campaigns and for measuring sales force effectiveness. At a perhaps more strategic level, claims data can offer insights for evaluating unmet medical needs, understanding the cost of acquiring a drug in a broader context, pricing new products, gaining favorable formulary position, and convincing prescribers about the value of a drug.

The development of healthcare marketing data is also informed by the FDA's encouragement of peer review. Strategic marketing goals can be accomplished by publishing material that meets peer-review standards. A substantial group of researchers address this need. The International Society for Pharmacoeconomics and Outcomes Research (www.ispor.org), for example, promotes the science of pharmacoeconomics (health economics) and outcomes research (the scientific discipline that evaluates the effect of healthcare interventions on patient well-being, including clinical outcomes, economic outcomes, and patient-reported outcomes) and facilitates the translation of this research into useful information for healthcare decision makers to ensure that society allocates scarce healthcare resources wisely, fairly, and efficiently. A combination of private/public, not-for-profit/for-profit entities contribute to this literature. Overall, the process means that for-profit entities that contribute data and research must develop strategies that are consistent with academic standards.

Given that the collection of such data can be expensive, there must be a revenue stream to offset data collection costs. In the case of the Thomson PULSE Survey, for example, the revenue stream comes from the use of the data in marketing and planning tools sold to providers and suppliers. Revenue also covers licensing of general marketing information. As for the funding of administrative data, the costs of retail and product-switch data are largely covered by pharma. Health plan and employer data are largely covered by the operations of payer organizations, with additional support from consultants serving many organizations, including pharma, government, benefits consultants, and reinsurance companies.

Licensing data in a for-profit setting has considerable benefits. Licensing helps customers achieve their goals by making data easy to use and to be sorted based on their interests in particular aspects of marketing's "four Ps." For license holders, the process of licensing offers the capability to market assets developed at considerable expense, and to recoup some of the costs of developing the data. This area raises considerations about how to best manage the intersection and interaction of private data assets and academic research. Although there is an inherent challenge in balancing the costs and benefits of making data available to students and researchers,

it is important that channels be maintained to make data available at no charge to academic audiences and to ensure access to data for replication of results.

In terms of applications of clinical data for marketing, not many such data are now available. The best of what are currently available are lab-result data linked to claims. Health plans are in the best position to acquire data from national labs. The comprehensiveness of such data can be checked relative to claims. At the same time, however, population-based comprehensive clinical data are not "right around the corner," as sometimes suggested, but are likely to be realized only at some point in the future. Registry data are not generally available for marketing purposes.

In sum, marketing data can be seen as a synergy of inputs and interests from a variety of entities. The public sector provides raw material and models of data collection, at minimal cost. The private sector builds databases with clear commercial value that fill needs suggested by, but not covered by, public sources. As electronic medical record systems become more common, one can envision a blend of databases that draws on both public and private data sources—the mix will depend on government willingness to fund aggregation.

REFERENCES

Baigent, C., A. Keech, P. M. Kearney, L. Blackwell, G. Buck, C. Pollicino, A. Kirby, T. Sourjina, R. Peto, R. Collins, and R. Simes. 2005. Efficacy and safety of cholesterol-lowering treatment: Prospective meta-analysis of data from 90,056 participants in 14 randomized trials of statins. *Lancet* 366(9493):1267–1278.

Barter, P. J., M. Caulfield, M. Eriksson, S. M. Grundy, J. J. Kastelein, M. Komajda, J. Lopez-Sendon, L. Mosca, J. C. Tardif, D. D. Waters, C. L. Shear, J. H. Revkin, K. A. Buhr, M. R. Fisher, A. R. Tall, and B. Brewer. 2007. Effects of torcetrapib in patients at high risk for coronary events. *New England Journal of Medicine* 357(21):2109–2122.

Benson, K., and A. J. Hartz. 2000. A comparison of observational studies and randomized, controlled trials. *New England Journal of Medicine* 342(25):1878–1886.

Berenson, A. 2007. Cardiologists question delay of data on two drugs. *New York Times*. 2007. http://www.nytimes.com/2007/11/21/business/21drug.html# (accessed August 10, 2008).

Cannon, C. P., C. M. Gibson, C. T. Lambrew, D. A. Shoultz, D. Levy, W. J. French, J. M. Gore, W. D. Weaver, W. J. Rogers, and A. J. Tiefenbrunn. 2000. Relationship of symptom-onset-to-balloon time and door-to-balloon time with mortality in patients undergoing angioplasty for acute myocardial infarction. *Journal of the American Medical Association* 283(22):2941–2947.

Caulfield, T., A. L. McGuire, M. Cho, J. A. Buchanan, M. M. Burgess, U. Danilczyk, C. M. Diaz, K. Fryer-Edwards, S. K. Green, M. A. Hodosh, E. T. Juengst, J. Kaye, L. Kedes, B. M. Knoppers, T. Lemmens, E. M. Meslin, J. Murphy, R. L. Nussbaum, M. Otlowski, D. Pullman, P. N. Ray, J. Sugarman, and M. Timmons. 2008. Research ethics recommendations for whole-genome research: Consensus statement. *PLoS Biology* 6(3):e73.

Chaudhry, B., J. Wang, M. Maglrare, W. Mojica, E. Roth, S. C. Marten, and P. G. Shekelle. 2006. Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care. *Annals of Internal Medicine* 144(10):742–572.

Cole, C. R., E. H. Blackstone, F. J. Pashkow, C. E. Snader, and M. S. Lauer. 1999. Heart-rate recovery immediately after exercise as a predictor of mortality. *New England Journal of Medicine* 341(18):1351–1357.

Concato, J., N. Shah, and R. I. Horwitz. 2000. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine* 342(25):1887–1892.

D'Agostino, R. B., Jr., and R. B. D'Agostino, Sr. 2007. Estimating treatment effects using observational data. *Journal of the American Medical Association* 297(3):314–316.

D'Agostino, R. B., Sr., S. Grundy, L. M. Sullivan, and P. Wilson. 2001. Validation of the Framingham coronary heart disease prediction scores: Results of a multiple ethnic groups investigation. *Journal of the American Medical Association* 286(2):180–187.

Davidoff, F., C. D. DeAngelis, J. M. Drazen, M. G. Nicholls, J. Hoey, L. Hojgaard, R. Horton, S. Kotzin, M. Nylenna, A. J. Overbeke, H. C. Sox, M. B. Van Der Weyden, and M. S. Wilkes. 2001. Sponsorship, authorship, and accountability. *New England Journal of Medicine* 345(11):825–826; discussion 826–827.

DeAngelis, C. D., and P. B. Fontanarosa. 2008. Impugning the integrity of medical science: The adverse effects of industry influence. *Journal of the American Medical Association* 299(15):1833–1835.

Detrano, R., A. D. Guerci, J. J. Carr, D. E. Bild, G. Burke, A. R. Folsom, K. Liu, S. Shea, M. Szklo, D. A. Bluemke, D. H. O'Leary, R. Tracy, K. Watson, N. D. Wong, and R. A. Kronmal. 2008. Coronary calcium as a predictor of coronary events in four racial or ethnic groups. *New England Journal of Medicine* 358(13):1336–1345.

Diez Roux, A. V., S. S. Merkin, D. Arnett, L. Chambless, M. Massing, F. J. Nieto, P. Sorlie, M. Szklo, H. A. Tyroler, and R. L. Watson. 2001. Neighborhood of residence and incidence of coronary heart disease. *New England Journal of Medicine* 345(2):99–106.

Executive summary of the third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (adult treatment panel III). 2001. *Journal of the American Medical Association* 285(19):2486–2497.

Ezzati, M., A. B. Friedman, S. C. Kulkarni, and C. J. Murray. 2008. The reversal of fortunes: Trends in county mortality and cross-county mortality disparities in the United States. *PLoS Medicine* 5(4):e66.

Frolkis, J. P., C. E. Pothier, E. H. Blackstone, and M. S. Lauer. 2003. Frequent ventricular ectopy after exercise as a predictor of death. *New England Journal of Medicine* 348(9):781–790.

Gelfand, J. M., A. L. Neimann, D. B. Shin, X. Wang, D. J. Margolis, and A. B. Troxel. 2006. Risk of myocardial infarction in patients with psoriasis. *Journal of the American Medical Association* 296(14):1735–1741.

Graham, D. J., J. A. Staffa, D. Shatin, S. E. Andrade, S. D. Schech, L. La Grenade, J. H. Gurwitz, K. A. Chan, M. J. Goodman, and R. Platt. 2004. Incidence of hospitalized rhabdomyolysis in patients treated with lipid-lowering drugs. *Journal of the American Medical Association* 292(21):2585–2590.

Gregg, E. W., Q. Gu, Y. J. Cheng, K. M. Narayan, and C. C. Cowie. 2007. Mortality trends in men and women with diabetes, 1971 to 2000. *Annals of Internal Medicine* 147(3):149–155.

Hannan, E. L., M. J. Racz, G. Walford, R. H. Jones, T. J. Ryan, E. Bennett, A. T. Culliford, O. W. Isom, J. P. Gold, and E. A. Rose. 2005. Long-term outcomes of coronary-artery bypass grafting versus stent implantation. *New England Journal of Medicine* 352(21):2174–2183.

Hannan, E. L., C. Wu, G. Walford, A. T. Culliford, J. P. Gold, C. R. Smith, R. S. Higgins, R. E. Carlson, and R. H. Jones. 2008. Drug-eluting stents vs. coronary-artery bypass grafting in multivessel coronary disease. *New England Journal of Medicine* 358(4):331–341.

Hemingway, H., A. McCallum, M. Shipley, K. Manderbacka, P. Martikainen, and I. Keskimaki. 2006. Incidence and prognostic implications of stable angina pectoris among women and men. *Journal of the American Medical Association* 295(12):1404–1411.

Hozawa, A., A. R. Folsom, A. R. Sharrett, and L. E. Chambless. 2007. Absolute and attributable risks of cardiovascular disease incidence in relation to optimal and borderline risk factors: Comparison of African American with white subjects—atherosclerosis risk in communities study. *Archives of Internal Medicine* 167(6):573–579.

IOM (Institute of Medicine). 2001. *Crossing the quality chasm*. Washington, DC: National Academy Press.

IOM. 2009. *Beyond the HIPAA privacy rule: Enhancing privacy, improving health through research*. Washington, DC: The National Academies Press.

Jha, A. K., T. G. Ferris, K. Donelan, C. DesRoches, A. Shields, S. Rosenbaum, and D. Blumenthal. 2006. How common are electronic health records in the United States? A summary of the evidence. *Health Affairs (Millwood)* 25(6):w496–w507.

Kannel, W. B. 1995. Clinical misconceptions dispelled by epidemiological research. *Circulation* 92(11):3350–3360.

Kelvin, W. T. 1883. Electrical units of measurement. *Physics Letters A* 1.

Krumholz, H. M., Y. Wang, J. A. Mattera, Y. Wang, L. F. Han, M. J. Ingber, S. Roman, and S. L. Normand. 2006. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation* 113(13):1683–1692.

Laine, C., R. Horton, C. D. DeAngelis, J. M. Drazen, F. A. Frizelle, F. Godlee, C. Haug, P. C. Hebert, S. Kotzin, A. Marusic, P. Sahni, T. V. Schroeder, H. C. Sox, M. B. Van der Weyden, and F. W. Verheugt. 2007. Clinical trial registration—looking back and moving ahead. *New England Journal of Medicine* 356(26):2734–2736.

Lauer, M. S., G. S. Francis, P. M. Okin, F. J. Pashkow, C. E. Snader, and T. H. Marwick. 1999. Impaired chronotropic response to exercise stress testing as a predictor of mortality. *Journal of the American Medical Association* 281(6):524–529.

Lauer, M. S., C. E. Pothier, D. J. Magid, S. S. Smith, and M. W. Kattan. 2007. An externally validated model for predicting long-term survival after exercise treadmill testing in patients with suspected coronary artery disease and a normal electrocardiogram. *Annals of Internal Medicine* 147(12):821–828.

Lee, I. M., N. R. Cook, J. M. Gaziano, D. Gordon, P. M. Ridker, J. E. Manson, C. H. Hennekens, and J. E. Buring. 2005. Vitamin E in the primary prevention of cardiovascular disease and cancer: The women's health study: A randomized controlled trial. *Journal of the American Medical Association* 294(1):56–65.

Lindenauer, P. K., P. Pekow, K. Wang, D. K. Mamidi, B. Gutierrez, and E. M. Benjamin. 2005. Perioperative beta-blocker therapy and mortality after major noncardiac surgery. *New England Journal of Medicine* 353(4):349–361.

Lunshof, J. E., R. Chadwick, D. B. Vorhaus, and G. M. Church. 2008. From genetic privacy to open consent. *Nature Reviews Genetics* 9(5):406–411.

Madden, R., C. Sykes, and T. Ustun. 2007. *World Health Organization family of international classifications: Definition, scope and purpose*. http://www.who.int/classifications/en/FamilyDocument2007.pdf (accessed September 23, 2008).

McPherson, R., A. Pertsemlidis, N. Kavaslar, A. Stewart, R. Roberts, D. R. Cox, D. A. Hinds, L. A. Pennacchio, A. Tybjaerg-Hansen, A. R. Folsom, E. Boerwinkle, H. H. Hobbs, and J. C. Cohen. 2007. A common allele on chromosome 9 associated with coronary heart disease. *Science* 316(5830):1488–1491.

Miller, K. A., D. S. Siscovick, L. Sheppard, K. Shepherd, J. H. Sullivan, G. L. Anderson, and J. D. Kaufman. 2007. Long-term exposure to air pollution and incidence of cardiovascular events in women. *New England Journal of Medicine* 356(5):447–458.

Moore, T. J. 1989. The cholesterol myth. *Atlantic Monthly* 264:37.

Mukherjee, D., S. E. Nissen, and E. J. Topol. 2001. Risk of cardiovascular events associated with selective COX-2 inhibitors. *Journal of the American Medical Association* 286(8):954–959.

Murray, C. J., S. C. Kulkarni, C. Michaud, N. Tomijima, M. T. Bulzacchelli, T. J. Iandiorio, and M. Ezzati. 2006. Eight Americas: Investigating mortality disparities across races, counties, and race–counties in the United States. *PLoS Medicine* 3(9):e260.

Nallamothu, B. K., E. H. Bradley, and H. M. Krumholz. 2007. Time to treatment in primary percutaneous coronary intervention. *New England Journal of Medicine* 357(16):1631–1638.

National cardiovascular data registry home page. (2008). http://www.accncdr.com/WebNCDR/Common/ (accessed April 23, 2008).

National Heart, Lung, and Blood Institute creates national network to study cardiovascular disease. 2007. http://www.eurekalert.org/pub_releases/2007-11/kpdo-nhl112807.php (accessed April 23, 2008).

NCVHS (National Committee on Vital and Health Statistics). 2001. *Information for health: A strategy for building the national health information infrastructure*. http://www.ncvhs.hhs.gov/nhiilayo.pdf (accessed August 20, 2008).

NCVHS. 2002. *Shaping a health statistics vision for the 21st century*. http://www.ncvhs.hhs.gov/21st%20final%20report.pdf (accessed August 20, 2008).

NCVHS. 2007. *Enhanced protections for uses of health data: A stewardship framework for "secondary uses" of electronically collected and transmitted health data*. http://www.ncvhs.hhs.gov/071221lt.pdf (accessed August 20, 2008).

NIH public access policy details. 2008. http://publicaccess.nih.gov/policy.htm (accessed April 23, 2008).

Nissen, S. E., and K. Wolski. 2007. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine* 356(24):2457–2471.

Nkomo, V. T., J. M. Gardin, T. N. Skelton, J. S. Gottdiener, C. G. Scott, and M. Enriquez-Sarano. 2006. Burden of valvular heart diseases: A population-based study. *Lancet* 368(9540):1005–1011.

Peeters, A., J. J. Barendregt, F. Willekens, J. P. Mackenbach, A. Al Mamun, and L. Bonneux. 2003. Obesity in adulthood and its consequences for life expectancy: A life-table analysis. *Annals of Internal Medicine* 138(1):24–32.

Pocock, S. J., and D. R. Elbourne. 2000. Randomized trials or observational tribulations? *New England Journal of Medicine* 342(25):1907–1909.

Psaty, B. M., D. Arnett, and G. Burke. 2007. A new era of cardiovascular disease epidemiology. *Journal of the American Medical Association* 298(17):2060–2062.

Radford, M. J., and J. M. Foody. 2001. How do observational studies expand the evidence base for therapy? *Journal of the American Medical Association* 286(10):1228–1230.

Rathore, S. S., Y. Wang, and H. M. Krumholz. 2002. Sex-based differences in the effect of digoxin for the treatment of heart failure. *New England Journal of Medicine* 347(18):1403–1411.

Rossouw, J. E., G. L. Anderson, R. L. Prentice, A. Z. LaCroix, C. Kooperberg, M. L. Stefanick, R. D. Jackson, S. A. Beresford, B. V. Howard, K. C. Johnson, J. M. Kotchen, and J. Ockene. 2002. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the women's health initiative randomized controlled trial. *Journal of the American Medical Association* 288(3):321–333.

Safran, C., M. Bloomrosen, W.E. Hammond, S. Labkoff, S. Markel-Fox, P. C. Tang, D. E. Detmer, with input from the expert panel. 2007. Toward a national framework for the secondary use of health data: An american medical informatics association white paper. *Journal of the American Medical Informatics Association* 14(1):1-9. http://www.healthlawyers.org/Members/PracticeGroups/HIT/Toolkits/Documents/5_Health_Data_AMIA_Summary.pdf (accessed August 18, 2008).

Snader, C. E., T. H. Marwick, F. J. Pashkow, S. A. Harvey, J. D. Thomas, and M. S. Lauer. 1997. Importance of estimated functional capacity as a predictor of all-cause mortality among patients referred for exercise thallium single-photon emission computed tomography: Report of 3,400 patients from a single center. *Journal of the American College of Cardiology* 30(3):641–648.

Stukel, T. A., E. S. Fisher, D. E. Wennberg, D. A. Alter, D. J. Gottlieb, and M. J. Vermeulen. 2007. Analysis of observational studies in the presence of treatment selection bias: Effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. *Journal of the American Medical Association* 297(3):278–285.

Topol, E. J., and R. M. Califf. 1994. Scorecard cardiovascular medicine. Its impact and future directions. *Annals of Internal Medicine* 120(1):65–70.

Turner, E. H., A. M. Matthews, E. Linardatos, R. A. Tell, and R. Rosenthal. 2008. Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine* 358(3):252–260.

3

Changing the Terms: Data System Transformation in Progress

INTRODUCTION

Compared to even just a few years ago, today's research questions are often dauntingly complex, a characteristic reflected in the data required for such research. Data are often needed from multiple data sources, including laboratory values, unstructured text records of clinical findings, cost and quality information, and genetic data. In addition, research on less common clinical conditions—those with low incidence or prevalence—inherently demand larger data sets with greater geographic and demographic diversity. Data from a single organization are generally insufficient for many research questions aimed at gaining the depth of understanding required to support evidence-based practices tailored to individuals. Thus there is a growing need for data sharing across research entities and collaborators. Approaches will need to be fast, inexpensive, sustainable, secure, and customized to meet the differing needs of both patients and researchers.

With significant volumes of clinical data housed in public and private repositories across the nation, pioneers are seeking opportunities to use these data to gain powerful insights by synthesizing elements of multiple data sources. Portions of the data stored by healthcare organizations are undergoing transformation—linking large datasets, aggregating health databases, networking for standardized reporting, and developing and interpreting registries. The growth in number and scope of large, linked datasets, aggregated data, and registries will likely benefit care delivery and research. However, different approaches to organizing and aggregating data

generate a unique set of limitations and challenges—all of which seem to be responsive to unique incentives and drivers.

This chapter highlights some notable existing and emerging efforts to coordinate clinical data into more readily available and usable resources; describes incentives for these activities; examines the shortfalls, limitations, and challenges related to various approaches to organizing and aggregating data; and looks at the dynamics pushing integration.

The National Cancer Institute (NCI), for example, has determined that the scale of its enterprise has reached a level that demands new, more highly coordinated approaches to informatics resource development and management. As discussed by Peter Covitz, chief operating officer of NCI, at NCI's Center for Bioinformatics, the Cancer Biomedical Informatics Grid (caBIG) program was launched to meet this challenge. The caBIG infrastructure is a voluntary network that facilitates data sharing and interpretation with aims to translate knowledge from the laboratory bench to patient bedside. As a tool designed to link resources within the cancer research community, caBIG would ultimately function as a template for sharing and communicating in a common language as well as a platform for building tools to collect and analyze information. The caBIG project is an essential resource to complement other cancer research projects. Moreover, Covitz suggests, caBIG might serve as a possible model for engaging the broader challenge of developing nationwide, interoperable health information networks.

Translational health research draws information from institutional entities as the primary source of analysis. Such an approach has historically enabled researchers to compare outcomes and differences in practice patterns within organizations. Pierre-André La Chance, chief information officer and research privacy officer at the Kaiser Permanente Center for Health Research, offers strategies on cross-institution data sharing through local, interoperable data warehouses and data networks. With an interconnected approach to data, researchers can access data resources more efficiently; the data have higher quality and reliability for generating analyses and decisions that affect both treatment and policy. At Kaiser Permanente, work is underway to develop sharable administrative, disease registry, and clinical data resources as well as a biolibrary to increase access to Kaiser Permanente tumor registries and histology data.

One attainable goal of health information technology (HIT) is the ability to continuously enhance quality and safety in the delivery of health care. Current healthcare financial incentives, which encourage high-cost, high-volume care, steer clinicians away from fully achievable low-cost, high-quality care. Steven Waldren, director of the Center for Health Information Technology at the American Academy of Family Physicians (AAFP), notes that the AAFP highlights the principle that data aggregation can drive and support multiple aspects of healthcare delivery, including quality initiatives, health services and

clinical research, public health, and transparency in reporting practices. As an advocate for the broader use of electronic health records (EHRs), a tool central to data aggregation and sharing, the AAFP encourages members to understand the power of EHRs as more than tools for administrative data. Confidentiality, standardization, and system use can be barriers to data aggregation and must be addressed. Several potential avenues for such improvements were described, including work in support of the American Society for Testing and Materials (ASTM) Continuity of Care Record (CCR) standard and work with the AQA Alliance to articulate the concept of a National Health Data Stewardship Entity.

Offered as a promising model for measuring impact on outcomes, the Society of Thoracic Surgeons' (STS) Adult Cardiac Surgery Database (ASCD) is highlighted by Peter Smith, a professor and the chief of thoracic surgery at Duke University. The ASCD, the largest of three distinct databases in the STS National Database, is a clinical registry aimed at continuous quality improvements in cardiac surgery. In an effort to push information to the bedside, the STS makes an individual risk calculator, including current risk adjustment information, available to the public. Smith highlighted multiple studies indicating that feedback from repository data can change physician behavior. In addition, ASCD data, in combination with administrative data, have been employed to illustrate the cost-effectiveness of continuous improvement initiatives, which could be shared with, and possibly replicated by, other medical specialties.

EMERGING LARGE-SCALE LINKED DATA SYSTEMS AND TOOLS

Peter Covitz, Ph.D.
Chief Operating Officer, National Cancer Institute

The mission of the National Cancer Institute is to reduce suffering and death from cancer. NCI leadership has determined that the scale of its enterprise has reached a level that demands new, more highly coordinated approaches to informatics resource development and management. The caBIG program was launched to meet this challenge. Its participants are organized into work spaces that tackle the various dimensions of the program. Two cross-cutting work spaces—one for Architecture and the other for Vocabularies and Common Data Elements—govern syntactic and semantic interoperability requirements. These work spaces provide best practices guidance for technology developers and for conducting reviews of system designs and data standards. Four domain work spaces build and test applications for Clinical Trials, Integrative Cancer Research, Imaging, and Tissue Banks and Pathology Tools, representing the highest priority areas defined by the

caBIG program members. Strategic-level work spaces govern caBIG requirements for Strategic Planning, Data Sharing, Documentation, and Training & Intellectual Capital.

In its first year, caBIG defined high-level interoperability and compatibility requirements for information models, common data elements, vocabularies, and programming interfaces. These categories were grouped into degrees of stringency, labeled as caBIG Bronze, Silver, and Gold levels of compatibility. The Silver level is quite stringent, and demands that systems adopt and implement standards for model-driven and service-oriented architecture, meta-data registration, controlled terminology, and application programming interfaces. The Gold level architecture consists of a data and analysis grid, named "caGrid." caBIG systems register with and plug into caGrid, which is based on the Globus Toolkit and a number of additional technologies, such as caCORE from the NCI and Mobius from Ohio State University.

Cancer: A Disease from Within

Cancer is a disease that comes from within, and researchers have to tease out the difference between cancer cells and normal cells. This task requires a molecular approach; in other words, we have to analyze things that are too small to see with light microscopes or the naked eye. Such an endeavor requires many specialties and subspecialties and areas of inquiry that cannot be practiced by any one individual, laboratory, or institution. In other words, this is not a local problem. This is not a regional problem. This is a problem for the entire cancer research community. It must be tackled on a national—and even an international—scale if it is going to be solved. This breadth of resources is needed because the disease (actually, many widely differing diseases that are categorized as "cancer") is complex. The fact that any given institution hasn't quite achieved what we would have liked is no one's fault.

The vision for caBIG arises from such considerations. The caBIG vision is to connect the cancer research community through a sharable, interoperable structure to employ and extend standard rules, in a common language, to more easily share information and to build or adapt tools for collecting, analyzing, integrating, and disseminating data and knowledge.

The challenges faced by caBIG are quite substantial, but are similar to those faced by those addressing the broader agenda of trying to share health information to improve patient care. caBIG is focused on the immediate problem of cancer. However, because cancer is so complex, it forces one to confront many general problems of biology and medicine. We therefore believe that caBIG is breaking ground and creating a path forward for many biomedical and healthcare disciplines. At the least, we

believe we are creating a possible model or prototype for the broader challenge of creating an interoperable health information network across the nation.

When we launched the caBIG program, we knew there would be tremendous technical challenges, but we were not, in the immediate sense, most concerned with those issues. Rather, we were more concerned about whether the organizational structure was going to be able to scale to the national level. We were keenly aware of failed attempts to consolidate the biomedical information technology market in the past, and of the challenges in deploying technology in the biomedical setting.

From Feudalism to Democracy

In our preliminary thinking, we looked at large conglomerations of people who have tried to organize themselves before. We hoped we could learn something from their models and their experiments, even if they were unrelated to health care, cancer research, or technology. So we looked at national governance models—starting with feudalism.

Feudalism is the prevailing model under which the National Institutes of Health (NIH) operates when it allocates grant-funded research money. Basically, the idea is to have a relatively limited monarchy of sorts, which the NIH represents, that is mostly engaged in the activity of finding and supporting the most capable members of the nobility out in the national terrain. Right now, of course, those lords are getting older and older, and the younger ones are not really able to access the monarch, so we have this problem of the aging research force in America.

Nonetheless, this system has some good qualities. It has been a great system for fostering the creativity of investigator-initiated research. That is why it grew the way it did and why the system has produced such notable successes. But remember, we are trying to build a national network for sharing data, which is not a research program; it's a technology implementation program. So we believed this model, although it is prevalent at the NIH, would not be appropriate for this particular activity, even though we are at the NIH and the NCI. The problem is that feudalism creates a warlord culture that simply offers too little incentive to cooperate. Regional medical networks share data, but there is no national-scale example that you can point to that has been successful, and this is because most of those programs operate under a feudal structure. The reason feudalism fell to the wayside in the course of human events was because it was inadequate for dealing with national levels of organization.

Today, many programs are in place that say, in effect, let's just get all the data in one place, then we will appoint one group as the coordinating center. Such efforts always seem to have a nice collegial name, such as the

coordinating center or central database. The idea is that these are the data experts and they will take care of all our data needs. We'll give them the money, and somehow or other, everything will just work itself out. But of course in practice this does not work. It's not that it never works, but that it only works for a small number of data types that have wide utility across the community.

This model does have some success stories, such as Genbank, a database of genetic sequences and DNA codes that researchers like to access. The codes are the same for all organisms, using the same four letters as part of the code, so it does make sense for that all to be centralized. Genbank has been a very successful, centralized database. People send in their DNA sequences when they sequence genes, and no one feels forced because it is just the norm, although during its inception there was a big debate over whether this should be a central database. There are several other examples of successful central national databases in biomedicine, but you can count them on the fingers of your two hands.

In cancer research, we deal with thousands of data types across a huge variety of studies; much of those data are subject to privacy restrictions. Thus, a collectivized, centralized database on a national level will not work. So we rejected that approach, and looked for another model—this time considering the notion of federal democracy. Why? In part because it strikes a balance between centralized management and local control. It's the best fit for what we were trying to achieve. It has worked pretty well in the United States and in other nations, and it's really the way to go if you want to get community participation and ownership but retain a mechanism for central leadership.

The *Federalist Papers* are a series of essays in which some of the founding fathers debated the pros and cons of overly centralized control versus overly dispersed and delegated authority. These were the beginnings of the debates about states' rights versus federal control. Those debates raged for 100 years and, in fact, you still see cases about states' rights. So it is an ongoing debate and there is no perfect comfort point. But in that tension, in that pull and that tug between centralized authority and local control, you actually wind up moving an entire field forward. This is not as comfortable a model as saying that we'll give the money out and the lords will take care of it. It's not nearly as comfortable as saying we'll just create the central database and then they'll take care of it. This model forces everybody to participate. This model is the most successful model in our view, not just because it's cancer research, but because it applies to biomedical data exchange in general.

From such deliberations, we created an organization that elaborates on this notion of control and oversight at the center, but nonetheless

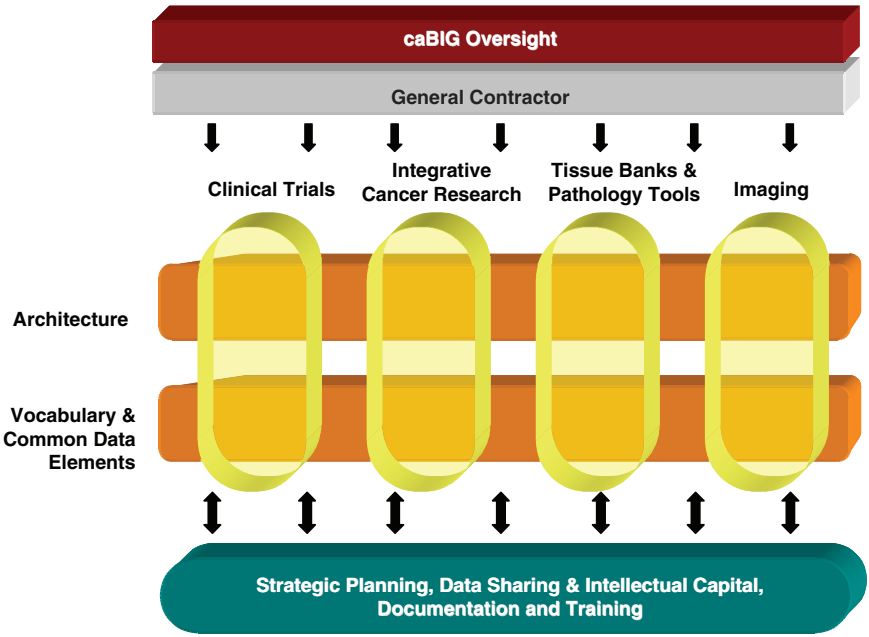


FIGURE 3-1 caBIG organizational model.

includes ample local participation and inclusion in decision making (see Figure 3-1).

caBIG Today

Our design currently includes four domains in cancer research. We can add more if necessary, but the ones in place now are based on a national discussion of priorities—they include clinical trials; a category of all-inclusive integrated cancer research that includes genomics, proteomics, and other molecular-level dimensions; tissue banking and pathology; and imaging.

We could have stopped there. We could have said we were going to get everyone in the nation who cares about clinical trials and then deal with technology issues for trials. But we decided we needed additional elements that could in essence hold the model together—hence, we created an architecture group, which deals with issues related to system architecture and technology choices and, importantly, a group devoted to vocabulary and common data elements, which deals with data standards and semantics. The idea was not for us just to build a closed system where everything works

internally. Rather, we wanted to specify the data and the semantics so that they could be exchanged with other such programs and systems, such as a National Medical Records System. If you specify the data independently of the system architecture, you get to do that. If you tie it all together, then you have created an internally closed system that isn't going to work well with others. We did not want to make that mistake.

We also have several groups focused on strategic planning. We realized we are not just a technology program; we also have to deal with issues of privacy, licensing, and public–private partnership. CaBIG is not just an academic, federally funded program—it involves and includes the private sector, and has from the start. We are actually beginning to see even greater interest in uptake by the private sector in caBIG program activities and technologies.

For the first 3 years, from 2004 to 2007, the program was a pilot managed by a general contractor, Booz Allen Hamilton. A report is available that summarizes the results of the pilot.[1] We attracted the interest of about 190 organizations. There were approximately 300 software projects and subprojects; 40 actual end-user applications were developed for those different domains referred to above. In a real technical tour de force, we built on a number of existing projects that were dealing with the issue of interoperability to create a semantic data grid called caGrid.[2] caGrid connects the disparate caBIG community systems.

Some caBIG applications are in the area of clinical trials and have been deployed in a variety of cancer centers. For these software projects, a very modular approach was taken. We tried not to create the massive, central, one-size-does-it-all for everybody. We broke the clinical trials problem down to a number of components: adverse events, data exchange, study participation, and a number of others. That gives a site flexibility to pick which components are going to be necessary for their operation and allows the use of different components. That's an important feature of the program. It means "rip and replace" is not a requirement, but adhering to standards is. At the same time, the user has multiple ways to access systems that adhere to the standards, including adapting existing systems. For the more basic life sciences, we have solutions for biobanking, for genomewide association study data management, imaging, and microarrays. The overall idea is to address a wide range of different disciplines and tie them together. Finally, we have a major activity in data-sharing security, with policy documents, templates, licenses, and other features that people in the program can use to wrap around and include in their own project.

[1] See https://cabig.nci.nih.gov/overview/pilotreport.

[2] See https://cabig.nci.nih.gov/workspaces/Architecture/caGrid.

A Changing Landscape and New Momentum

The landscape is changing and there is new momentum for creating publicly accessible registries of clinical trials. We are working closely with other groups who have been chosen to conduct this activity. In particular, the National Library of Medicine has a key role. There is also a role for different communities with additional requirements beyond the minimum standards needed for a national trial registry. Our plan is to implement cancer trial registries that are completely in conformance with, and contribute to, national and even international trial registries.

Having successfully completed the pilot, the caBIG management is now focused on expansion, roll-out, and deployment. Ultimately we'd like to connect all biomedical researchers, not just cancer researchers. In the development of caBIG, we heard loud and clear that people are at institutions that conduct research in many areas, not just cancer. Thus the overarching goals are to increase the speed and volume of data aggregation and dissemination; broaden the community; and really serve as a model for a scalable national infrastructure for molecular medicine.

caBIG enjoys support from key NIH leaders. Dr. John Niederhuber has been extremely supportive of the program, which was created prior to his taking up the leadership of the institute by the previous NCI director, Andrew von Eschenbach. The director of the entire NIH, Dr. Elias Zerhouni, has suggested that caBIG can serve as a model for other areas.

The NCI is rising to the challenge of cancer by recognizing it cannot just do business as usual; it must change the game. The NIH is often criticized for being conservative and safe, but I would submit that is an overgeneralization, and the spark of leadership and creativity can be found in programs such as caBIG.

NETWORKED DATA-SHARING AND STANDARDIZED REPORTING INITIATIVES

Pierre-André La Chance, B.S.
Chief Information Officer, Kaiser Permanente

The need for sharing data across research entities and among collaborators continues to grow at an astonishing pace. To meet these needs, data sharing solutions must be fast, cheap, sustainable, high quality—with understood meanings—and safe. Current work with the Center for Health Research (CHR) at Kaiser Permanente Northwest and its research collaborators successfully meets these criteria. This work also provides data sharing across entities that compose the NIH's Clinical and Translational

Science Award (CTSA) program, a consortium of academic health centers that is transforming the discipline of clinical and translational science. This paper will discuss these successes and how they can be extended to support data sharing across the CTSA program, even as many entities create their own clinical data repositories.

Data sharing has changed substantially over the past two decades at Kaiser Permanente. Research questions and data are far more complex. Today we are discovering the increased necessity of data sharing.

Data sharing is essential when identifying potential participants or related issues that might be available to researchers once a study has been approved. It is vital when information from one entity does not meet the researchers' needs. For example, we know that low-frequency events are important to study, but these demand larger pools of data. We also have strong needs for geographic and demographic diversity. In Portland, Oregon, for instance, we don't have a great deal of demographic diversity, so we are strongly motivated to find collaborators who can inject diversity into our research data.

Data sharing is required when studying varying practices and outcomes using entities as the unit of analysis. Translational research, for example, uses institutional entities as the unit of analysis so researchers can compare outcome differences and patterns of practice.

My point is that one entity simply does not have enough data to fulfill our research needs. This is just on the data side, of course; in today's global society, more and more researchers want to work with fellow scientists of experience and renown, who are spread across a number of different entities. We are always seeking additional ways to share and pool data. Therefore, out of necessity and over the years, we have been crafting a vision for data sharing.

CHR's Vision for Data Sharing

It Must Be Fast

The first requirement for any data-sharing solution is that it must be fast; it must be able to move data quickly. In past research projects, we may have taken up to 2 years to develop sharable data standards. Now we cannot afford such luxuries.

It Must Be Cheap

Data sharing must be done cost-effectively. Fifteen years ago we had a large vaccine safety datalink study that required us to share data among a number of institutions. When we first started the project, we found that

we used 90 percent of the research dollars just to create the poolable data. That left only 10 percent of the funding to pursue the science and the knowledge that the data produced. Clearly something had to change.

It Must Be Sustainable

We have to conduct data sharing in a sustainable manner. As CHR's chief information officer, I have viewed countless initiatives toward data warehousing and data sharing. Typically, they get about 2 or 3 years out then fail because the person who championed the project leaves the organization or because the effort became so grand and expensive that it simply collapsed.

It Must Be High Quality—with Understood Meanings

If our findings are used to make decisions that affect treatment and policy, we must be certain that our data are correct and of the highest quality. We also have to present the data in a way that is understandable. In other fields, we have seen extraordinary advancements in data warehousing and shared data, but often they have relatively simple operational data that are well understood. Health data can be very complex, so our bar is set even higher. The onus is on us to find methods and processes to inject higher quality into our data; furthermore, this involves providing metadata for a fuller understanding of the data's meanings and limits.

It Must Be Safe

Because we are largely a collaborative research entity, nearly all of our research projects rely on sharing or pooling data across institutions. However essential, we also acknowledge that sharing data can be dangerous. Therefore, tight controls are imperative. Out of respect for study participants—and this is largely a compliance issue—we can no longer proceed as some did in the 1960s, 1970s, 1980s, and early 1990s, sharing data without restriction, sometimes even recklessly.

We need to be able to share data with specific use guidelines. We know that some of the regulatory restrictions placed on us (for good reasons) have made performing preparatory research difficult. Our data sharing solutions must allow for such preparatory work, largely in the area of compliance. Once we have the appropriate Institutional Review Board (IRB) approvals and participant authorizations, we need assurances that we have reasonable ways of sharing data—even fully identified data, not just limited data sets.

CHR's Strategy: Virtual Data Warehousing

As stated earlier, our vision was to make data sharing fast, cheap, sustainable, high quality—with understood meanings—and safe. We needed to create a strategy to support that vision—and that was virtual data warehousing. The goal was to construct research-friendly, locally controlled data warehouses and associated data marts, without becoming prisoner to data warehousing methodologies. We also wanted to create networks of local interoperable data warehouses across collaborators to provide virtual data warehouses that would be well defined, at the byte level, the format level, and the standards and coding level.

How did we do that? At Kaiser Permanente, one of our largest areas and sources for data is our electronic health record. Unfortunately, at the operational level those data are in a hierarchical structure—MUMPS (Massachusetts General Hospital Utility Multi-Programming System), to be precise, which is virtually useless for querying and reporting. Initially, the MUMPS-based system was created to make things go fast so that the clinician wouldn't have to wait up to 3 seconds for a computer response.

We needed to fix this problem. To draw the value we need from hierarchical data, we must have some version of those data in a relational state that is optimized for querying and reporting. Operational data, while necessary, do not sufficiently meet our needs. Relational data are *sufficient*, but not *efficient*; that is, they support payment, treatment, and operations, but they do not support research. Relational data, for example, tend to be departmental, designed to clarify what is happening in the inpatient or outpatient setting.

By contrast, what we need for research to be thoroughly optimized are data stores that focus on participants, patients, or disease areas. Therefore, we must take those data and create a second-level, locally controlled data warehouse that is optimized and research friendly. Without this step, we would have to go back to the relational data structure for payment, treatment, and operations, and do the acquisition, transformation, and publishing to derive research-friendly data—for every research study. This would be like going back in time to the vaccine study we did 15 years ago, when 90 percent of the research dollars went to answering the research question—and we know we don't want to repeat that.

With research-based, locally controlled data warehouses, we are able internally to use those data quickly, cheaply, and expertly. Our ultimate goal is to include data pooling with partners who share our methods, so that when we have the appropriate approvals, we can pool those data and use them across entities. In fact, we have done this successfully within the Health Maintenance Organization Research Network (HMORN), which has 15 members. Not all of those members have committed to this practice,

but many have. The important thing is that we can collaboratively show data quickly, successfully, for more than 10 million patients per month.

We have also shared data successfully within the Oregon Clinical and Translational Research Institute and between a health maintenance organization and a university medical setting. A key question for us now is this: Do we want to take the HMORN methodologies and use them within and across CTSAs to find out which data are most useful and how we can advance data-sharing methodologies?

Creating Sharable Versions of Data

CHR is working with Oregon Health & Science University (OHSU) to recreate sharable versions of data that include enrollment, demographics, a tumor registry, a pharmacy, vital signs, procedures, diagnosis, and laboratories. These will be locally constructed and protected by governance rules and honest brokers to ensure that they are not shared without the appropriate approvals in place.

At the same time, we are building a biolibrary that allows researchers from both institutions to go across Kaiser Permanente tumor registries and histology data. They are able to access electronic inventories of slides and blocks in a way that quickly and easily helps scientists identify patients with specific diagnoses and stages of cancer—and then connect to those patients' respective tissues. We believe this will save time dramatically; when we supplied similar requests in the past without such streamlining, the process took twice as long.

We also have set up the ability to work with OHSU scientists so they can find those retrospective fixed-formalin samples and identify participants of interest. This information will facilitate the acquisition of fresh tissue, appropriate authorization and consent from the patient, and successful collection and delivery to the research scientist.

Deidentified Data Marts, Counters, or Cubes

CHR is also focused on a specific aspect of data warehousing that involves counters or specific data marts. These can be shared across entities because the resulting data are deidentified, even for preparatory research purposes. The most important thing about counters or data marts, especially for preparatory research, is the speed with which these data can be shared. Data may be shared within an hour or, in most cases, in less than a day. This time line is crucial in meeting the lifecycle of a proposal that cannot wait 2 or 3 months to determine whether there are enough participants and appropriate tissue samples to move forward. Our cancer counter is an excellent example: We have a deidentified cancer cube that researchers can

use to see which participants might be usable within a study in response to a proposal.

This strategy has been practically tested within HMORN and the cancer research network. Simply put, it works—in terms of speed, quality, and especially compliance. With data pooling, compliance is a key issue and is vital to the process. The other key issue is defining and getting agreement on data ownership. We must contractually define how data are used and whether additional disclosures are required; then we must enact data use agreements, use agreements (if not limited data), IRB approvals, or some other sort of agreement. This is as vital as the technology; sometimes the political and social issues are more of a roadblock than the technical issues.

There is a fierce urgency now to advance on these ideas—especially among those who are ill. I have experienced years of discussions on the need to share data, and I look forward to the blossoming of some of the tools we have today. At this moment we have patients, citizens, and research participants who are depending on us to deliver information to them right now to improve or extend their lives. We cannot wait any longer to develop and roll out these tools. As we wait for this strategic unfolding, let's not lose track of the tactical tools we already have today to move this work forward.

LARGE HEALTH DATABASE AGGREGATION

Steven Waldren, M.D., M.S.
Director of the Center for Health Information Technology,
American Academy of Family Physicians

The holy grail of health information technology is its ability to drive rapid improvement in the quality and safety of healthcare delivery. To reach this goal, data must be correctly entered into applications in a structured and coded form. From there, data can be analyzed locally and aggregated into large health databases for further analysis at the population level. Given the current healthcare system, including rapidly growing health spending and increasing chronic disease burden, why has the industry not adopted and used this technology? The reason, like nearly any application of information technology, is not with the technology, but rather the business model.

The current financing of health care rewards high-cost, high-volume care—not low-cost, high-quality care. This produces a disconnect between those that need to invest in the technology and those that will financially benefit from the technology. Despite this misalignment, physicians are adopting health information technology to improve the efficiency in their offices and to improve the quality of care delivery. Physicians see the value of adopt-

ing EHRs to help them with documentation and managing complex patients, although this vision is tempered by the misalignment of payment. To realize this value, physicians are not required to highly structure and code clinical information, which is a prerequisite to data analysis and aggregation. As an industry, we need to produce products and services that allow physicians to value entry of highly structured and coded clinical information. Increasing this value can be through providing a means to manage and effectively document patient care or by lowering the financial barriers to entry of coded clinical information (e.g., increased payment for quality).

The American Academy of Family Physicians has been working to move the industry forward on these fronts: providing valued services, lowering technology costs, and advocating for quality-based payment. Standardization is an important tool to lower the costs of these systems and to provide a platform to build value. AAFP has worked to establish and promote HIT standards that are focused on clinical data, such as the ASTM Continuity of Care Record standard. The CCR has become the first widely available HIT content standard for core clinical patient data. The AAFP is now exploring the next step in establishing a clinical data repository for its members. The purpose of the CDR would be to promote three areas: quality improvement, pay for performance, and patient-centered medical home transformation. This repository would give physicians a set of services so they appreciate the value in entering highly structured, coded data. These services will likely borrow from the success of our medical association colleagues such as the American College of Cardiology and the Society of Thoracic Surgeons.

The first obstacle for data aggregation in the ambulatory environment is the adoption of standards-based HIT in the practice. The AAFP has spent 4 years to bring our members to the point that we can start aggregating data. Our work has been facilitating adoption of HIT in our membership and driving data standards. Membership EHR adoption is now between 40 and 50 percent. A member survey in late 2007 found that approximately 37 percent have a fully (by their definition) implemented electronic medical record system. Another 13 percent said they had purchased one or were in the process of implementation. When we examined the survey results more closely regarding individual functionalities used, the results were not particularly surprising. The functionalities adopted were in large part to help physicians with documentation, billing, and remote access to the EHR. Functionalities adopted did not focus on quality improvement, e-prescribing, or population management. In short, physicians in the field are adopting the EHR and other technologies, not in the interest of data aggregation but because they are under business constraints to obtain complete and accurate documentation and to maintain productivity. Additional functionalities must be adopted to achieve data aggregation at any level other than administrative data.

Data Collection Lifecycle

Obviously, if you do not have the data you cannot aggregate them. The data collection process has its own lifecycle. Today, for many healthcare organizations, the lifecycle does not include the entry of structured and coded data. Some organizations are still dictating all clinical documents. One can aggregate those documents, but that will not produce meaningful data aggregation or analysis.

Once data are coded, structured, and entered into an information system, the data needs to be in some type of standardized format so they can be aggregated across multiple systems and organizations. We found there is no good set of standards to do that. Many standards apply to health care, yet they are all about messages and documents; they are not about datasets and aggregating data. However, the ASTM CCR is an exception as it represents a patient-centric dataset that can be aggregated from multiple sources.

Another question is about policies relating to data aggregation, where privacy and confidentiality are concerns not only for the patient, but also the provider. Privacy and confidentiality, as well as security, issues can be a real impediment to data aggregation. If data privacy is not ensured, patients, clinicians, and healthcare organizations will not share their data. A national entity is needed to address these issues by establishing best practice standards for data aggregation. We have been working in the AQA Alliance to articulate the concept of a National Health Data Stewardship Entity. Two documents at http://www.aqaalliance.org describe the concept of the Entity, which is intended to define operating rules for data aggregation to ensure quality. We are finding that many of our members and hospitals say they cannot share data with aggregators because it would be a violation of the *Health Insurance Portability and Accountability Act*, but this is not always the case. We must get rid of the myths about privacy and security concerns and establish best practices for these issues to get buy-in from all stakeholders.

After Data Are Aggregated

Another issue is what happens to aggregated data. You have to do something with the data—you may have created and shared them, but they have no value until you start to use them; one of the Roundtable themes is to ensure that publicly funded data are used for the public benefit. A fundamental question is, why aggregate these data? Reasons include:

- Quality improvement
- Public reporting

- Health services research
- Clinical research
- Healthcare value analysis
- Biosurveillance
- Population management
- Public health

This is just a small list of some of the factors that AAFP members value. One of their top priorities is quality improvement, yet business constraints make it difficult for them to implement new technologies and develop new ways to improve quality.

Another priority of data aggregation is clinical research. We have established a practiced-based research network with thousands of physicians. We have a subnetwork of those with EHRs, and we are starting to aggregate those data. This does present challenges; for example, different electronic medical record systems codify data in different ways. The data models of the EHR differ not only across vendors but also across practices using the same EHR product. To deal with this diversity, we must either define a standard dataset for the vendor to produce or map each EHR database to a standard dataset. For now, at least, the decision is to map each EHR database. We have been working with a company to actually map the individual data structures to a common data structure based on the ASTM CCR.

Healthcare data aggregation and analysis is in the best interest of the U.S. healthcare system. In the short term, the potential financial and privacy risks to individual stakeholders must be managed appropriately. Without access to data and their analysis, one cannot improve quality, increase safety, or appropriately provide incentives for cost-effective care.

REGISTRIES AND CARE WITH EVIDENCE DEVELOPMENT

Peter Smith, Ph.D.
Professor and Chief of Thoracic Surgery, Duke University

Background

Cardiac surgery procedures are the most well-studied, evidence-based procedures performed today. Because of the high-risk, high-benefit, high-cost nature of the procedures, especially coronary artery bypass grafting (CABG), considerable research effort has been expended to ensure appropriateness of use. The Department of Veterans Affairs Cardiac Surgery Advisory Group was formed in 1972, creating the first multi-institutional database monitoring cardiac surgery outcomes; it monitored volume and unadjusted operative mortality. Seminal randomized clinical trials first dem-

onstrated life prolongation in patients with left main coronary disease. Subsequently, this benefit was demonstrated in patients with disease in two or more vessels when there was involvement of the left anterior descending coronary artery. Furthermore, CABG was shown to provide more benefit to patients with impaired ventricular function and to those with diabetes mellitus. These findings highlight the key principle that increased prospective risk of surgical mortality in CABG goes hand in hand with increased, demonstrated longevity benefit compared to medical therapy for all patients, and compared to percutaneous intervention for patients with multivessel disease (Smith et al., 2006).

In 1986 the Health Care Financing Administration (now the Centers for Medicare & Medicaid Services, or CMS) released a list of hospitals that had high risk-adjusted mortality rates for Medicare patients. This list received a great deal of notice because the risk-adjustment algorithm employed was, of necessity, based on administrative rather than clinical data, and was likely to be relatively ineffective. The potential for the unintended consequence of reducing the overall benefit of cardiac surgery through individual avoidance of risk by surgeons was viewed as a serious problem (DeLong et al., 2005). Accordingly, this initiative by the federal government stimulated the establishment of The Northern New England Consortium and the Society of Thoracic Surgeons' Adult Cardiac Surgery Database to ensure that our profession had accurate risk-adjusted information on surgical performance.

The characteristics of the STS ASCD, a *clinical database*, was a response to the weaknesses inherent in *administrative databases* that should be clearly understood. In general, administrative databases contain prospectively collected demographic and financial information and retrospectively collected diagnostic and therapeutic information developed by professional hospital coders from chart review. Because of the way in which diagnoses are encoded, it cannot be determined if they were preexisting (and therefore are risk factors) or occurred as a result of the encounter (and therefore are complications). Additionally, there are financial incentives to "upcode" diagnoses, a process that will degrade their utility in risk adjustment. The coding process is designed to detect all possible diagnoses, which are then frequently sorted by financial importance and truncated in transmission, further reducing the applicability of the risk profile created for the encounter. Because the coding process is not specific to an encounter or disease, the absence of a diagnosis does not definitively mean that it was not present, a distinct liability. Finally, many diagnoses used for risk adjustment are "synthesized" from financial events occurring in the encounter, through questionable methodologies.

By contrast, a clinical database is designed to collect prespecified risk factors prospectively. Prespecification results in knowledge regarding the

definitive presence or absence of the most important risk-adjustment variables. Coupled with clinical definitions, the presence or absence of the specific diagnosis is free of bias. Complications and risk factors are separately recorded, assuring that the important distinction can be made. Finally, database entry is generally performed by clinical staff rather than professional coders, promoting accuracy of the data entered.

Database Description

The STS ACSD is the largest database of the three distinct databases that make up the STS National Database. The ACSD is a voluntary clinical registry developed for the purpose of continuous quality improvement in cardiac surgery. It contains more than 3 million surgical procedure records from 857 participant groups in 49 states, representing approximately 80 percent of adult cardiac surgical procedures performed nationally.

The database contains more than 250 data elements for each patient encounter. The data elements are prespecified and associated with clear definitions, particularly for medical conditions that affect surgical risk and for known complications of cardiac surgical procedures. The data dictionary has been modified extensively over the years as the nature of important risk factors has become clarified.

Data Entry and Verification

The current method of data entry is through certified software vendors, who must provide key-entry verification and real-time access to the field definitions. Several sites, including our own, have legacy systems for database predecessors to the STS ACSD. These system front-ends were unaffected by the creation of the STS ACSD, which harvests information from locally developed back-ends that are compliant with the STS data standards. Our system has been modified to a web-based application that can be pushed to a handheld PC. The web-based and handheld applications support patient lists for mid-level providers, who populate the local database as part of the care process. At the time of entry, most of the fields are thus capable of entry checking to increase data validity and reduce data entry error. STS is exploring the development of a web-based front-end for national use.

The STS data elements are important characteristics of each patient encounter, and are defined clinically for entry by care providers at the point of care. At Duke, interim reports are generated to become part of the ongoing patient medical record. The most important of these is the operative note, which is completely generated from the constructed database and provides a primary motivation for timeliness and accuracy.

Data are harvested quarterly and are subjected to extensive evaluation

against predefined norms to ensure accuracy. Risk adjustment algorithms have been created for mortality, morbidity, length of stay, prolonged ventilation, deep sternal wound infection, stroke, and renal failure. These are the most common serious complications patients encounter, and they contribute to both efficacy and expense of these procedures. The risk adjustment algorithms are updated with each data harvest and the included variables, coefficients, and algorithm intercepts are published for public scrutiny and use.

Sharing: Aggregate Data

The ACSD has been studied extensively using observational analytic methods, resulting in 69 publications and presentations. It has been shown to be more accurate than administrative databases and has been selectively audited and endorsed for public use in several states.

The ACSD was queried from 1995 to 1999 on the use of the internal mammary artery (IMA) as a conduit in bypass grafting. That has been well known to provide a survival advantage, but the penetration of its use has not been as great as it might be—only 60 percent in 1995. With publication of these observations (Ferguson et al., 2003), there was increasing prevalence of IMA use and improved outcome. These data have now been employed to support IMA use as a process measure in the CMS Physician Quality Reporting Initiative (PQRI).

In a similar project, supported by an Agency for Healthcare Research and Quality (AHRQ) grant, there was a focus on the use of beta blockers before cardiac surgery (Ferguson et al., 2002). In this study, we found that intervention (reporting actual use by the institution compared to national use, accompanied by evidence supporting use) caused an increase in beta blockade that was sustained through the 60 months of the study. This was also identified as a CMS PQRI measure, based on this evidence.

Several publications are available that are relevant to the concept of regionalization of surgery to promote high-quality outcomes (DiSesa et al., 2006; Peterson et al., 2004; Welke et al., 2007). Comparing coronary bypass grafting risk-adjusted mortality to annual hospital volume, for example, low-volume programs have a wide dispersion and high variance in risk-adjusted mortality compared to higher volume centers. The trend line is for higher volume centers to have better results. This has some important public policy implications in that evidence supporting expansion of primary percutaneous coronary intervention (PCI) sites for acute myocardial infarction is also promoting the fragmentation of cardiac surgery. The need for site-of-service cardiac surgery presence to promote better public access to primary PCI programs appears to be unsupported by evidence, yet it is commonly a regulatory requirement in many states.

ASCD data have been used to improve the accuracy of the Medicare

Physician Fee Schedule (Smith et al., 2007). A variety of inaccuracies had accrued in the Fee Schedule due to the survey methodology used to determine (estimate) physician procedure time. In the regular 5-year review, 600,000 cases from the STS database were analyzed to provide the actual times. Employing these times allowed these procedures to be accurately valued, removing many payment anomalies that resulted in inappropriate payment incentive, sometimes for inferior procedures! This use has engendered great discussion by all professional societies, and provided an important stimulus to develop similar clinical databases.

Finally, the ASCD has been employed to analyze the cost of complications, and resulted in the creation of the Virginia consortium through which a group of Virginia hospitals who participate in the STS database self-report results. By linking STS data to hospital cost systems, the institutional average cost per case has been compared to the observed/expected mortality ratio. This has shown that cost actually goes down as quality improves, principally due to the reduction of the cost of complications, which can be reduced by continuous quality improvement methods.

Sharing: Individual Surgeon and Program Data

The founding philosophy of the ACSD was to collect aggregate data, risk adjust nationally, and feed the information back to individual surgeons. Individual data are provided as well as deidentified group and regional and national data for benchmarking. As a result, over the past 7 years, STS/AHRQ grant programs have demonstrated that the use of a clinical data repository and feedback can rapidly change physician behavior on a national scale. This scientifically validated process cannot be accomplished using administrative data alone. The impact of this shared knowledge has been profound. Between 1994 and 2003, predicted coronary bypass grafting operative mortality has increased while observed mortality has declined (Figure 3-2) (Shroyer et al., 2003; Welke et al., 2004). In other words, there has been increased risk *and* improved performance. Its success is best demonstrated by a display of the "observed" divided by the "expected" ratio. The trend, using a constant model over time, has been downward, thus showing improvement in mortality outcome. Also, it is important to note that this constitutes professional self-regulation rather than mandatory external regulation (Figure 3-2).

STS currently has a set of 21 performance measures developed using the database and endorsed by the National Quality Forum (NQF). Compliance with this set of measures is specifically followed through a unique reporting mechanism to database participants. Recognizing the importance of composite measures, STS has developed a composite measure based on the NQF individual measure set. This hospital-level composite measure is

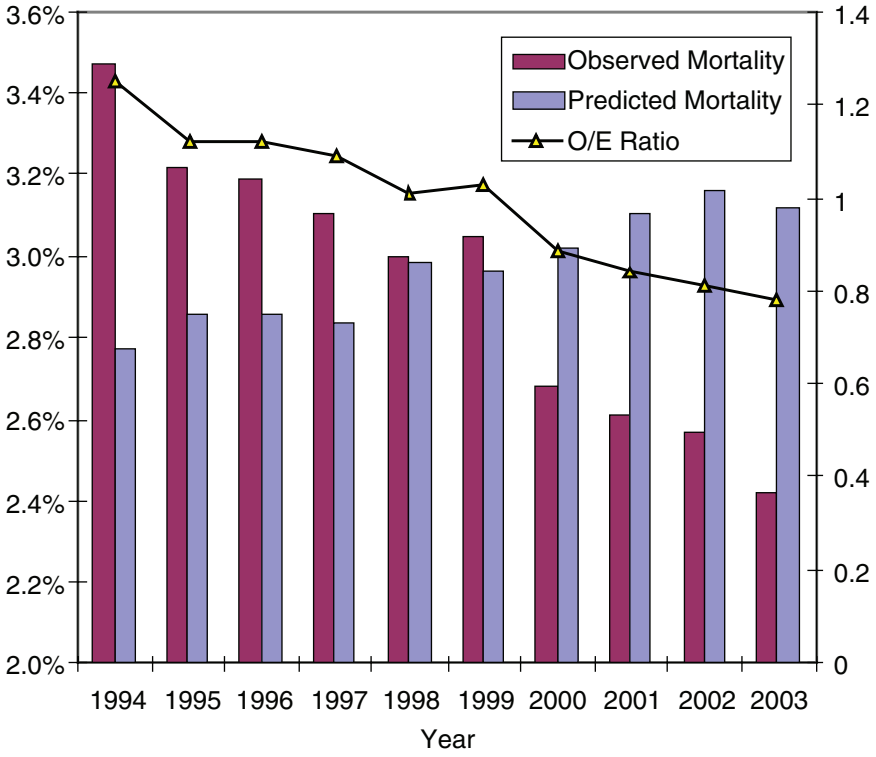


FIGURE 3-2 Observed and risk-adjusted coronary bypass grafting operative mortality trend between 1994 and 2003.
NOTE: O/E = observed/expected.
SOURCE: STS Adult Cardiac Database Isolated CABG Cases, 1994–2003.

reported as a one-, two-, or three-star rating system and is provided to all database participants.

Sharing: With Patients and Providers

The publicly available individual patient STS risk calculator,[3] based on the most recent risk-adjustment algorithms, has been developed to rapidly disseminate knowledge to the bedside. This web-based online risk calculator can be employed to calculate prospective risk for a potential surgical patient with similar characteristics. Because it is publicly available, individual patients and their primary physicians can access the information

[3] See http://66.89.112.110/STSWebRiskCalc261/.

to improve their decision-making capability. This site, and other resources such as the risk algorithm coefficients and intercept, are examples of transparency, which is essential in evaluating all information sources that provide outcome corrected for inherent patient risk.

Sharing

With Other Databases

The value of linking this robust and accurate clinical database to other administrative and clinical databases is unlimited. The establishment of linkages can enhance the overall knowledge base, extend follow-up for adverse outcomes, and permit comparisons of providers, provider environments, and alternative therapies.

With Administrative Databases

The STS database currently follows patients only for 30 days or in-hospital, thus reporting short-term mortality and morbidity results. The relationship of long-term outcome to procedural intervention is of particular value in assessing chronic diseases. Therefore it is imperative that we expand the outcome horizon using the National Death Index or the Social Security Death Index. The expense of the National Death Index may be prohibitive unless a payment mechanism is found. At this time, only a few STS sites can afford to independently accrue long-term information.

Parallel linkage, adding new information by matching STS patient encounters to administrative databases, can also add tremendous value. An enormous amount of accurate, detailed information is available from hospital systems regarding drug use and other diagnoses and procedures that are not prespecified for inclusion in the STS database. These data can be matched directly, or through linkage to the secondary proprietary industry that has evolved to promote continuous improvement through benchmarking hospitals. The melding of these databases would vastly improve the ability to risk-adjust outcomes and to detect low-probability events that occur with too little frequency to justify prospective collection by STS. The recently announced Food and Drug Administration (FDA) Sentinel Initiative is likely to achieve this result through the Reagan-Udall Foundation, which will create public–private consortiums to share data, with the goal to provide postmarketing surveillance of drug safety.

The CMS longitudinal database for Medicare beneficiaries is perhaps the most critical linkage that has yet to be fully realized. Here, two-way interaction would provide the ability to more accurately risk-adjust outcome, provide access to all professional and hospital encounters following

a cardiac procedure, provide an unlimited outcome horizon, and provide an avenue for direct physician participation in CMS quality improvement initiatives such as the PQRI. Furthermore, this linkage would provide the ability to transform the PQRI from a process-oriented, budget-neutral (within Part B) program into an outcome-oriented, budget-neutral (within Parts A and B) program that would provide meaningful and appropriate incentives for physicians. Ensuring that publicly funded data are used for the public benefit is one of the themes of the Institute of Medicine's Roundtable.

The critical importance of the role of CMS cannot be overemphasized. The powerful linkage of physician payment to the documentation of medical necessity remains essentially unexplored. CMS has the capability to develop beneficiary problem lists, and therefore national problem lists, through beneficiary-registering evaluation and management services. If the national will is to promote better payment for evaluation and management services, an initiative to pay for accurate clinical databasing would best serve the national interest and provide a better ability to assess the impact of fiscal interventions.

With Other Clinical Databases

Other existing clinical databases, administered by other professional societies or disciplines, have the potential to provide insight into other aspects of a patient's health status.

An example of this potential has been realized at our Duke University STS site, where we have linked an enhanced STS dataset to the local Duke cardiology registry, which is a subset of the American College of Cardiology and its national registry, the National Cardiovascular Data Registry (NCDR). By adding longitudinal outcome data through locally supported patient contact and interaction with the National Death Index, we have been able to evaluate the three main treatment alternatives for patients with documented coronary artery disease (coronary artery bypass grafting, percutaneous coronary intervention such as stenting, or optimal medical therapy).

We published data on 26,000 patients with significant coronary disease (Smith et al., 2006). We looked at the percentage of patients treated with PCI, and found that it increased from 20 percent in 1986 to about 60 percent in 2005. The data show the introduction of the bare metal stent, and then the drug-eluting stent, were associated with these trends, and with the overall downward trend of coronary bypass grafting. In evaluating the outcome impact of these trends in treatment selection, we found that there was a longevity benefit with CABG compared to PCI (Figure 3-3) and that the CABG advantage was *increasing* despite theoretical advances in PCI (Figure 3-4) and PCI's increasing application. These findings were replicated

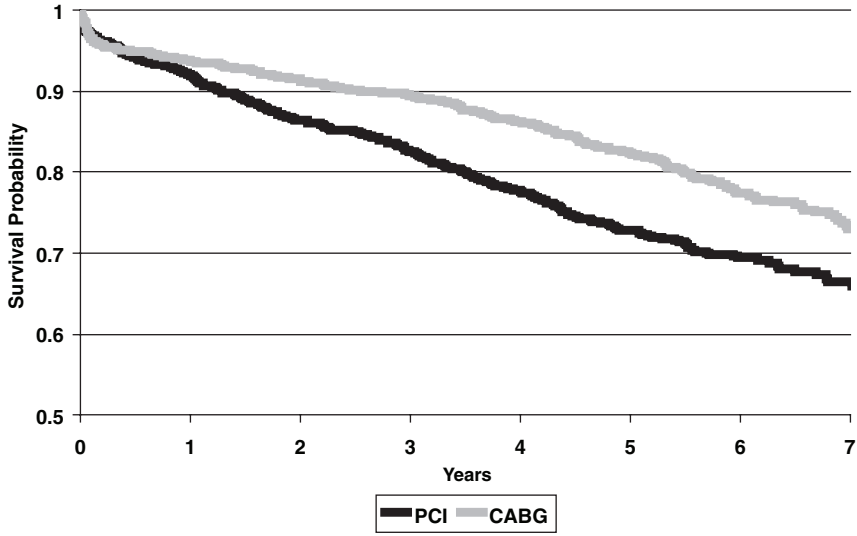


FIGURE 3-3 Trend comparison of CABG to PCI.

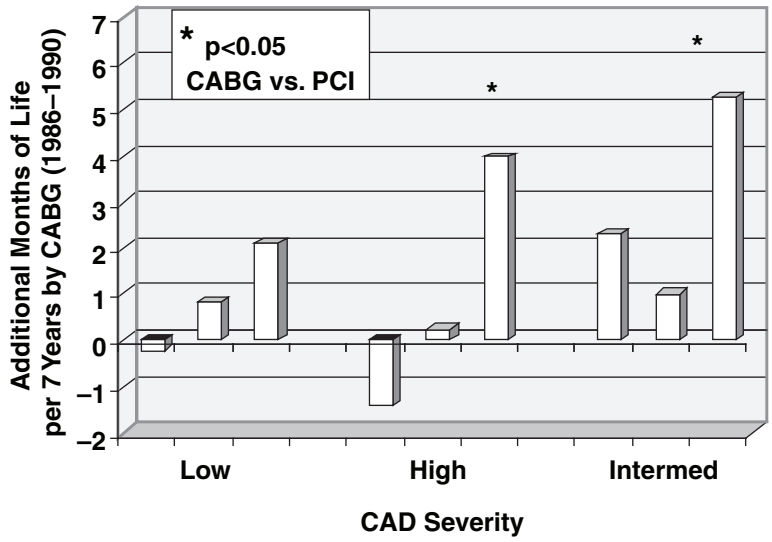


FIGURE 3-4 Increase in CABG advantage despite theoretical advances in PCI. SOURCE: This article was published in the *Annals of Thoracic Surgery*, Vol. 82, Smith, PK et al., "Selection of Surgical or Percutaneous Coronary Intervention Provides Differential Longevity Benefit," pp. 1420–1429, © Elsevier (2006).

in three other regions: northern New England (Malenka et al., 2005), the Midwest (Brener et al., 2004), and New York (Hannan et al., 2005)—made possible by clinical data sharing among specialties.

The logical extension of this work is to develop a national partnership with the American College of Cardiology NCDR, so that all sites participating in both STS and NCDR could share data. If augmented, at a minimum, by the National Death Index, it would become possible to better evaluate competitive/complementary therapies for cardiovascular disease. In addition, this would enable protection of the public health via liaison with the FDA to follow postmarketing outcomes of a variety of devices being introduced as treatment options for cardiac patients (e.g., new coronary stents and percutaneous valve repair/replacement devices).

Challenges to Database Use for the Public Good

The first challenge to data sharing for the public good is the willingness of data owners to be transparent and enabling. In large part this is because of valid concerns about proper use of aggregate information, primarily because of the *difficulties inherent in observational data analysis*. National standards and consensus regarding reliable analytic methodology and publication requirements are lacking, and much of the public reporting relies on administrative data adjusted by proprietary methods that prevent verification. Standards and consensus are essential to achieve the stated goals.

The second challenge is to develop systems that collect data as coded clinical information as a natural component of the patient care process, along with the resources to make this cost neutral to care providers. This will require cooperation among professional societies to standardize medical language, and cooperation with payers to reward participation in this process because it is for the common good.

The Role of Payers and the Professions

We are entering an era when payers are leveraging their control of physician and hospital payments to promote improvement in the nation's health. Their reliance on administrative data and process measures, and their reluctance to compensate for participation in the reporting of reliable clinical information and for outcome improvement, is concerning. The pathway to the least common denominator, which today is physicians and physician groups who lack access to reliable and pertinent aggregate patient information, will not suffice.

The development and utility of the STS database is hopefully an example of professional self-regulation that should be promoted. Only through transparency of information and self-reporting through databases will we

restore the public trust in physicians. Clinical databasing and data sharing, promoted by the health system, will empower the fulfillment of professional responsibility.

REFERENCES

Brener, S. J., B. W. Lytle, I. P. Casserly, J. P. Schneider, E. J. Topol, and M. S. Lauer. 2004. Propensity analysis of long-term survival after surgical or percutaneous revascularization in patients with multivessel coronary artery disease and high-risk features. *Circulation* 109(19):2290–2995.

DeLong, E. R., L. P. Coombs, T. B. Ferguson, Jr., and E. D. Peterson. 2005. The evaluation of treatment when center-specific selection criteria vary with respect to patient risk. *Biometrics* 61(4):942–949.

DiSesa, V. J., S. M. O'Brien, K. F. Welke, S. M. Beland, C. K. Haan, M. S. Vaughan-Sarrazin, and E. D. Peterson. 2006. Contemporary impact of state certificate-of-need regulations for cardiac surgery: An analysis using the Society of Thoracic Surgeons' National Cardiac Surgery Database. *Circulation* 114(20):2122–2129.

Ferguson, T. B., Jr., L. P. Coombs, and E. D. Peterson. 2002. Preoperative beta-blocker use and mortality and morbidity following CABG surgery in North America. *Journal of the American Medical Association* 287(17):2221–2227.

Ferguson, T. B., Jr., E. D. Peterson, L. P. Coombs, M. C. Eiken, M. L. Carey, F. L. Grover, and E. R. DeLong. 2003. Use of continuous quality improvement to increase use of process measures in patients undergoing coronary artery bypass graft surgery: A randomized controlled trial. *Journal of the American Medical Association* 290(1):49–56.

Hannan, E. L., M. J. Racz, G. Walford, R. H. Jones, T. J. Ryan, E. Bennett, A. T. Culliford, O. W. Isom, J. P. Gold, and E. A. Rose. 2005. Long-term outcomes of coronary-artery bypass grafting versus stent implantation. *New England Journal of Medicine* 352(21):2174–2183.

Malenka, D. J., B. J. Leavitt, M. J. Hearne, J. F. Robb, Y. R. Baribeau, T. J. Ryan, R. E. Helm, M. A. Kellett, H. L. Dauerman, L. J. Dacey, M. T. Silver, P. N. VerLee, P. W. Weldner, B. D. Hettleman, E. M. Olmstead, W. D. Piper, and G. T. O'Connor. 2005. Comparing long-term survival of patients with multivessel coronary disease after CABG or PCI: Analysis of BARI-like patients in northern New England. *Circulation* 112(9 Suppl): I371–I376.

Peterson, E. D., L. P. Coombs, E. R. DeLong, C. K. Haan, and T. B. Ferguson. 2004. Procedural volume as a marker of quality for CABG surgery. *Journal of the American Medical Association* 291(2):195–201.

Shroyer, A. L., L. P. Coombs, E. D. Peterson, M. C. Eiken, E. R. DeLong, A. Chen, T. B. Ferguson, Jr., F. L. Grover, and F. H. Edwards. 2003. The Society of Thoracic Surgeons: 30-day operative mortality and morbidity risk models. *Annals of Thoracic Surgery* 75(6):1856–1864; discussion 1864–1865.

Smith, P. K., R. M. Califf, R. H. Tuttle, L. K. Shaw, K. L. Lee, E. R. DeLong, R. E. Lilly, M. H. Sketch, Jr., E. D. Peterson, and R. H. Jones. 2006. Selection of surgical or percutaneous coronary intervention provides differential longevity benefit. *Annals of Thoracic Surgery* 82(4):1420–1428; discussion 1428–1429.

Smith, P. K., J. E. Mayer, Jr., K. R. Kanter, V. J. DiSesa, J. M. Levett, C. D. Wright, F. C. Nichols, III, and K. S. Naunheim. 2007. Physician payment for 2007: A description of the process by which major changes in valuation of cardiothoracic surgical procedures occurred. *Annals of Thoracic Surgery* 83(1):12–20.

Surgeons, Society of Thoracic. 1994–2003. Adult Cardiac Database Isolated CABG Cases.

Welke, K. F., T. B. Ferguson, Jr., L. P. Coombs, R. S. Dokholyan, C. J. Murray, M. A. Schrader, and E. D. Peterson. 2004. Validity of the Society of Thoracic Surgeons National Adult Cardiac Surgery Database. *Annals of Thoracic Surgery* 77(4):1137–1139.

Welke, K. F. , E. D. Peterson, M. S. Vaughan-Sarrazin, S. M. O'Brien, G. E. Rosenthal, G. J. Shook, R. S. Dokholyan, C. K. Haan, and T.B. Ferguson. 2007. A comparison of cardiac surgical volumes and mortality rates between the Society of Thoracic Surgeons and Medicare datasets. *Annals of Thoracic Surgery* 84:1538–1546.

4

Healthcare Data: Public Good or Private Property?

INTRODUCTION

By virtue of the origins of clinical data with individual patients, and because these data are often compiled with public funds, they have many characteristics of a public good or public utility. This situation suggests implicitly that these data should be shared widely and used for the common good of improving the nation's health and healthcare system. However, private entities also collect and analyze clinical data, often at great expense; place a proprietary value on clinical data; and protect these data as their own intellectual property. One of the goals of *Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good* was to evaluate the nature of goods, both public and private, in the healthcare data marketplace and to propose concepts, opportunities, and guidance for improving access to and sharing of medical data. This chapter reviews perspectives on clinical data; effects of the medical care data marketplace on research priorities, gaps, and possibilities; characteristics of a public good or utility—and on which dimensions healthcare data compare; distinctions that can be made within data types or sources; barriers to broader sharing of and access to medical data; and the conceptual advances, guidance, or policy needed.

David Blumenthal, director of the Institute for Health Policy at Massachusetts General Hospital and Partners Health System, and now the U.S. national coordinator for Health Information Technology, describes the theoretical concept of a public good as a way to guide practical policy development around clinical data. Using biomedical research as an example,

Blumenthal explores how research data can have characteristics of a public good while simultaneously holding significant value and inherent costs as a private database asset. In addition, he discussed how taxpayer-funded data, collected and stored in a variety of public and private institutions, provide another opportunity to consider such data a public good. With taxpayer-funded data there is an obligation to evaluate the incentives of data aggregation along with the benefits of making such data more available. Ultimately this may lead to opportunities to have the public, legal, and legislative arenas address the future utility of clinical data.

The potential to support evidence-based medicine through the wide variety of prescription drug and medical databases continues to grow because these data can offer greater insight into the practices of care delivery and safety surveillance. Current data sources have been constructed to serve as potential resources for research and commercial endeavors. William Crown, president of i3 Innovus, offers ideas on the elements to consider in building large, multifaceted data assets. From a private-sector perspective, Crown outlines some of the potential standardization, privacy, and statistical challenges associated with data aggregation and provides insight into the variety of sources of clinical data. As guidance for future database developments, he characterizes the increasing demand for a data resource that draws information from multiple, diverse sources of medical data and, in turn, synthesizes those data into a tool available for a wide range of healthcare activities, including research and evidence generation.

Given the growing complexity of data gathering, access, and pooling, many legal issues must be considered. Nicolas Terry, Chester A. Myers Professor of Law and codirector of the Center for Health Law Studies at Saint Louis University School of Law, provides an overview of legal rules and regulations that preclude effective data sharing and aggregation. Terry elaborates on concepts of property and inalienability rules, the disconnect between federal and state regulations, and the continued development of legal models protecting privacy of health data. One area discussed is the notion of a combined National Committee on Vital and Health Statistics (NCVHS) secondary stewardship model with that of the European data directives, which might guard against data misuse while addressing the growing need for access to patient clinical information by supporting strong obligations for data stewards.

CHARACTERISTICS OF A PUBLIC GOOD AND HOW THEY ARE APPLIED TO HEALTHCARE DATA

David Blumenthal, M.D., M.P.P.
Director, Institute for Health Policy at
Massachusetts General Hospital/Partners Health System
Physician, Massachusetts General Hospital

This paper will review the classic definition of a public good and discuss how that definition applies to healthcare information under varying circumstances. The paper will also examine rationales for making data publicly available even when they do not meet the classic definition of a public good. These rationales apply to situations in which government has supported—through financial or other means—the development of the data in question or in which making the data publicly available has major benefits (i.e., positive externalities) that are not captured in normal market transactions. The purpose is to provide a framework for evaluating the case for public release of varying types of data; one Roundtable theme is to ensure that publicly funded data are used for the public benefit.

Some say that in theory, there is no difference between theory and practice, but in practice there is. The observations that follow are made on the assumption that theory still matters. I will start with rather esoteric, highly abstract observations, then make the argument that theory actually provides practical guidance in the development of policy and action.

Theory is of practical importance because it strongly affects the political thinking of key actors, whether they are aware of this effect or not. Ideology is a powerful force in our highly partisan political environment. This ideology concerns first and foremost the role of markets in our polity, and assumptions about what markets do well and do not do well. Decisions about the management of health information will involve politics at many levels and they will, consequently, involve ideology. If information is to be treated in any respect as a public good, it will be necessary to keep in mind that nonmarket mechanisms have a role in its management and distribution.

What is a public good? Many of us think we know what a public good is, and the term is often used, but it is a term in which intuition is not a good guide. There is a common supposition that a public good is something that is so good that somebody—usually the government—should make sure that anyone who wants and needs it can get it. This intuitive definition of a public good, however, is a reflection of moral, not economic, reasoning.

The Economics 101 definition of a good is that it is a product or service. It has no intrinsic merit; its goodness is determined by its value in the marketplace. Diamonds and gold have goodness in that they have market value. Straw is not so good unless you believe, as in the fairy tale, that straw can be turned into gold; alchemists broke many wands over that effort.

But markets don't always work. When markets fail it means they have ceased to be efficient at determining or recognizing the value of a good or service. When that happens, socially optimal rates of production or patterns of distribution are not achieved.

Where and when do markets fail? One time where they fail is when there is a moral revolt against the operation of markets. Markets can also fail in the cases of pure public goods and quasi-public goods. An example of a market that has failed because of public abhorrence is the slave market. This market was common at one time, but of course it is no longer permitted in the United States. Many other markets are banned in the United States on moral grounds: contract killing, sale of human organs, selling votes, illicit substances such as heroin, and sale of cigarettes to minors. The fact that restrictions on gambling and alcohol have declined shows that moral restrictions can change over time. Another example is prostitution, which is not only allowed, but flourishes in a few parts of this country.

Another type of market failure is more relevant to the discussion at hand and defines the pure public good in economic terms. Pure public goods cannot be efficiently traded for two reasons: they are nonrival and they are nonexcludable. Defining a good as nonrival means that using the good does not preclude others' use of the good. In effect, the marginal cost of nonrival goods use is zero, and therefore an efficient market should price the marginal use of that good at zero. No money should be made in an efficient market from the sale of that product. In accordance with a Roundtable theme, we must correct the market failure for expanding electronic health records.

A good is nonexcludable if, even if wholly owned and paid for, its use and benefit by others cannot be prevented. Both Einstein's theory of mass-energy equivalence and the double-helix nature and structure of the DNA molecule are examples of public goods. Both are the products of fundamental research. The theory of special and general relativity and the structure of molecules of DNA are nonrival because, no matter how often the research results are used, their value remains and they are available for general use. Furthermore, it would have been virtually impossible for these findings to have had any value if they were not widely shared. Einstein did not just dream up his theory in isolation—he validated it by sharing it broadly within the community of physics to allow it to be critiqued. Similarly, Watson and Crick broadly shared their findings. In fact, replication of this kind of research result is critical to establishing its value. The scientific method in itself requires broad dissemination of results to confirm their validity, and once disseminated, their use can't be restricted. That is, they are fundamentally nonexcludable.

The public-good nature of basic research is something that people of every political persuasion in this country accept. Across the spectrum of

ideological opinions about markets, from those who cherish them to those who revile them, there is no question that basic research is nonrival and nonexcludable, and that its support is an appropriate and necessary role of government. That is why the National Institutes of Health (NIH) and the National Science Foundation (NSF) are not controversial public programs. Differences on degrees of support may vary in terms of the inherent value of their basic research, but few argue that markets could achieve what these agencies do. In fact, the government produces many examples of classic public goods without controversy. One is national defense. For example, an individual's use of an aircraft carrier for protection does not exclude general use of it, nor does it diminish the value to others.

There are also quasi-public goods. These are a little more relevant to our discussion of healthcare data. A quasi-public good is one whose production or consumption generates or might generate effects on third parties. It might be a case in which my contract with you has an effect on somebody else in this room who is not a party to that contract. It could be a positive or a negative effect. The consequences of the effect on the third party are not captured in the market transactions between the individuals who participate in the private purchase and sale.

There are also cases where goods may be nonrival, but not nonexcludable, or perhaps nonexcludable, but not nonrival. These are things that don't exactly fit the definition of a public good, but there is a public feel to them. There are also many examples of goods that are quasi-public because of their externalities—their effects on third parties not directly involved in the market transactions involving the goods. Energy production and, to an even greater extent, energy consumption are cases in point. Another is chemical production with the pollution of air and water. Clearly this has externalities.

Applied biomedical research has aspects of a public good as well as aspects of a quasi-public good. Knowledge concerning research related to a particular drug or device can be appropriated up to a point. It is excludable within limits and it is rival within limits. Clearly one can keep this kind of information secret and benefit from it in a marketplace, and many medical device companies make their living without patenting by keeping secret how their devices are produced. In some such cases making the information available would broadly benefit society, leading to the advancement of other knowledge. Keeping knowledge private causes a loss of efficiency, but we tolerate this loss for the gain that is created by the incentives for innovation resulting from the opportunity for economic gain.

The purpose of patent law is to mitigate the efficiency loss. The critical feature of patent law is that in order to get a patent, one must reveal the science and practice that led to the patent. To obtain exclusivity, the monopoly that is granted by the government requires making public the

information underlying the patent. That makes scientific progress based on protected information possible, while individuals enjoy economic fruits of innovation.

In the course of this workshop's discussions of large private clinical databases, some of the examples used are of groups that have made public the data they collect available to third parties essentially free of charge, usually out of altruism. That is wonderful, of course, but as a society we have not organized ourselves around altruism as a guarantee of any particular outcome. The real-world, large clinical databases have an aspect of a quasi-public good because they are not pure public goods in any sense. They are definitely excludable. Kaiser could exclude others from the use of their database, as could my own institution, the Partners Health System. These data create opportunities for private gain. The data create competitive advantage by enabling organizations to learn from their experience, perhaps to achieve better outcomes than their rivals; for example, they may learn how to treat a certain disease better. Organizations might be willing to sell clinical information, but are probably not sharing information for free. This is the equivalent of using trade secrecy for medical practice, and it is possible today.

As with applied research, there is a nonrival aspect to such information. In a local market, it is true that people compete with others on a new treatment for diabetes or coronary artery disease; however, someone in Singapore probably could not compete with me, and there might be enormous gains to sharing the knowledge with people in Singapore. When discussing large clinical databases, the fact is that the marginal cost of using the database is virtually nothing. The data runs and bytes of information are there whether used or not. They are not used up.

The principal questions for this discussion concern what to do with privately maintained databases that have private costs and value: databases, in other words, that given parties will neither construct nor share out of altruism, but for which large externalities exist. Therefore, in effect, a way must be found to realize these benefits without losing the incentive to put the database together.

Also relevant to this discussion is the existence of another type of less controversial informational public good. Data found at the NIH, for example, or developed through the National Health Interview Survey or National Census represent situations where the taxpayer has paid for the information to be collected. The data may not be very useful by the time they actually move into the public domain, but eventually they become available. Restrictions are placed on such data—for example, for national security purposes, not all defense-related data are public. Additionally, at certain times, making data available to the public is inefficient. The *Bayh-Dole Act* of 1980 was meant to remedy underuse of results of publicly funded research when there was no way to privatize the resulting intel-

lectual property. The legislation sought to encourage, by creating private incentives, the use of publicly developed information.

In general, there are two solutions to determine whether information is a quasi-public good or a public good. The first solution is to increase the appropriateness or excludability of information. We have used patents and copyrights to do that. The second is to have the government produce the good in question. NIH and NSF are examples of those. In closing, we need to acknowledge that this is not going to be simple. Joseph Stiglitz, who won the Nobel Prize in economics, recently wrote this about public goods: "The concept of intellectual property . . . is not just a technical matter. There are judgment calls and trade-offs, with different people . . . affected differently by alternative decisions. . . . In practice, decisions are made on a case-by-case basis."

CHARACTERISTICS OF THE MARKETPLACE FOR MEDICAL CARE DATA

William Crown, Ph.D.
President, i3 Innovus

This paper will consider the various sources of medical and prescription drug data that are available to support real-world safety surveillance and other types of evidence-based medicine. It will consider why these databases are initially constructed, the implications this has for their use as research tools, and their commercial applications. The paper will conclude with some thoughts about how it might be possible to construct a data asset that would represent the broad experience of patients from several national databases.

Although we often lament the inadequacies of research databases in the United States, we are data rich compared with most other nations. We have national probabilistic surveys (e.g., National Medical Expenditure Survey [NMES]/Medical Expenditure Panel Survey [MEPS]; Medicare Current Beneficiary Survey [MCBS]; Surveillance, Epidemiology, and End Results [SEER]; National Long Term Care Survey [NLTCS]; and National Health and Nutrition Examination Survey [NHANES]); hospital discharge data from payers (e.g., Healthcare Cost and Utilization Project [HCUP]/National Immunization Survey [NIS]); pharmaceutical claims (e.g., drug-switch data from IMS and Wolters-Kluwer); linked enrollment; medical and drug claims databases from commercial health plans and large self-insured employers; and combinations of such databases assembled and made available in the form of commercial databases by data aggregators such as Ingenix, Medstat, and Pharmetrics. In addition, there are a variety of government databases, including state Medicaid files (Medicaid Statistical Information System [MSIS] and State Medical Research Files [SMRF]) and the Medicare

5 percent sample available from the Centers for Medicare & Medicaid Services (CMS). Integrated medical claims, prescription drug data, and enrollment information are also assembled by the Department of Defense. Although not necessarily available to outside researchers, electronic medical record (EMR) databases exist for several health plans. Large physician practices also frequently have such data and, as a by-product of providing EMR software, several vendor organizations have built aggregated EMR databases. Finally, numerous patient registries follow patients longitudinally if they have a given condition or have been treated with a particular therapy. This is not an exhaustive list, of course, but it provides a sense of the breadth of data available in the United States (Box 4-1).

BOX 4-1
Sources of U.S. Medical Care Data

- National Probabilistic Surveys
  - NMES/MEPS
  - MCBS
  - SEER
  - NLTCS
  - NHANES
- Hospital Discharge Data, all payer
  - HCUP/NIS
- Pharmaceutical Claims
  - Drug Switches
    - IMS
    - Wolters-Kluwer
- Linked Medical and Drug Claims
  - Data aggregators, commercial
    - Ingenix
    - Medstat
    - Pharmetrics
  - Government
    - MSIS
    - Medicare 5% sample
    - VA
- Disease Registries

The national probabilistic surveys and, to a lesser extent, the disease registries are probably the only types of data in our list that are collected specifically for research purposes. We end up using the other types of data for research, but that was not their original purpose. Although voluminous for their service types, the large inpatient databases such as HCUP and the NIS, as well as the prescription drug databases from organizations such as IMS and Wolters-Kluwer, typically are not linked to other data types such as outpatient medical claims. As a result, they are not as useful for most questions regarding safety surveillance, comparative effectiveness, or evidence-based medicine.

My focus here is on the linkage of drug data with medical claims because those sources provide the most comprehensive view of drug treatment or treatment of a patient, whether it is with a procedure or with a pharmaceutical, and then capture all the other healthcare uses of those patients. This is also true with respect to medical records in certain settings (staff model health maintenance organizations), and not necessarily true in other settings (e.g., specialized oncology or cardiovascular clinics).

A variety of sources of linked medical and drug claims data are available. Large commercial health plans and employers often have such data, but for their own systems. There are companies that provide data aggregation services, pooling data from multiple sources, such as health plans and employers. These organizations reformat and standardize the data and feed back to the contributing sources information on healthcare use and benchmarking. Sometimes they also measure correspondence with practice guidelines. Data aggregators also provide similar services to government agencies. For example, for many years CMS has funded the development of state Medicaid databases (tape-to-tape, MSIS, and SMRF) that combine longitudinal deidentified data on beneficiary enrollment, medical claims, and drug claims. Similarly, CMS has historically built linked inpatient and outpatient claims datasets for the entire Medicare population and made a 5 percent sample of these files available to researchers.

With a few notable exceptions, the kinds of rich, longitudinal data that we have in the United States either do not exist or are not accessible to researchers in other parts of the world. Examples of what is available include the Nordic registries in Norway, Sweden, and Finland. Probably the best known research database outside of the United States is the United Kingdom's General Practitioner Research Database, an outpatient encounter database. In addition, the Saskatchewan data in Canada are similar to U.S. claims data.

The federal government, actuarial consulting firms, academic researchers, and pharma are among those who license commercial U.S. research databases. Claims aggregators spend literally hundreds of millions of dollars entering these data, pooling them, standardizing them, and trying to turn

them into something useful. It is easy to underestimate what a big job that is. Consider, for example, that there is no such thing as a hospitalization in a claims database—all you have is many events happening in the hospital. Every encounter generates a claim. You have to figure out what constitutes a hospital stay, when the person was admitted, and when he or she was discharged, then roll up all the events that take place between the dates of admission and discharge and, in essence, create a hospital stay. Researchers have to cull bad information (e.g., men who show up as having had hysterectomies). It takes a lot just to process the volume of claims information in such databases even when the format is similar. One might think that data from commercial health plans would be formatted in similar ways, but they actually have different formats and different record layouts. Data aggregation is further complicated when you start thinking about pooling across different data sources. For example, how do you effectively combine information from different health plans, where one plan has fee-for-service dollar amounts attached to each service provided and another plan is fully capitated and only the service encounters are recorded? Similar issues arise in the pooling of medical records data across multiple sources when these sources use different medical record systems. The complexity of pooling claims data with medical records data is even more complicated. In short, an enormous amount of effort and considerable financial investment are required to develop large databases that pool information across multiple sources.

After putting all this effort into creating these databases, it is not surprising that claims aggregators commonly create deidentified research databases that they then license to third parties. They license the data to the federal government and to actuarial consulting firms that use the data to develop benefit designs. They also license the data to academic researchers, generally at reduced rates. The largest market for these commercially licensed databases undoubtedly is pharma, which uses the data for a variety of purposes, including outcomes research, safety monitoring, tracking market trends, and many others.

In terms of content, the standard medical claims database includes enrollment information and pharmacy, physician, and facility claims (Figure 4-1).

These data elements are linked together via a unique member ID. This is extraordinarily rich information in one sense; in a health plan context, it captures every interaction with the healthcare sector reimbursed by the patient's insurance—every procedure and date of service, diagnosis, prescription drug filled, use of the emergency room or hospital, and so forth. By rolling all of this information together for each plan member, one can create episodes of care and can follow patients longitudinally in the data, collecting all the information about adverse events, total healthcare use, and more. Of course, all of this is subject to drawbacks given that these data

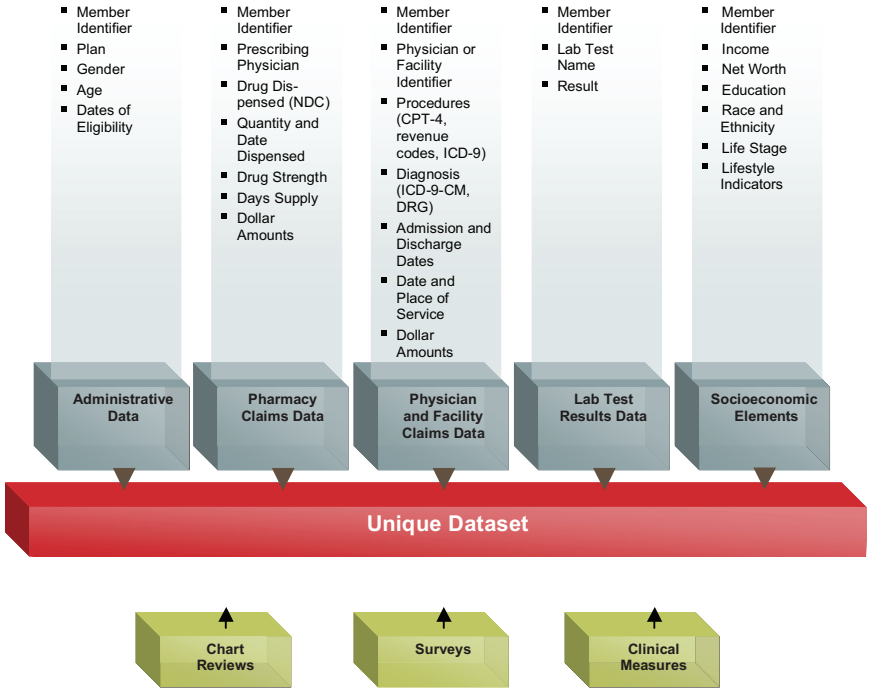


FIGURE 4-1 Data availability in a large health plan.

were never intended for research purposes, so there are inherent pitfalls, such as falsely inferring causality or erroneous coding. Depending on the database, there may also be information on laboratory test results or the social demographics of patients. The addition of sociodemographics such as income, net worth, education, self-reported race/ethnicity, and so forth represents an attempt to further enhance the data that are typically available just from the administrative claims and to try to find additional sources of variation and healthcare use that one wouldn't be able to observe from the health claims themselves.

The discussion thus far has focused on deidentified databases. However, in certain situations, such as in large health plans or physician practices, there is also the possibility to access protected health information (PHI), under Institutional Review Board (IRB) approval, to abstract medical charts and conduct surveys of the patients related to reasons for medication behaviors and health-related quality of life. Those data can then be combined with the administrative data already captured for these patients in the system. For example, when patients are found to be switching medications, it

is always possible to try to analyze this move with deidentified data, looking at factors such as changes in benefit design on subsequent drug switches. However, the real reasons for the switching behavior may not be observable in the claims data. A patient may discontinue an antihypertensive drug because he or she feels no direct physical evidence showing the drug is effective in lowering blood pressure. Alternatively, the patient may not like the idea of being dependent on a drug, may be concerned about side effects, or may they think the drug is not really working. In short, a variety of reasons for behaviors cannot be observed in the claims data and can only be gathered from the patient. Similarly, to fully understand behaviors of physicians, it is necessary to interview them.

Large retrospective claims databases can be particularly useful for safety signal detection. However, because of a variety of issues about the reliability of diagnostic coding in such databases, it is desirable to have access to medical records for the patients represented in the data. Again, this requires access to the PHI.

Aside from data quality issues themselves, there are challenges in drawing reliable inferences about cause and effect from observational data of any sort. Nonetheless, with current statistical methods, we can do a lot with observational data to control for confounders. Moreover, these data represent the real world, as opposed to the carefully controlled settings of clinical trials, which typically cover only small, carefully selected patient populations and therefore do not necessarily represent the patient populations that ultimately are going to use the drug. Also, follow-up periods are often short.

From the standpoint of forecasting what is going to happen in the real world, there are at least as many dangers in the use of clinical trials data to predict real-world outcomes as there is danger in inferring cause and effect from observational data. At least in the latter case, we observe real-world outcomes in actual patient populations using the drug. Both types of information have a role. Randomization enables reliable statistical inferences about cause and effect to be drawn for the patient population in the trial. Real-world observational studies allow us to see what transpires in actual clinical practice.

From the standpoint of payers, one issue from a health economics standpoint is establishing the value of pharmaceutical treatments and other interventions. This is particularly the case given that so many of the new products emerging from clinical development programs now are biologics, with very high price points, 10, 20, 30, or even 100 times the price of existing conventional pharmaceuticals. In many countries (especially northern Europe, Canada, and Australia), coverage and reimbursement of new therapies has been predicated on the demonstration of relative cost effectiveness. However, estimates of comparative cost-effectiveness have generally combined efficacy and safety data from clinical trials with real-

world cost data from other sources. The reality is that historically we have not had good data from clinical trials regarding effectiveness in real-world patient populations in order to figure out relative cost-effectiveness. This is going to drive a need for real-world data collection—going beyond retrospective data to collect clinical data that we typically do not have in these administrative databases to get the clinical endpoints about effectiveness. It is a significant challenge to collect information about patient-reported outcomes—in particular, health-related quality of life.

Where are we headed? In terms of the trade-offs between a pooled mega-database and pulling data from different data aggregators, the need is growing for a mega-database that would pull data from different health plans, the Department of Veterans Affairs, Medicare, and so forth. Then we would need to standardize the data, and create a public good that will be available for research, for cost-effectiveness studies, and for real-world drug safety to be able to understand guideline compliance of physician practices and other issues. There are some real challenges (Box 4-2).

BOX 4-2
Mega-Database or Distributed Network?

- Pooled Database
  - — Advantages
    - o Facilitates deduplication of patient data
    - o Statistical analysis of rare outcomes
  - — Disadvantages
    - o Concerns of data contributors make it unlikely that they will contribute their data to a pooled database
      - ■ Investment of data aggregators
      - ■ Competitive concerns of health plans
      - ■ Concerns about protecting patient confidentiality
    - o Very costly to build and maintain

- Distributive Network
  - — Advantages
    - o Effectively deals with concerns of data contributors
    - o Prior examples of successful collaborations of multiple health plans (e.g., HMO Research Network)
    - o Less expensive to build and maintain
  - — Disadvantages
    - o Deduplication of patient data

One of the biggest challenges is the need to deduplicate the data. For example, in order for the database to be valuable, it is necessary to recognize that person A—who originally starts out as an enrollee with United HealthCare and then 6 months later switches to Aetna—is the same person. That is a big challenge. Conceptually, the simplest way to accomplish deduplication is to pool the data. In terms of statistical analysis of rare outcomes, this would result in huge sample sizes. As a consequence, the ability to follow large cohorts of people longitudinally would provide researchers with the statistical power to detect extremely rare events that could not be detected, or certainly not in a statistically significant way, within smaller subsets of the data. There are some real advantages to that.

As discussed earlier, however, a significant challenge to constructing a pooled mega-database for the purposes of evidence-based medicine is the cost of building and maintaining such a database. Even the construction of a pooled database built from similar data streams (e.g., commercial health plans) is a huge task. Pooling data from sources with very different data structures (e.g., medical claims and EMRs) would be a monumental effort.

Institutions contributing information to a pooled database generally have a variety of concerns. For example, although health plans generally have an interest from a public health standpoint in contributing their data to a pooled database, they may have concerns about the potential to inadvertently provide their competitors with information on charges and payments for different types of services, benefit design, and more. Health plans are also very concerned about protecting patient confidentiality. Any time patient-level data (even though they are deidentified) are made available to third parties, the potential exists for reidentification of a patient either intentionally or by accident.

Finally, data aggregators who have invested tremendous amounts of money in creating their databases will be hesitant to turn them over to a pooled database unless they have a commercial incentive that enables them to recoup the value of the investments they have made in constructing these databases.

For all of these reasons, we need an alternative to the pooled mega-database model. One alternative is a distributive research network. Databases already exist in different health plans that use a common EMR provider. As a result, the data are fairly similar and could be combined. Similarly, large commercial health insurance databases already exist that link patient enrollment, medical claims, and drug claims. Rather than trying to pool everything in one massive file, these large subsets of data could be kept as separate nodes in a distributed network. It would be possible to conduct research within the databases represented by each of the nodes through a standard research protocol, and then to pool the results

afterward. Such an approach is potentially a way to address the possible concerns of data contributors. The chief disadvantage of this approach is the patient deduplication issue, which, of course, is a major strength of the pooled mega-database model.

One issue currently being considered by the Food and Drug Administration (FDA) Sentinel Network is exactly this issue—how to develop methods either with restricted databases or with IRB approval for the deduplication function to pool data across multiple sources. Although the FDA is interested primarily in drug safety, it seems clear that the Sentinel Network might also meet the objectives of a public good database for the purposes of evidence-based medicine and comparative effectiveness research. It is fair to say that broad interest is coalescing around the virtues of creating a public database to support evidence-based medicine and safety research. We will get there a lot faster if we recognize the practical challenges raised by alternative data models, as well as the issues and concerns of all the stakeholders.

LEGAL ISSUES RELATED TO DATA ACCESS, POOLING, AND USE

Nicolas P. Terry, LL.M.
Chester A. Myers Professor of Law,
Codirector, Center for Health Law Studies
Saint Louis University Law School

Introduction

The legal system enters the "public good" debate because it reflects and thus perpetuates the current "excludability" state of clinical data with property and intellectual property models. Furthermore, market exchanges or shifts to public good "nonexcludability" face legal barriers (e.g., privacy, confidentiality, and security) that are designed to reduce or eliminate negative externalities suffered by data subjects. This paper identifies the major clusters of legal rules that create barriers to clinical data morphing into a public good: property or inalienability rules, federal–state disconnects, and evolving data protection models. The paper concludes with some observations on approaches to resolving the current excludability rules and, somewhat counterintuitively, argues that a more rigorous data protection model will be required as a prerequisite for greater access to patient data.

Three clusters of legal issues potentially create barriers to the adoption of a public good model for clinical data. They are described herein as property rules, federal–state vectors, and data protection. However, to understand them requires some initial observations regarding the legal sys-

tem and its historically unsatisfactory interaction with health information technologies (HITs).

First, in this debate the legal system is neither a spectator nor an independent actor. Legal models enter the equation because they reflect and so perpetuate the intended or perceived current state of public policy. The frame for "public goods" analysis begins with the recognition that they exhibit the characteristics of "nonexcludability" and "nonrivalrous" consumption (Cowen, 2002). These characteristics may cause market failures in information properties because they encourage free riders and can create positive externalities (Cowen, 2002). As Yochai Benkler explains, we apply legal protection to information properties because of our willingness "to have some inefficient lack of access to information every day, in exchange for getting more people involved in information production over time" (Benkler, 2006). That is, we "trade off some static inefficiency to achieve dynamic efficiency" (Benkler, 2006). By enabling excludability regimes (via property or, more typically, intellectual property laws), we seek to promote a dynamic efficiency model. Today, the legal system considers itself as being under a mandate to create or support structures that treat clinical data as a *private* good (NRC, 2003c). It follows that any move away from clinical data excludability to public good status must deal not only with the technical legal barriers also with their economic underpinnings.

Similarly, exchanges between data stewards that facilitate nonrival consumption of information properties (including interinstitution sharing of records data for outcomes research and the sale of clinical data for marketing purposes) or novel public goods exceptions to nonexcludability regimes may impose negative externalities on the data subjects. Data protection laws are designed to eliminate or reduce these externalities.

Second, the legal system is rife with uncertainties. To the befuddlement of "real" scientists, lawyers seems to spend less time providing efficient "yes/no" answers, and far more billable time delivering annoyingly inefficient "maybe" responses (Solum, 1987).[1] Consider some of the real or perceived barriers to HIT and how many physicians have been discouraged from improving access to care by using telemedicine because of uncertainties about the impact of state licensure laws, the standard of care, or the application of malpractice insurance (Terry, 2004). This sense of uncertainty or indeterminacy is increased by the legion of "legacy laws," such as records laws predating electronic clinical data collection and the potential for data mining of records to improve outcomes and effectiveness.

Overall, it is tempting to recall a well-known phrase coined by an Australian judge discussing the interaction of law and medicine: "Law,

[1] There is a long tradition in the legal literature of examining this phenomenon at various levels of abstraction.

marching with medicine but in the rear and limping a little" (Windeyer, 1970). As discussed below, these tendencies toward indeterminacy are exacerbated by the relationship between federal and state regulatory and statutory models. Furthermore, interlocking problems of indeterminacy, outdatedness, and overlapping or contradictory legal regimes applied to HIT and health information exchange (HIE) reinforce the sense of unintended consequences. Such consequences range from the technical (e.g., regulatory safe harbors notwithstanding, stark and antikickback barriers to market transactions between providers to accelerate the adoption of e-prescribing and electronic health records, or EHRs[2]) to the conceptual (e.g., the *Health Insurance Portability and Accountability Act's*, or HIPAA's, compliance-based, provider-centric data protection model that tends to confirm the proprietary, or private goods, nature of clinical data by encouraging providers to wall off the data as "theirs" rather than treat it as held in trust for their patients or the public).

Property/Inalienability Rules

Ownership of Medical Records

State law continues to dominate the records space. Licensure laws create duties of accuracy, completeness, legibility, and timeliness (Nev. Rev. Stat., 2008; N.M. Stat., 1978; Wyo. Stat. Ann., 2005), while other state laws regulate the alteration of records (Nev. Rev. Stat., 2008) or their retention (La. Rev. Stat. Ann., 2001; N.M. Stat. Ann., 2008). Providers who breach these standards may face disciplinary sanctions (Nieves vs. Chassin, 1995; Schwarz vs. Bd. of Regents, 1982). State malpractice principles weigh in on some questions about the sufficiency of records (Thomas vs. United States, 1987), while the emerging tort of spoliation of evidence increasingly deters alteration or destruction of records (Pikey vs. Bryant, 2006; Rosenblit vs. Zimmerman, 2001). Increasingly, national accreditation standards such as the Joint Commission on Accreditation of Healthcare Organizations (JCAHO) rules[3] and federal Medicare standards (CMS, 2005) are entering the records space by mandating record retention rules.

This mixed legal basis carries over to the question of property rights in clinical data. It is generally accepted that doctors own the medical records they keep about patients (**Regensdorfer vs. Orange Regional Medi-**

[2] See http://www.hhs.gov/healthit/certification/stark/.

[3] See Joint Commission on Accreditation of Healthcare Organizations, IM.6.10: "The medical record contains sufficient information to identify the patient; support the diagnosis/condition; justify the care, treatment, and services; document the course and results of care, treatment, and services; and promote continuity of care among providers."

cal Center, 2005; Waldron vs. Ball Corp., 1994). Statutes in some states (Fla. Stat., 2009) and practices endorsed by the American Medical Association (AMA, 1983) confirm this position. State statutes have extended this model to hospitals and their ownership of records (Fla. Stat., 2009; Tenn. Code Ann., 2004).

Arguably, the position on ownership of records is slightly more complicated. Although patients may not own the actual paper records, they may have some ownership rights in the information contained in the records (although this position has been rejected by the High Court of Australia [Breen vs. Williams, 1996]). The federal HIPAA confidentiality rule sought to be agnostic on the issue, purporting to govern only use and disclosure of certain records. However, HIPAA grants quasi-property interests to patients in their records by recognizing rights of access (CMS, 2003a) and modification (CMS, 2003b).

Although it did not address the issue of property rights in records or other clinical data, the well-known case of Moore vs. Regents, University of California,[4] is a useful starting position from which to predict the likely judicial attitude to property-based arguments by patients (Moore vs. Regents, University of California, 1990). In Moore, the Supreme Court of California held that a leukemia patient did not have a property-based interest in his removed tissue from which the defendant established a potentially profitable cell line. The court reasoned that the extension of the conversion tort, by which such interests may be protected, would potentially "punish innocent parties" (downstream researchers) or "create disincentives to the conduct of socially beneficial research."[5] However, another reason for the conclusion flowed from the same court's recognition of causes of action for breach of fiduciary duty or lack of informed consent that could provide remedies against the physician. Moore, therefore, begs the question of whether a physician or hospital should obtain consent from a patient regarding the use of the patient's data beyond the point of care. Similarly, would a fiduciary duty be breached if a provider sought to monetize data extracted from a record?

One final property-related complication regarding EMR may require attention. Some EMR technology providers (the owners of the enabling software platform) may retain proprietary rights in that technology and so to an extent the records built on that platform (Harty-Golder, 2007). In most cases this issue should be resolved in advance by licensing agreements between the healthcare provider and the software vendor. However, as

[4] For subsequent cases dealing with the disputes between researchers and healthcare institutions over issues of property and other rights in cells and genetic materials, see Greenberg vs. Miami Children's Hosp. Research Inst., Inc., 264 F. Supp. 2d 1064 (S.D. Fla. 2003); Wash. Univ. vs. Catalona, 490 F.3d 667 (8th Cir. Mo. 2007).

[5] See 793 P.2d at 144.

interoperability (e.g., a standardized export format) increases, the importance of this issue should diminish.

IP and Trade Secret Protections

In most situations, and as is generally the case with information properties, the legal system will treat the content of records and derived clinical data as more abstract properties that potentially are protected by the law of intellectual property (IP). Clinical data claims for IP protection could arise under patent or copyright laws. In practice, however, related protections under the law of trade secrets may be more important.

In 2007 the Supreme Court took some initial steps away from limitless patentability by tightening the requirement of "obviousness" as applied to patent claims (KSR vs. Teleflex, 2007). Notwithstanding, there is authority that medical records software is patentable (Micro Chem., Inc. vs. Lextron, Inc., 2003), although it is doubtful whether that status extends to the actual record created using the software.

The law of copyright protects most written works. However, records data or derived clinical data may be viewed as merely factual, and copyright law does not protect facts, including medical and biographical facts (N.Y. Mercantile Exch., Inc. vs. Intercontinental Exchange, Inc., 2007). In limited situations, the question of copyright protection might be satisfied by the way the facts are arranged (Inc. vs. Chinatown Today Pub. Enterprises, Inc., 1991; Matthew Bender & Co. vs. West Publ. Co., 1998) or by their supplementation with creative work (CCC Information Servs. vs. MacLean Hunter Mkt. Reports, 1994; Harper & Row, Publishers, Inc. vs. Nation Enterprises, 1985). Furthermore, although facts themselves can never be the subject of copyright, their organization may be protected as a "compilation" (United States Code, 1992). In Feist Publications, Inc., v. Rural Telephone Service Co., a case dealing with lists of subscribers in a telephone directory, the Supreme Court limited copyright protection for nonoriginal, noncreative works (Feist Publications, Inc. vs. Rural Telephone Service Co., Inc., 1991). Notwithstanding, the creativity threshold is quite low. The Feist court famously held that "'Original,' as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity" (Feist Publications, Inc. vs. Rural Telephone Service Co., Inc., 1991). Thus, it has been stated, "a compilation of preexisting facts . . . can still meet the constitutional minimum for copyright protection if it features original selection, coordination or arrangement of those facts" (Victor Lalli Enterprises, Inc. vs. Big Red Apple, Inc., 1991).[6]

[6] See generally David E. Shipley, *Thin but Not Anorexic: Copyright Protection for Compilations and Other Fact Works,* 15 J. Intell. Prop. L. 91 (2007).

Traditional intellectual property systems protect inventors or authors with a term-limited monopoly when they market their expressions or inventions. It is as likely, however, that most "owners" of records or derived clinical data will seek to protect them by keeping them private or "secret." Indeed, the application of data protection rules, discussed below, essentially mandates "secrecy." Looked at this way, patient records and derived clinical data are more analogous to customer lists and other business records (Unistar Corp. vs. Child, 1982). Therefore, data with economic value that are kept secret and reasonably secure may be treated as trade secrets under state law.[7]

As in the case of customer lists, patient records are often tangentially protected through contract. Thus, when doctors become employees, shareholders, or members of a medical practice, they often enter a contractual agreement stating that records are owned by the practice and containing a restrictive covenant not to compete if they leave the practice. The terms of these contracts (particularly the covenant not to compete) are often not enforced by the courts when to do so would adversely affect the public interest in patient choice and physician mobility (Ohio Urology, Inc. vs. Poll, 1991; Valley Med. Specialists vs. Farber, 1999).

The stewards of clinical data (again, in part motivated by data protection laws) likely will protect their data with security and related rights management systems. Those who circumvent these protections may face actions brought under the *Computer Fraud and Abuse Act* (United States Code, 2009) or the *Digital Millennium Copyright Act* (United States Code, 1998).

An Escalating Federal-State Legal Vector

Traditionally the regulation of medical records has been a creature of state law. Given the national initiatives for HIPAA transactions, privacy and security, and HIE, it should follow that the legal environment would be a cohesive federal one. That model seems increasingly unlikely. To take just the issue of privacy, the federal HIPAA code is less than comprehensive, leaving unprotected large swathes of patient data. Yet, unlike its security code, HIPAA's Privacy Rule does not preempt more stringent state protections (a phenomenon leading to what is known as the HIPAA *floor*). Therefore, if anything, the post-HIPAA years have seen an increase in state legislation impacting HIT. Even in the case of HIE, an area once considered purely national policy, states have started to address incentives and disincentives as federal legislative activity and funding have slowed.

[7] Most states have adopted the *Uniform Trade Secret Act*, National Conference of Commissioners on Uniform State Laws (1985).

State Restrictions on Data Collection, Processing, or Security

Several states have clinical data protection legislation that is more protective of patient data than is HIPAA. The most likely area for additional protection is to apply data protection to a more expansive list of custodians than HIPAA's narrow "current entities" model. A few seek to go considerably further. For example, a recently defeated New Hampshire bill (House kills medical privacy bill, 2008) would have increased data protection considerably beyond HIPAA protection standards by restricting data use to the point of care, thereby potentially outlawing many marketing and research uses (Guay, 2008; U.S. House of Representatives, 2008).

Other state legislation operates on the periphery of HIPAA. For example, an Arizona bill would prohibit non-U.S. outsourcing of medical data processing absent patient consent,[8] while several states (Georgia, 2008; South Dakota, 2008) are set to join the ranks of those controlling data acquisition through subcutaneous Radio Frequency Identification (RFID) tags.[9]

Increasingly, states regulate the use of either individual or aggregated (even de-identified) clinical data. Thus, a majority of states have legislation prohibiting employment discrimination based on genetic information.[10] A minority of states go farther and prohibit the genetic testing of employees (NCSL, 2008c). Most states apply a similar model to applications for health insurance (NCSL, 2008a), and some states have extended that to disability and life insurance (NCSL, 2008b). As is the case with health privacy, much of this activity in the states is a function of Congress's apparent inability to pass comprehensive legislation dealing with the issues.[11]

Recent legislation in New Hampshire (**N.H. Rev. Stat. Ann., 2007**), Maine (Maine State Legislature, 2005), and Vermont (Vermont, 2007) has placed varying levels of restrictions on the secondary uses of pharmacy information (New Hampshire, YEAR; MRSA, 2007). In the aftermath, pharmacy data aggregators have successfully challenged such legislation in the federal courts for violation of protected commercial speech principles (IMS Health vs. Sorrell, 2007; IMS Health, Inc. vs. Ayotte, 2007; IMS

[8] See generally Nicolas P. Terry, *Under-Regulated Healthcare Phenomena in a Flat World: Medical Tourism and Outsourcing,* 29 W. N. Eng. L. Rev. 421-72 (2007) and H.B. 2401.

[9] See, e.g., Cal. Civ. Code § 52.7.

[10] For example, New Jersey law requires: "No person shall obtain genetic information from an individual, or from an individual's DNA sample, without first obtaining informed consent from the individual. . . . " N.J. Stat. § 10:5-45. However, the statute does not apply to "anonymous research where the identity of the subject will not be released" N.J. Stat. § 10:5-45 (a)(5).

[11] The *Genetic Information Nondiscrimination Act* has twice been approved by the House (in 2007 and 2008). See generally National Human Genome Research Institute, *Legislation on Genetic Discrimination,* http://www.genome.gov/10002077#2.

Health Corp. vs. Rowe, 2007). Notwithstanding this unfriendly reception, Maryland introduced H.B. 50 and Washington is considering similar legislation. An Arizona bill may go even farther in that it extends the prohibition to commercial uses of "records relating to prescription information that contain patient-identifiable and prescriber-identifiable data."[12]

Finally, breach notification statutes demonstrate one of the most rapid explosions of regulation in the privacy-confidentiality-security constellation. California passed the first such statute in 2003. By 2008, with 9 often-conflicting bills languishing in various congressional committees, 39 additional states had passed similar legislation requiring a data steward to inform data subjects when their data have been compromised. Most of these statutes apply only to financial identity theft. However, a growing number also seem to apply to cases of medical identity theft, only granting providers a safe harbor when they are subject to *and in compliance with* HIPAA.[13]

State Initiatives in HIT and HIE Policy

In February 2008 the Government Accountability Office (GAO) reported that there is still no national strategy for HIT (GAO, 2008). This echoed the previous month's conclusion by the California Healthcare Foundation that "The President's HIT adoption agenda has raised consciousness about HIT and EHRs. Beyond the laying of a conceptual foundation, however, there is as yet no measurable increase in HIT or EHR adoption." Dealing with the specifics of the Administration's proposed National Health Information Network (NHIN), the Foundation concluded, "[t]hough it represents a worthy goal, the NHIN is impractical and cannot be implemented" (Fried, 2008).

As they observe the failure of high-profile regional health information organization (RHIO) projects (Miller and Miller, 2007), state actors may now perceive that the Office of the National Coordinator for Health Information Technology (ONCHIT) lacks a coherent, sustainable strategy or, at least, that its flat budgets confirm there will be no centralized funding that goes beyond demonstration projects. Into this real or perceived vacuum, some states are floating their own "carrots" and "sticks" designed to provide new impetus toward HIE projects. Recent state initiatives (few of which have met with legislative approval) have included the funding of a pilot program for clinical data sharing (West Virginia, 2008a), mandating the use of EMRs (Indiana, 2008), prohibiting providers from buying EMRs

[12] See S.B. 1251.

[13] See, e.g., Haw. Rev. Stat. § 487N-2(g)(2); Ore. 2007 S.B. 583, Chapter 759; Mich. Comp. L.S. § 445.72.

that lack interoperability (New Jersey, 2006), and granting providers a state tax credit to offset investments in EMRs (West Virginia, 2008b).

Overall, the Agency for Healthcare Research and Quality's (AHRQ's) 2007 Health Information Security and Privacy Collaboration Privacy and Security Project report (AHRQ, 2007) noted more than 300 current state initiatives relating to HIT and HIE. As these initiatives continue and state data-related laws run faster and deeper than federal protections do, unraveling them and finding agreement on a federal model for data protection and sharing will become immeasurably more difficult. At the same time, such local activity steadily increases legal indeterminacies and risks for interstate data stewards and processors.

Data Protection

The need to protect the privacy of health information is broadly accepted, yet the mechanisms for its assurance continue to be controversial. Pre-HIT protection for medical records was formally achieved with a patchwork of state statutory and common-law rules. In practice, however, the fragmentation of paper records across innumerable data silos provided the greatest protection. Electronic aggregation of records, powerful data-mining technologies, and a growing market for secondary uses have exponentially increased the negative externalities faced by data subjects (Terry, 2008). As is well known, in 2000 the federal government sought to address the costs to data subjects imposed by the HIPAA transactional standards with apparently comprehensive privacy and security regulations (HHS, 2000). In the years that followed, innumerable critical questions have been raised about the HIPAA codes, primarily regarding the costs (to data custodians), reach, and enforcement of the privacy regulations. These questions have multiplied[14] since President George W. Bush announced the federal EHR initiative (House, 2004).

With regard to data protection, the clinical data as a public good question requires a threshold issue to be addressed. HIPAA only applies to *identified* clinical data (HHS, 2002a). If a public good use of clinical data to improve outcomes and effectiveness research only contemplates the use of *deidentified* data, is there any substantial implication of data protection?

The short answer is that legal indeterminacies surrounding deidentification, variations in institutional data policies and practices, and uncertainties about the contemplated "public" secondary uses place data protection front and center as a potential barrier to outcomes research. The NCVHS has noted that some data custodians erroneously believe they have satisfied the

[14] See, e.g., Nicolas P. Terry & Leslie P. Francis, *Ensuring the Privacy and Confidentiality of Electronic Health Records*, 2007 U. Ill. L. Rev. 681-735.

deidentification safe harbor with anonymity or pseudoanonymity. Equally, some highly ethical (or risk-averse) providers use anonymity or pseudoanonymity when they could legally use identified data.[15] Furthermore, business associate agreements frequently lack clarity on the requirements for adequate deidentification, while complications arise from HIPAA's "Limited Dataset" Safe Harbor for research, public health, or healthcare operations and its required data use agreement (NCVHS, 2007). However, perhaps the most difficult challenge regarding reliance on deidentification is the increased use of data in records (e.g., genetic information and geocoding data) that exposes apparently deidentified datasets to reidentification.

The HIPAA Privacy Model

HIPAA's data protection regulation is technically complex and obstinately opaque. Its basic concept, however, is quite simple—arguably too simple. The model imagines a provider-controlled zone where individually identifiable patient data can flow quite freely (referred to as the "green" zone). In HIPAA-speak this green zone is referred to as "treatment, payment, or health care operations" (HHS, 2002a). There are two regulatory "walls" between the green zone and the "red" zone in which patient data generally should not circulate. First, the HIPAA Privacy Rule includes a general rule that green zone data custodians may not disclose data into the red zone. This is a confidentiality rule, albeit one that HIPAA mislabels as one of privacy (Terry and Francis, 2007). Second, the HIPAA Security Rule imposes technical and process obligations on data stewards to build a security wall that impedes those in the red zone (e.g., hackers) from accessing data stored in the green zone. Using regulated contracts, health insurers and providers may extend the green zone to include their "business associates" (e.g., law firms) (HHS, 2002b).

This HIPAA data protection model has several important (and unsatisfactory) properties. First, as follows from the description above, the protective model is almost exclusively disclosure-centric. That is, HIPAA does not limit or regulate *collection* of data. Thus, there are no controls over what or how much data can be collected, for example, by reference to a proportionality rule. Similarly, there is no requirement that patients must opt in or may opt out regarding the collection of certain types of data (e.g., psychiatric or gynecological records). Furthermore, HIPAA does not place any restrictions on secondary uses of data, other than simple patient consent that is mainly oblivious to the informational and bargaining asymmetries between the parties (HHS, 2002c). Additionally, for example, there is no prohibition

[15] Data do not have to be deidentified when used in healthcare operations (e.g., intraentity quality improvement research), 45 C.F.R. 164.506.

(consent notwithstanding) on the sale of clinical data for commercial purposes (that would be described as an inalienability rule). Finally, the HIPAA model shows little respect for its own red–green zone boundaries because it features broad exceptions (e.g., public health, judicial, and regulatory) to its protective model that do not require patient consent to data processing and that are susceptible to "function creep" (HHS, 2002e).

Personal Health Records and Consumer-Directed Health Care

Perhaps the greatest flaw of the HIPAA data protection model is how quickly it has been rendered wanting by new technologies. Designed to reduce the negative externalities imposed on data subjects in HIPAA transactions, it became obvious that the model was flawed in its applicability to emerging interoperable health record systems (Terry and Francis, 2007). Now, the emergence of personal health record (PHR) models raise far more serious issues. In contrast to the more familiar charts, paper records, and electronic medical records maintained by healthcare providers, PHRs are medical records created and maintained by patients. PHRs are provided by the patient's health insurer, healthcare provider, or employer, or even on an independent, commercial site potentially supported by advertising (collectively, PHR providers). However, HIPAA only applies to a relatively narrow range of healthcare entities (HHS, 2002d) that engage in certain types of transactions (HHS, 2002f). It is highly unlikely that most PHR providers will be directly subject to HIPAA data protection rules, although some state privacy statutes may apply and the Federal Trade Commission could exert some general control over PHR providers that promulgate their own privacy policies.

Currently, only 2 percent of the population uses PHRs (California HealthCare Foundation, 2008). However, robust growth is likely if the national EHR initiative slows and as major technology companies such as Microsoft, Google, and Inuit enter the PHR space (Lohr, 2007).

Working Toward Solutions

Information Property

Arrayed against the IOM public goods goal, the current aggregation of IP and related laws and technologies maintains excludability and so denies public use; what, in a broader context, James Boyle has referred to as "the second enclosure movement" (Boyle, 2003). There are certainly signs that the clinical data enclosure movement already has momentum. Many forthcoming health-quality initiatives, such as pay-for-performance (P4P) or consumer-directed health care (Jost, 2007), that seek to resolve

outstanding market failures in health care are heavily data driven; those who control the data are likely to have disproportionate control over the metrics. The AMA seems to have taken the position that their members should seek to monetize records data (O'Reilly, 2005). Indeed, a member of the AMA Board of Trustees noted, "there is tremendous economic value to the cumulative data in terms of analyzing patterns," and suggested that control of such data is central to doctors having influence on emerging P4P programs (NRC, 2003b).

Of course, property and IP debates are not new to the scientific community, as evidenced by the worldwide literature on gene patenting and attempts to balance research incentives and public goods arguments (Caulfield et al., 2006; ORNL, 2008). One theme that has emerged in the legal literature is to inquire whether IP rights holders should owe concomitant "public" duties. For example, Patricia Roche and George Annas have called for a comprehensive genetic privacy law that goes beyond the current model seen in state antidiscrimination laws (Roche and Annas, 2001, 2006). Jacqueline Lipton has argued more broadly that proprietary rights in "information property," while necessary to provide incentives and protect private property, must be balanced by broad new duties placed on rights holders, such as obligations of accuracy, confidentiality, and "an obligation to facilitate scientific, technical, and educational uses of information."[16] Agreement on how to operationalize such an approach has been elusive. Presumably, the IOM or NIH could explore with data rights holders the possibility of publishing clinical data under a creative commons license that permits noncommercial research (NRC, 2003a). Similarly, some form of compulsory licensing model may be possible (NRC, 2003a), although this approach has gained no traction in the gene-patenting arena and policy makers have tended to take the opposite approach in related areas, such as stimulating public benefits by *increasing* proprietary rights under the *Orphan Drug Act*.[17]

As actors seek to find a place for "public goods" considerations at the clinical data table, they must learn from broader (and not wholly successful) experiences recalibrating private and public interests in intellectual property. Of course "[c]opyright protection . . . is not available for any work of the United States Government,"[18] but that does not apply to most federally funded research. Furthermore, the *Bayh-Dole Act* changed the rules of the game for patent rights flowing from government-funded research.[19] More recently, the GAO, in examining contracting issues with

[16] Jacqueline Lipton, *Information Property: Rights And Responsibilities,* 56 FLA. L. REV. 135, 172 (2004).

[17] P.L. 97-414, as amended.

[18] 17 U.S.C.S. § 105.

[19] *University and Small Business Patent Procedures Act,* 35 U.S.C.S. § 200-212.

funded researchers, has discussed the tensions inherent in the monetizing of publicly funded research (GAO, 2008). Perhaps predictive of a future public goods regime for clinical data, the NIH has addressed the nature of the results of funded research with a policy addressing data sharing (NIH, 2003) and has required the public availability of manuscripts prepared by funded investigators (NIH, 2008).

Data Protection Models

The HIPAA privacy and health records debates have been marked by a serious disconnect between data custodians and government policy makers on one side and privacy advocates on the other. In the context of the federal NHIN project, the Bush Administration has narrowly framed the privacy–confidentiality issue, merely identifying divergent state laws as impeding implementation. This has been translated into a mandate to replace the HIPAA "floor," whereby more stringent state privacy protections are not preempted, with existing or reduced HIPAA protections as the new "ceiling" (Terry and Francis, 2007). The issue has even been raised in some tense exchanges between the GAO and ONCHIT. In its June 2007 report on HIT and privacy, the GAO recommended that "The Secretary . . . define and implement an overall approach for protecting health information" (GAO, 2007). The Department of Health and Human Services (HHS) responded that it already had a "comprehensive and integrated approach for ensuring the privacy and security of health information. . ." (GAO, 2007). When the GAO pushed back, ONCHIT agreed that an "overall approach" was required and instituted further study.[20] Yet, in its February 2008 report on nationwide HIT implementation by the HHS, the GAO noted, "Our recommendation for protecting health information has not yet been implemented" (GAO, 2008).

It may be seen as counterintuitive and as propagating still more legal barriers to public goods access to clinical data, but a stronger data protection model for medical privacy is a necessary predicate for greater sharing of patient data. Patients who lack trust in how data stewards or researchers treat their records will hide information from their doctors. Many of those doctors will perceive HIE as inconsistent with their professional standards of confidentiality or as creating liability "traps," and either refuse to participate or, if given no choice, reduce or distort their charting (Terry and Francis, 2007). In the words of the NCVHS, "*Erosion of trust* in the healthcare system may occur when there is divergence between what individuals reasonably expect health data to be used for and when uses are made for other purposes without their knowledge and permission" (NCVHS, 2007).

[20] Id. at 4.

A system that tolerates a lack of trust and the exposure of patients to discrimination, embarrassment, or stigma (NCVHS, 2007) will face reactions that compromise individual care and distort the data required for outcomes research.

The United States is not alone in confronting this tension between data protection and public utility. For example, a recent report by the New Zealand Law Commission noted, "there remains an outstanding issue as to whether there is a strong enough public mandate for the use of personal health information without consent for research in the public good." Similarly, a 2007 Canadian study noted very high (89 percent) public support for health quality research, yet only 11 percent of respondents felt no need for notification or consent regarding such research. Only 4 percent of respondents would deny all use, 32 percent would require consent for each use, 29 percent would be satisfied with a broad notification model, and 24 percent wanted notification and opt-out processes (Willison et al., 2007; Woolley and Propst, 2005).

If trust and transparency are the ideals in structuring a data protection model for medical data, what should follow? These ideals reflect patient expectations on how their data are processed and information they require before permitting unexpected uses. Current patient expectations likely are limited to point-of-care and continuum-of-care uses. There is probably also an increasing expectation of data access and use for personal health management (e.g., interoperability of EMR and PHR systems and any necessary processing).

Patients' expectations are unlikely to include the array of possible secondary uses for their clinical data, including the generation of outcomes and effectiveness research. As the NCVHS has suggested, "secondary use" is "an ill-defined term" that should be abandoned "in favor of precise description for each use of health data" (NCVHS, 2007). Indeed, the *Stewardship Framework* report identified a variety of such health-related (in contrast to law enforcement or regulatory) uses, including (1) payment; (2) healthcare operations (including internal quality assessment); (3) quality measurement, reporting, and improvement; (4) clinical research; (5) public health research; and (6) sale or barter of the data for commercial uses, including marketing (NCVHS, 2007).

A simplistic data protection model would simply outlaw some or all of these uses, but thereby deny patients the benefits of appropriate uses and resulting research. A more robust yet sophisticated protective model must be able to distinguish between these uses and adjust the responsibilities of data stewards and processors accordingly.

Although not sufficiently granular for our current purposes, the European Union data directive suggests a data protection model that imposes far more powerful obligations on data stewards and "chain of trust" data

processors (e.g., HIPAA's business associates). It is a model that rotates around a proportionality rule ("adequate, relevant and not excessive"), that applies to both *collection* and *disclosure* of data, and that limits the reprocessing of data for purposes incompatible with the original purpose of collection.

Applying a focused version of this model to clinical data, patient expectations would be met by relatively unimpeded use of data for point-of-care and continuum-of-care purposes (a far more restricted green zone). Trust will be further earned by permitting patient opt-out or data sequestering. Patient acceptance of *some* secondary uses will be more likely secured with strict limitations on commercial uses (a larger red zone). Between these extreme groupings, patient trust about research must be earned through transparency. The more patients learn about the uses projected for their data, are informed of its level of deidentification, and understand the feedback loop whereby outcomes research will improve their own care, the more likely they are to support a broader public agenda. The guarantee of transparency to the patient when data are applied for research is another theme of the Roundtable.

The difficult question, however, is how to implement this more robust data protection model. In its *Stewardship Framework* report (NCVHS, 2007), the NCVHS did an admirable job in suggesting tweaks to the HIPAA model by calling for stronger guidance, strengthening of business agreements and their parties' expectations, and calling on the Federal Trade Commission to increase its footprint in areas not regulated by HIPAA (such as PHRs). It is an approach that is perhaps attuned to the current political and legislative realities that apply to the data protection debate. However, it is hard to see how a patchwork of additional protections, particularly when based on a model as flawed as HIPAA, can deliver the robust model that is a predicate to patient-supported outcomes research. The NCVHS also recommended that "HHS should work with other federal agencies and the Congress . . . *for more inclusive, federal privacy legislation* so that all individuals and organizations that use and disclose individually identifiable health information are covered by the data stewardship principles inherent in such legislation, including a range of organizations not currently covered by HIPAA" (NCVHS, 2007). This is a difficult and likely unpopular agenda. However, it is difficult to see any alternative if there is to be a long-term accommodation of patient and researcher interests.

In conclusion, in the context of data protection and ownership, and patient expectations, it follows that there are two broad sets of legal barriers to a public goods future for clinical data. The first set of issues is somewhat process oriented. Thus, policy makers and legislators dealing with HIT and HIE issues must be better informed of the technologies and future technologies they seek to regulate to better reduce indeterminacies

and unintended consequences. Equally, few HIT or HIE issues are particularly local or are amenable to local or state legislative solutions. The federal legislative logjam on matters such as genetic discrimination, HIT funding, and effective data protection must be cleared to reduce the barriers posed by an escalating number of state "solutions."

However, the larger and more substantive barriers are as much a function of underlying policies as their legal transcription. Information properties in data and an inability to agree on an effective data protection model create immensely difficult barriers. These barriers will not be reduced with better legislative or regulatory drafting. Instead, they require a sober appreciation by all stakeholders in the clinical data space that they must support fundamental reforms.

REFERENCES

AHRQ (Agency for Healthcare Research and Quality). 2007. *Privacy and security solutions for interoperable health information exchange: Impact analysis*. http://healthit.ahrq.gov/portal/server.pt?open=514&objID=5554&mode=2&holderDisplayURL=http://prodportallb.ahrq.gov:7087/publishedcontent/publish/communities/k_o/knowledge_library/features_archive/features/impact_analysis_of_the_privacy_and_security_solutions_for_interoperable_health_information_exchange_project.html (accessed August 20, 2008).

AMA (American Medical Association). 1983. *Sale of a Medical Practice*. Medical Code of Ethics, E-704. http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion704.shtml (accessed February 3, 2010).

Benkler, Y. 2006. *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT, and London, UK: Yale University Press.

Boyle, J. 2003. The second enclosure movement and the construction of the public domain. *Law & Contemporary Problems* 33.

Breen vs. Williams. 1996. 186 CLR 71.

California HealthCare Foundation. 2008. *The state of health information technology in California: Consumer perspective*. http://www.chcf.org/documents/chronicdisease/HITConsumerSnapshot08.pdf (accessed August 20, 2008).

Caulfield, T., R. M. Cook-Deegan, F. S. Kieff, and J. P. Walsh. 2006. Evidence and anecdotes: An analysis of human gene patenting controversies. *Nature Biotechnology* 24(9):1091–1094.

CCC Info. Servs. vs. MacLean Hunter Mkt. Reports. 1994. 44 F.3d 61 (2nd Cir. Conn.).

CMS. 2003a. *Access of individuals to protected health information*. 45 C.F.R. § 164.524. http://edocket.access.gpo.gov/cfr_2003/octqtr/pdf/45cfr164.524.pdf (accessed February 4, 2010).

———. 2003b. *Amendment of protected health information*. 45 C.F.R. § 164.526. http://edocket.access.gpo.gov/cfr_2003/octqtr/pdf/45cfr164.526.pdf (accessed February 4, 2010).

———. 2005. 42 C.F.R. § 482.24(b)(1). http://edocket.access.gpo.gov/cfr_2007/octqtr/pdf/42cfr482.24.pdf (accessed February 3, 2010).

Cowen, T. 2002. *The concise encyclopedia of economics: Public goods and externalities*. http://www.econlib.org/library/Enc/PublicGoodsandExternalities.html (accessed August 21, 2008).

Desktop Marketing Systems Pty Ltd vs. Telstra Corporation Limited. 2002. FCAFC 112.

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. 1995. *Official Journal* L 281:31–50.

Feist Publications, Inc. vs. Rural Telephone Service Co., Inc. 1991. 499 U.S. 340, 111 S.Ct. 1282.

Fla. Stat. 2009. Ownership and control of patient records; report or copies of records to be furnished. § 456.057. http://www.leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=Ch0456/Sec057.HTM (accessed February 3, 2010).

Fried, B. M. 2008. *Gauging the progress of the National Health Information Technology Initiative: Perspectives from the field*. http://www.chcf.org/documents/chronicdisease/GaugingTheProgressOfTheNationalHITInitiative.pdf (accessed August 20, 2008).

GAO (Government Accountability Office). 2007. *Testimony before the Subcommittee on Information Policy, Census, and National Archives Committee on Oversight and Government Reform, U.S. House of Representatives. Health information technology: Efforts continue but comprehensive privacy approach needed for national strategy, GAO-07-988t*. http://www.gao.gov/new.items/d07988t.pdf (accessed August 20, 2008).

———. 2008. *Testimony before the Senate Committee on the Budget: Health information technology—HHS is pursuing efforts to advance nationwide implementation, but has not yet completed a national strategy.* GAO-08-499t.

Georgia. 2008. *The Microchip Consent Act of 2008*. H.B. 940. http://www.legis.state.ga.us/legis/2007_08/pdf/hb940.pdf (accessed February 24, 2010).

Guay, V. 2008. Health record privacy rules may get tighter, *Foster's Daily Democrat*. http://fosters.com/apps/pbcs.dll/article?AID=/20080210/GJNEWS01/67136800 (accessed February 23, 2010).

Harper & Row, Publishers, Inc. vs. Nation Enterprises. 1985. 471 U.S. 539.

Harty-Golder, B. 2007. EMR Ownership Questions. *Medical Laboratory Observer*. July 2007: 54. http://www.mlo-online.com/articles/0707/0707liab&lab.pdf (accessed February 23, 2010).

HHS (Department of Health and Human Services). 2000. *Security & Privacy*. 45 C.F.R. Part 164. http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=d63854405945c9ab56b93612fcc5e089&rgn=div5&view=text&node=45:1.0.1.3.71&idno=45 (accessed February 24, 2010).

———. 2002a. *Other requirements related to uses and disclosures of protected health information*. 45 C.F.R. § 164.514. http://edocket.access.gpo.gov/cfr_2002/octqtr/pdf/45cfr164.514.pdf (accessed February 24, 2010).

———. 2002b. *Uses and Disclosures: Organizational* Requirements.45 C.F.R. § 164.504. http://edocket.access.gpo.gov/cfr_2002/octqtr/pdf/45cfr164.504.pdf (accessed February 24, 2010).

———. 2002c. *Uses and Disclosures for which an authorization is required*. 45 C.F.R. § 164.508. http://edocket.access.gpo.gov/cfr_2002/octqtr/pdf/45cfr164.508.pdf (accessed February 24, 2010).

———. 2002d. *Definitions*. 45 C.F.R. § 160.103. http://edocket.access.gpo.gov/cfr_2002/octqtr/pdf/45cfr160.103.pdf (accessed February 24, 2010).

———. 2002e. *Uses and disclosures for which consent, or authorization, or opportunity to agree or object is not required*. 45 C.F.R. § 164.512. http://edocket.access.gpo.gov/cfr_2002/octqtr/pdf/45cfr164.512.pdf (accessed February 24, 2010).

———. 2002f. *Applicability*. 45 C.F.R. § 164.104(a). http://edocket.access.gpo.gov/cfr_2002/octqtr/pdf/45cfr164.104.pdf (accessed February 24, 2010).

House, T. W. 2004. *Transforming health care: The President's health information technology plan*. http://www.whitehouse.gov/infocus/technology/economic_policy200404/chap3.html (accessed August 20, 2004).

House kills medical privacy bill. 2008. *New Hampshire Business Review*.
IMS Health Corp. vs. Rowe. 2007. 532 F.Supp.2d 153 (D. Me.). Suit has also been filed regarding the Vermont legislation.
IMS Health, Inc. vs. Ayotte. 2007. 490 F.Supp.2d 163 (D. N.H. 2007).
IMS Health, Inc. v. Sorrell, No. 1:07-CV-188 (D. Vt. Apr. 23, 2009).
Inc. vs. Chinatown Today Pub. **Enterprises, Inc., 1991. 945 F.2d 509 (2nd Cir. N.Y.)**
Indiana. 2008. H.B. 1342.
Jost, T. S. 2007. *Health care at risk: A critique of the consumer-driven movement*. Durham, NC: Duke University Press.
KSR vs. Teleflex, 2007. 127 S. Ct. 1727.
La. Rev. Stat. Ann. 2001. Legislation around retention of medical records. § 40:2144(F) (Supp. 2006).
Lohr, S. 2007. Google and Microsoft look to change health care. The *New York Times*.
Maine State Legislature. 2005. *Confidentiality of prescription drug information*. 22 M.R.S.A. § 1711-E. http://www.mainelegislature.org/legis/bills/bills_123rd/billpdfs/HP000502.pdf (accessed February 24, 2010).
Matthew Bender & Co. vs. West Publ. Co. 1998. 158 F.3d 693 (2nd Cir. N.Y.).
Micro Chem., Inc. vs. Lextron, Inc. 2003. 317 F.3d 1387 (Fed. Cir.).
Miller, R. H., and B. S. Miller. 2007. The Santa Barbara county care data exchange: What happened? *Health Affairs (Millwood)* 26(5):w568–w580.
Moore vs. Regents University of California. 1990. 51 Cal. 3d 120.
MRSA 1711. An act to amend the prescription privacy law. State of Maine Legislature. June 18, 2007. http://janus.state.me.us/ legis/LawMakerWeb/summary.asp?ID=280022219. (accessed May 18, 2009).
NCSL (National Conference of State Legislatures). 2008a. *Genetics and health insurance state anti-discrimination laws*. http://www.ncsl.org/programs/health/genetics/ndishlth.htm (accessed August 20, 2008).
———. 2008b. *Genetics and life, disability and long-term care insurance*. http://www.ncsl.org/programs/health/genetics/ndislife.htm (accessed August 20, 2008).
———. 2008c. *State genetics employment laws*. http://www.ncsl.org/programs/health/genetics/ndiscrim.htm (accessed August 20, 2008).
NCVHS (National Committee on Vital and Health Statistics). 2007. *Enhanced protections for uses of health data: A stewardship framework for "secondary uses" of electronically collected and transmitted health data*. http://www.ncvhs.hhs.gov/071221lt.pdf (accessed August 20, 2008).
Nev. Rev. Stat. 2008. *Failure to maintain proper medical records; altering medical records; making false report; failure to file or obstructing required report; failure to allow inspection and copying of medical records; failure to report other person in violation of chapter or regulations*. §630.3062. http://leg.state.nv.us/NRS/NRS-630.html#NRS630Sec3062 (accessed February 3, 2010).
New Jersey. 2006. A.B. 1390.
New York Mercantile Exchange, Inc. v. IntercontinentalExchange, Inc. 05-5585-CV, 2007 WL 2189129, C.A.2 (N.Y.), 8/1/2007.
N.H. Rev. Stat. Ann. 2007. *The Perscription Restraint Law.* § 318:47-f.
Nieves vs. Chassin. 1995. 625 N.Y.S.2d 344 (N.Y. App. Div. 3d).
NIH (National Institutes of Health). 2003. *Final NIH statement on sharing research data, NOT-OD-03-032*. http://grants.nih.gov/grants/policy/data_sharing/ (accessed August 20, 2008).
———. 2008. *Revised policy on enhancing public access to archived publications resulting from NIH-funded research, NOT-OD-08-033*. http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-033.html (accessed August 20, 2008).

N.M. Stat. 1978. *Medical Practice Act*. § 61-6-15 D. http://www.nmmb.state.nm.us/pdffiles/MedicalPracticeAct.pdf (accessed February 3, 2010).

N.M. Stat. Ann. 2008. Legislation around retention of medical records. § 14-6-2.

NRC (National Research Council). 2003a. Responses by the research and education communities in preserving the public domain and promoting open access. In *The role of scientific and technical data and information in the public domain: New legal approaches in the private sector*. Washington, DC: The National Academies Press.

———. 2003b. *The role of scientific and technical data and information in the public domain: New legal approaches in the private sector*. Washington, DC: The National Academies Press.

———. 2003c. Scientific knowledge as a global public good: Contributions to innovation and the economy. In *The role of scientific and technical data and information in the public domain: Proceedings of a symposium*. Washington, DC: The National Academies Press.

Ohio Urology, Inc. vs. Poll. 1991. 72 Ohio App. 3d 446, 594 N.E.2d 1027 (Ohio Ct. App., Franklin County).

O'Reilly, K. B. 2005. AMA to set guidelines on control of record data. AMNews, American Medical Association.

ORNL (Oak Ridge National Laboratory). 2008. *Human genome project information: Genetics and patenting*. http://www.ornl.gov/sci/techresources/Human_Genome/elsi/patents.shtml (accessed August 20, 2008).

Pikey vs. Bryant. 2006. 203 S.W.3d 817 (Mo. Ct. App).

Regensdorfer vs. Orange Regional Medical Center. 2005. 799 NYS 2d 571.

Roche PA, Annas GJ. 2001. Protecting genetic privacy. *National Review of Genetics* 2:392–396.

Roche PA, Annas GJ. 2006. *DNA Testing, Banking, and Genetic Privacy*, 2:392-396; N Engl J Med 355; Aug. 10, 2006.

Rosenblit vs. Zimmerman. 2001. 766 A.2d 749 (N.J.).

Schwarz vs. Bd. of Regents. 1982. 453 N.Y.S.2d 836, 836–37 (N.Y. App. Div. 3d).

Solum, L. B. 1987. On the indeterminacy crisis: Critiquing critical dogma, U. Chi. L. Rev. 54: 462.

South Dakota. 2008. *An Act to Prohibit the Implanting of a Microchip and Provide a Civil Recovery.* H.B. 1141.

Tenn. Code Ann. 2004. Records property of hospitals—Access—Not Public—Funding for medical record requests. § 68-11-304(a)(1).

Terry, N. P. 2004. Prescriptions sans Frontières (or How I Stopped Worrying about Viagra on the Web but Grew Concerned about the Future of Healthcare Delivery), 4 Yale JHPLE 183.

Terry, N. P., and L. P. Francis. 2007. Ensuring the Privacy and Confidentiality of Electronic Health Records. *University of Illinois Law Review* 681:707–708.

Terry, Nicolas, 2008. *What's Wrong with Health Privacy?* in Law and Bioethics, Eds. Sandra H. Johnson, Ana S. Iltis, and Barbara A. Hinze (NY: Routledge, 2008), 68-94.

Thomas vs. United States. 1987. 660 F. Supp. 216, 218. (D.D.C.)

Unistar Corp. vs. Child. 1982. 415 So. 2d 733 (Fla. Dist. Ct. App. 3d Dist.).

United States Code. 1992. *Definitions*. 17 U.S.C.S. § 101.

———. 1998. *Digital Millennium Copyright Act*. 17 U.S.C.S. § 1201 et seq.

———. 2009. *Fraud and related activity in connection with computers*. 18 U.S.C.S. § 1030 et seq. http://www.law.cornell.edu/uscode/18/1030.html (accessed February 4, 2010).

United States House of Representatives. 2008. An Act relative to patient health care information House Bill 1587. http://www.gencourt.state.nh.us/legislation/2008/HB1587.html (accessed August 31, 2010).

Valley Med. Specialists vs. Farber. 1999. 194 Ariz. 363, 982 P.2d 1277.

VICTOR LALLI ENTERPRISES, INC., vs. BIG RED APPLE, INC. No. 1577, Docket 91-7155. United States Court of Appeals, Second Circuit. Argued May 29, 1991. Decided June 13, 1991.

Vt. Stat. Ann. tit. 18, § 4631 (2007), Vermont Statute.

Waldron vs. Ball Corp. 1994. 619 N.Y.S. 2nd, 841, 210 A.D.2d 611 (A.D. 3rd. Dept.).

Washington State Legislation. 2008. *Prohibiting the sale or use of prescriber-identifiable prescription data for commercial or marketing purposes absent prescriber consent*. S.B. 6241. http://apps.leg.wa.gov/billinfo/summary.aspx?year=2008&bill=6241 (accessed February 24, 2009).

West Virginia. 2008a. S.B. 544.

_____. 2008b. H.B. 2177.

Willison, D. J., L. Schwartz, J. Abelson, C. Charles, M. Swinton, D. Northrup, and L. Thabane. 2007. Alternatives to project-specific consent for access to personal information for health research: What is the opinion of the Canadian public? *Journal of the American Medical Information Association* 14(6):706–712.

Windeyer, J., Mount Isa Mines vs. Pusey, 1970. 125 CLR 383, 395 (Supreme Court of Queensland).

Woolley, M., and S. M. Propst. 2005. Public attitudes and perceptions about health-related research. *Journal of the American Medical Association* 294(11):1380–1384.

Wyo. Stat. Ann. 2005. *Grounds for suspension; revocation; restriction; imposition of conditions; refusal to renew or other disciplinary action*. § 33-26-402(a) (xxvii) (G). http://legisweb.state.wy.us/statutes/statutes.aspx?file=titles/Title33/T33CH26.htm (accessed February 3, 2010).

5

Healthcare Data as a Public Good: Privacy and Security

INTRODUCTION

Any consideration of clinical data as a public good raises questions concerning the safety and security of individual patient records. Maintaining confidentiality of data records is of paramount importance. Public perceptions of privacy in the context of medical records links directly to the trust the public has in the entire healthcare establishment, and factors significantly into discussions of health data sharing. The complex issue has many challenging dimensions, from what happens after the initial intake of an individual's data to what happens in data aggregation and secondary use. This chapter provides commentary from four experts considering key legal and social challenges to privacy issues from a variety of perspectives, including public opinion, the implications of the *Health Insurance Portability and Accountability Act* (HIPAA), and institutions' experiences inside and outside of health care.

To provide insight into the public views on privacy issues in health care, Alan Westin, professor emeritus of public law and government at Columbia University and principal of the Privacy Consulting Group, presents outcomes of the 2007 national Harris/Westin survey that evaluates public attitudes toward the current state of health information privacy and security protection.[1] The survey examines attitudes about handling sensitive patient information, health research activities involving individual patient data, and

[1] This survey was commissioned by the Institute of Medicine as part of the work of the IOM Committee on Health Research and the Privacy of Health Information.

opinions on the extent to which trust is accorded to health researchers by the public. The results indicate that the public holds strong privacy concerns about how their personal health information is handled, especially uses of data not directly relevant to providing care. The survey also indicates that current laws and organizational practices may not provide adequate privacy protection for patients. Westin suggests that patient-controlled privacy policies, such as those offered through repositories of personal health records, might help with gaining traction on the issues of clinical data, privacy, and security with the public. He also recommends a scope of activities related to health privacy, patient notice, and public education on privacy and compliance as opportunities to provide evidence-based medicine (EBM).

Balancing patient privacy protections with advancing data-driven clinical research and care delivery is an ongoing challenge for many healthcare organizations. In 2003, the HIPAA Privacy Rule took effect, and early changes to the Rule permitted sharing healthcare data for restricted purposes, essentially easing some limitations on providers and health plans related to health services research. With the increased incorporation of electronic health records (EHRs) into care delivery and research, the growing volumes of valuable data for evidence-based research and care may eventually force significant changes to strike a balance between privacy and advancement. Marcy Wilder, a partner in the law firm of Hogan and Hartson, LLP, and former deputy general counsel at the Department of Health and Human Services (HHS), where she helped to develop HIPAA, comments on some important remaining legal barriers to effectively using clinical data for research. In particular, Wilder highlights the growing opportunity to address the confluence of future, unspecified research and individual rights regarding the use of individual data through policy. Also notable are her suggestions of formally reviewing HIPAA deidentification standards, safe harbor requirements, and distribution of liability burdens across covered and noncovered entities.

Providing examples of other sectors' approach to striking a balance between privacy and security and research innovation, Elliott Maxwell, a fellow in the communications program at Johns Hopkins University and distinguished research fellow at Pennsylvania State University, discusses the notion of data openness as demonstrated through projects such as the Human Genome Project. Examples of greater openness are also prevalent in the public registration of clinical trials and open-access journals. Greater digital openness has the potential to transform the use and application of clinical data in EBM, Maxwell suggests, but it must be tempered with determinations on the appropriate level of openness for given purposes. Maxwell provides an overview of the Committee for Economic Development's report *Harnessing Openness to Transform American Health Care*, including recommendations on patient consent requirements, electronic filing of device

and drug approvals, and EHR adoption incentives. The report advocates for increased federal support for large, clinical databases to accelerate advancements in EBM and standards development.

The quality of clinical care and access to care services are ubiquitous issues in American health care. The public demands higher quality care at lower costs with greater access. Healthcare data are uniquely positioned to provide deep insights into care delivery processes and outcomes. Simultaneously, provider organizations must secure individual patient health information and improve the coordination and quality of care. The tension between access to insight-generating data and security of health data continues to create significant barriers for organizations striving to provide clinical services. Alexandra Eremia, associate general counsel and corporate privacy officer at MedStar Health, Inc., discusses perceived and actual privacy or security hurdles experienced at healthcare delivery organizations nationwide. She elaborates on the opportunities for building trust in patients, structuring organizational policies and strategies to avoid adverse legal outcomes, and making strategic fiscal decisions associated with data retrieval and release for research. Addressing these opportunities through financial, strategic, regulatory, and public initiatives may advance access to healthcare data for research and EBM purposes.

PUBLIC VIEWS

Alan Westin, Ph.D., L.L.B.
Principal, Privacy Consulting Group

Based on a national Harris/Westin survey in 2007 sponsored by an IOM project, this paper will describe public attitudes toward the current state of health information privacy and security protection; health provider handling of patient data; health research activities; and trust in health researchers. The public is segmented into persons who have participated in health research projects, those who have been invited but declined (and why), and those never invited. Members of the public are identified who believe their personal health information has been disclosed improperly and by whom. Explaining the benefits and risks involved in having one's personally identified health records used in health research, the paper explores what kinds of advance patient/consumer notice and consent mechanisms are desired by various subsets of the public. Potential privacy harms are documented that patients see if their health records are used without notice and choice mechanisms, or disclosed improperly. The findings are applied to emerging large-scale health data systems, especially new online personal health record repositories and health data-mining programs. In terms of

positive actions suggested by these survey results, updated federal health privacy rights in legislation supporting information technology/EHR programs are discussed, as are national educational campaigns on the values of health research under robust health privacy rules or procedures, and new software tools to put direct control over the uses of health records into the hands of individual patients, through an individually driven "switch" mechanism between health data providers and health-research data seekers.

Privacy is pervasive in terms of the future of health information technology (HIT). How the public feels about privacy issues links directly to the trust level that people have in the entire healthcare establishment, and factors significantly in the move to EHRs, personal health records, interoperability exchanges, and so forth. Trust is a fragile commodity. Anything that profoundly threatens the trust that patients have in the healthcare system and in health researchers is a very dangerous step. We need to be careful, and my hope is that the survey data reported here will document this.

A national survey sponsored by an IOM working committee (Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule) investigated how the public feels about privacy in health care and the use of their information across the spectrum of healthcare operations. The survey's sample was 2,392 respondents who were 18 years of age and older. The data were adjusted to represent the entire population of 255 million persons age 18 years and older. We could analyze survey results not only by the majority, but also by health status groups, by standard demographics, by people who reported on their personal experiences in healthcare use, and by their policy attitudes. This paper presents only top-level results; the full 2007 survey project report is available through the Public Access Records Office of The National Academies (publicac@nas.edu).

The survey formulated four statements and asked people to agree or disagree with each statement. The first statement was about how much people trusted their own healthcare providers—doctors and hospitals—to protect the privacy and confidentiality of their personal medical records and health information. A significant 83 percent expressed such trust, a result confirmed by many other surveys. (See Appendix D in the full 2007 survey project report, available from the IOM as shown above.) These surveys have shown high trust in the healthcare provider establishment as manifested in the direct relationships among the patient, doctor, labs, hospital, and so forth.

However, when we asked people whether a healthcare provider ever disclosed their personally identified medical or health information in a way they believed was improper, 12 percent said yes. That represents roughly 27 million adults. The survey report shows how many said the information was disclosed by their doctor, their hospital, their pharmacy, their lab, their insurer, and others. This response indicates that a significant segment

of the public is really not comfortable with the way even their healthcare providers have handled their confidential information.

The second question was how much people agreed with this statement: "Health researchers can generally be trusted to protect the privacy and confidentiality of the medical records and health information they get about research subjects." Sixty-nine percent said they agreed with that statement; fewer than for the healthcare providers, but still, a two-thirds majority endorsement of the health research function as seen by the public.

Our third statement asked for agreement or disagreement with this presentation: "The privacy of personal medical records and health information is not protected well enough today by federal and state laws and organizational practices." In previous health and consumer privacy surveys, we have worded this statement both ways: Sometimes we asked people to agree or disagree with the statement that privacy is "well enough protected," and the results come out the same. Fifty-eight percent of the public in this IOM survey said they do not believe there is adequate protection today for their health information, either from laws or from organizational practice. This suggests that HIPAA has not created a sense of comfort and security in the majority of the population. My sense is that this judgment is being driven in part by the constant reporting of health data breaches taking place, such as theft of laptops with medical information, improper disposal of hardcopy medical records, and insiders leaking medical information. Such losses may not be at the same incidence level as the theft of financial information or identity theft through capture of consumer data. But reporting of medical data breaches contributes, in my view, to the judgment of a national majority that their medical information is not effectively secured today.

Finally, we asked people to agree or disagree with this statement: "Even if nothing that identifies me were ever published or given to an organization making consumer or employee decisions about me, I still worry about a professional health researcher seeing my medical records." The public is split right down the middle: 50/50. Half agree with the sense that there is an exposure that worries them and half are comfortable. Underlying this finding was probably the feeling that "if strangers are looking at my sensitive medical information, I am not quite comfortable with that." The full report shows that this is more strongly felt by people who have potentially stigmatizing health conditions, such as those who use mental health services, have HIV or sexually transmitted diseases, have taken a genetic test, and so forth. Demographics and health status would give some subsets of the public an even stronger than 50 percent concern about this.

Given the mission of the IOM committee that sponsored the survey, our prime focus was on how people would relate to health research per se. Consequently, we asked people how interested they would be in reading or hearing about the results of new health research studies, causes and preven-

tion of diseases, and effectiveness of new medications and treatments. We cast the net widely and did not limit it to narrow, clinical trial-type health research. Matching other surveys, three-quarters of the public (78 percent) said they were interested in tracking that kind of health research.

Perhaps the single most important focus of our study was when we asked people whether they were ready to have their personally identified health information used by health researchers, and, if so, what kind of notice and consent they would want to have provided. The fact that this was an online survey enabled us to ask a detailed and carefully crafted question that described how health research is done and gave the arguments of health researchers in favor of general advance consent or consents based on promises of confidentiality and human subject or Privacy Board oversight. We also put in comments of "some people" that only notices describing the researchers, the research topic, and the research result uses would ensure adequate privacy protection.

Having presented our lengthy question, we asked people to choose one of five alternatives that best expressed their view. These were randomly presented to mitigate any presentation-order bias. A miniscule 1 percent said that researchers would be free to use their personal medical and health information without their consent at all. We might characterize this group as "let it all hang out."

Eight percent said they would be willing to give general consent in advance to have their personally identified medical or health information used in future research projects without the researchers having to contact them. This small group might be characterized as a segment of the national population having a "high trust in the research establishment."

Nineteen percent said their consent to have their personal and medical health information used for health research would not be needed as long as the study never revealed their personal identity and was supervised by an Institutional Review Board (IRB). These respondents were ready to trust such general researcher assurances.

The largest group, 38 percent, equivalent to about 97 million adults in the population, chose the following response: "I would want each research study seeking to use my personal identified medical or health information to first describe the study to me and get my specific consent for such use." Clearly what is on the mind of this group is an insistence on knowing who is doing the research, what the topic is, and how the information is going to be used.

Finally, 13 percent said they do not want researchers to contact them or use their personal health information under any circumstances. This might be called the "no trust at all" segment of the public.

However, one in five, or 20 percent, of respondents simply could not make up their mind. The fact that they could not choose one of the five

alternatives suggests that a large group out there needs to be better informed or to have the choices put to them in a way that they can recognize and then make a choice. A 20 percent nonresponse rate is quite unusual in policy-related survey research of this kind.

We asked those people who would require notice and express consent why they were adopting this position, providing four possible reasons. As one might expect, 80 percent chose "I would want to know what the purposes of the research are before I consent." Sixty-two percent said, "knowing about the specific research study and who would be running it would allow them to decide whether I trusted them." Fifty-four percent said they "would be worried that their personally identified medical or health information would be disclosed outside the study," and 46 percent would want to know whether disclosing such information would help them or their family.

When we turned to what kind of harm the 38 percent believed could take place if personally identified health information was disclosed outside the study group, the answers primarily focused on discrimination. Privacy and discrimination values have always been closely linked. One claims privacy in order to protect oneself against being discriminated against in some benefit or opportunity. Here, results showed that people worry that distribution of their medical data could affect their health insurance, their ability to get life insurance, or their employment, or that it could result in their being discriminated against in a government program. The smallest number (33 percent) worried about embarrassment in front of friends, associates, or the public.

Now, here are some overall impressions about the survey results. First, this survey confirms, as many surveys have shown, that large majorities of the public hold strong concerns over the privacy and handling of their personal health information, especially concerning secondary uses of the data not in the direct-care setting. A strong majority, 58 percent, do not believe that current laws and organizational practices provide adequate privacy protection. The majority generally trust health researchers (albeit researchers undefined as to what kind they are) to maintain confidentiality, but what some researchers might hope for—that a promise of nonidentification and IRB review would persuade the public to give advance general consent—is not where the majority of the public is ready to come out at the present time. Also, even though we told people that researchers were concerned about the heavy costs in getting advance notice and consent, or that this might corrupt samples from statistical validity, that was not enough to persuade a majority. However, it is fair to say that surveys would get some different numbers if different kinds of researchers and topics were specified, so this is a variable to be understood.

What are the implications of our survey for expanded health data uses? Clearly, we are in transition from a part paper and part electronic record

realm to an interoperable world of electronic health records, personal health records, and huge new online personal health data pools. This opens some potentially valuable public-good health researcher possibilities. Privacy, however, is a make-or-break issue for whether we are going to be able to achieve those advantages from large-scale health data research through electronic communication and transmission.

Of course, privacy is not an absolute. Rather, privacy is a matter of balance and judgment, and it is very contextual. Still, unless we can create what the National Committee on Vital and Health Statistics called a new data stewardship responsibility for health data holders and secondary users, we are going to lose the balanced-privacy battle, with the risk of sharp limits being placed on using personal health data for very important health research.

What elements would provide a positive health privacy context for health research? First, we need new legislation. HIPAA is outdated, as many people have said. The late Senator Edward Kennedy proposed support for HIT and EHR systems but, already, bills have been introduced by Senator Patrick Leahy and Representative Ed Markey to add strong privacy protections to any bill that will support the health information technology cause. Without my endorsing any of those bills specifically, it is clear we will have to write a new code of privacy confidentiality and security into the legislation that is going to help to organize and finance EHRs.

Second, excellent models of voluntary patient control privacy policies are being offered by some new repositories of personal health records. Microsoft's HealthVault is one example; Google Health has indicated it will do the same when it issues its health product shortly. Such models need to be encouraged and modeled by many others.

Third, we need independent health privacy audits and compliance verification processes. Although no instrument is ready now to carry this out in the health information technology field, new organizations with the right mixture of nonprofit, for-profit, government, and consumer groups could be developed. Such meaningful audit and verification mechanisms are absolutely necessary for public acceptance and trust of the new large-scale health research enterprises.

Fourth, there are some new, easy-to-use technologies for implementing patient notice and choice—not "trust me, I am going to store your data, I will only give it to the people you want," but rather some new "switch, not store" programs. These will register patients and collect their privacy preferences. Then, they will connect data seekers—such as health researchers—with the data holders (providers, insurers, Regional Health Information Organizations, etc.) and facilitate the exchange of that information, without the data content ever being kept by the switch. This interesting idea

could revolutionize the ability of patients to make informed decisions about the use of their personal information in health research.

Fifth, we need to conduct serious field research into how privacy is unfolding in the EHR programs being developed. Researchers need to survey the patients involved in EHR programs as well as to talk, onsite, face to face, about what experiences they have had, what worries them, whether and how those worries have been solved, and so forth. Otherwise, one is back at 10,000 feet talking abstract principles about EHR programs and privacy satisfaction. It would be highly valuable to fund and manage a program of empirical studies of the impacts of EHR systems on privacy, confidentiality, and security values.

Finally, the health establishment needs to sponsor a major national educational campaign to promote privacy-compliant, evidence-based health research. Without such a national campaign, the danger is that the balance side—the public-good aspect of sharing patient medical data—will not be fully appreciated by the current privacy-sensitive public.

HIPAA IMPLICATIONS AND ISSUES

Marcy Wilder, J.D.
Partner, Hogan & Hartson, LLP

This paper will address the HIPAA Privacy Rule (45 C.F.R. § 164) and its effect on data research. As healthcare and HIT systems evolve, experience suggests that modifications are needed to strike the proper balance between protecting patient privacy and making data available for research to improve healthcare quality and to lower costs. Early advocacy efforts by the research community resulted in changes to the Privacy Rule that lightened some of the administrative burdens on healthcare providers and plans associated with making data available for research purposes. In addition, HHS revised the Rule to permit disclosures of limited datasets for research purposes. Identifying and developing policy alternatives for addressing the most significant barriers that remain, including those related to future unspecified research and data deidentification, will be essential to promoting the research enterprise.

The HIPAA Privacy Rule was the first comprehensive federal health privacy regulation. At the time of its drafting, HHS was focused on protecting privacy and ensuring that information would continue to be available within the healthcare system for appropriate uses. HHS set a baseline, making clear that health information could be used freely for treatment, payment, and healthcare operations. Policy makers were also clear that before health information could be used for marketing, an individual's

authorization would be required. The extent to which health information should, as a policy matter, be made available for research was far less clear. HHS, other federal agencies involved in the HIPAA rule making, healthcare stakeholders, and consumer advocates did not agree among themselves or with each other. Many believed research should not be placed in the same category as treatment, payment, and healthcare operations. But at the same time, they did not believe that individual authorization should always be required before protected health information (PHI) could be used for research purposes.

Some in the research community argued that HIPAA does not and should not regulate research per se and that the Privacy Rule simply should exempt research uses and disclosures.[2] For nearly 25 years the Common Rule for Protection of Human Subjects ("Common Rule")[3] had regulated research privacy. IRBs were already tasked with determining whether protocols contained provisions adequate to protect the privacy of subjects and the confidentiality of data. The notion was to leave the current Food and Drug Administration (FDA), Office of Human Research Protections, and state regulatory frameworks in place and undisturbed by HIPAA.

This argument, however, was ultimately rejected by regulators. HIPAA restricted access by researchers to PHI, which at that time was held by healthcare providers and health plans. These HIPAA-covered entities would need guidance on how they were to treat uses and disclosures of PHI for research purposes. In addition, although longstanding protections were in place, some privacy advocates believed current protections were not sufficient. When HHS ultimately did address issues related to research uses and disclosures, it did not attempt to harmonize HIPAA with the existing regulatory framework for human subjects' protection. It simply added yet another layer of regulation.

By 2002, 2 years after the Final Rule was issued, there was enough experience to suggest that the HIPAA Privacy Rule was unnecessarily creating barriers to medical research and that some provisions needed to change. The research community focused a great deal of effort on the deidentification safe harbor and the fact that data stripped of all requisite fields were not useful for many types of important research. The Department's response was to add provisions permitting the disclosure of limited datasets for research, provided that a HIPAA-compliant data use agreement was in effect.

Under HIPAA, as initially promulgated, before information could be freely used for research, it needed to be deidentified under strict standards.

[2] 67 *Fed. Reg.* 14776, 14793 (Mar. 27, 2002).
[3] 45 C.F.R. § 101.

In response to concerns expressed by the research community, HHS introduced the notion of a limited dataset, which is essentially deidentified data plus ZIP Codes, dates of service, and other dates related to the individual.[4] If a party wanted to use this partially deidentified information for research, it could enter into a data use agreement, the contents of which are prescribed by the regulation, promising to protect the information. Once the agreement was executed, the dataset could be released for research purposes. These provisions did enable researchers to obtain health data more easily. Although there is a question as to whether these provisions are sufficient, they clearly helped.

In addition, in 2002 HHS provided an alternative to the accounting of disclosure requirement.[5] The accounting of disclosure requirement mandates that when covered entities such as hospital systems and health plans disclose information for research purposes pursuant to an IRB waiver, they need to keep an accounting of these disclosures and make it available to individuals on request. Keeping individualized records about which records were disclosed for which research protocols operating under an IRB waiver of consent was seen as quite burdensome by the covered entities. As a result, many covered entities, and in particular smaller hospitals and those not affiliated with an academic institution, were restricting access to data.

HHS came up with an alternative. Instead of keeping track of every time data were disclosed pursuant to an IRB waiver, an institution could keep a list of all the research protocols for which information was disclosed pursuant to an IRB waiver for research purposes. Anyone requesting an accounting of disclosure would be given the entire list, which for institutions such as an academic medical center could be voluminous and burdensome to maintain. On request for an accounting of disclosures, the list would be provided and the individual would, in effect, be told that perhaps his or her information had been disclosed for one of the protocols on the list. The extent to which this is privacy protective or helpful to the individual is questionable at best. It seems to constitute an example of a privacy protection or a requirement that imposes cost and burden, yet does not deliver any meaningful privacy protection. Nonetheless, that is the current standard.

Experience over the past few years has helped highlight the need for further changes. The landscape surrounding research data has changed considerably, due in large part to significant technological changes that permit data aggregation on a scale that was previously unimaginable. In addition, emerging technology used by Google, Microsoft HealthVault, Dossia,

[4] 45 C.F.R. § 165.514(e).
[5] 45 C.F.R. § 164.508(c)(1).

WebMD, and others that will be aggregating data on behalf of consumers will further change the extent to which data are available for research.

One issue that should be revisited in this new context is the HIPAA deidentification standard.[6] HIPAA now includes a deidentification safe harbor, which says in essence that data are deidentified as long as 18 specified data elements are removed and the covered entity does not have actual knowledge that a person could be reidentified from the dataset under the safe harbor. All the demographic data, all dates related to the individual—including date of birth, dates of service, and ZIP Code—and all unique identifiers—such as medical record number—must be removed.

Alternatively, covered entities can use the statistician method, under which a qualified statistician can certify that a dataset is deidentified. At the time the Privacy Rule was drafted, it was believed that a cottage industry of statisticians who were willing and able to certify large deidentified datasets would emerge. In practice, however, only a handful of statisticians are available to provide these certifications. Although a number of large data aggregators are using statistically deidentified datasets, it is not the industry norm for research enterprises.

Some argue that the deidentification safe harbor is too narrow and that researchers should be able to freely use those data that include some elements on the safe harbor list. However, some privacy advocates argue that in the Internet age, there is no such thing as deidentified data—that because of the widespread availability of electronic information and the ability to aggregate it, personal data cannot ever be deidentified. The resolution of this debate will have profound implications for public health research, epidemiology, and the future of research on large datasets.

Another issue that should be reconsidered is whether individuals should be permitted to authorize the use of information about them for future unspecified research. Today, obtaining a HIPAA authorization for uses and disclosures for future unspecified research is not permitted under the Privacy Rule.[7] Privacy advocates argued at the time of the rulemaking that there is no way to adequately inform an individual about the privacy risks related to future unspecified studies. Research stakeholders, however, pointed out that individuals can and have authorized such uses under the Common Rule and that not permitting such authorizations unnecessarily limits and harms the research enterprise.

Another set of issues that needs to be discussed concerns whether liability burdens under the HIPAA Privacy Rule are properly distributed. Although research data involving PHI are held by both HIPAA-covered and -noncovered entities, liability risks reside largely with the HIPAA-

[6] 45 C.F.R. § 164.514(a).
[7] 45 C.F.R. § 164.508.

covered entities. In addition to confusing rules and administrative record-keeping that many covered entities—smaller hospitals in particular—find unduly burdensome, the effect of these liability risks is that covered data holders, including hospitals, health plans, and other large HIPAA-covered entities with sizable pools of data resist expending resources to create the deidentified data or limited datasets. They see little reason to spend the requisite time and money so that others can have large datasets on which to do research.

Finally, it is also true that HIPAA sometimes provides a convenient excuse for those who simply do not wish to share their data. Many stakeholders, for a variety of reasons, do not want to share. To successfully address the needs of both patients and the research community, policy makers need to understand which barriers regarding data sharing need legislative or regulatory solutions. This can be hard to discern because HIPAA is so often put up as a smoke screen to preclude the sharing of data.

Congress, agencies within HHS, and numerous advisory committees have recognized that the HIPAA research provisions need updating and improvement. As Congress and HHS examine and enhance federal health privacy protections, new opportunities for addressing the needs of the research community will emerge. Those who are ready with concrete and realistic proposals and solutions will be those most likely to succeed.

EXAMPLES FROM OTHER SECTORS

Elliot E. Maxwell, J.D.
Communications Fellow, Johns Hopkins University

This paper will attempt to put the use of healthcare data into the larger context of transforming health care by increasing openness. This means providing more access to more information to more people and allowing individuals to contribute their own expertise and insights to that information.

One of the many examples of increased openness in health care can be seen in the collaborative research model of the Human Genome Project, with results posted immediately, available to the world. Congress has mandated greater openness by requiring the public registration of more clinical trials. New models of disclosure and publication of research results in open-access journals and digital repositories provide greater openness. Greater access to information is transforming the relationship between doctors and patients and is increasing market incentives for improved health care.

Greater openness is not always better because of privacy and security issues. The challenge is to determine what level of openness is most

appropriate for the particular purpose to be accomplished. Clearly, greater openness has been made possible through the rise of the Internet and the increasing digitization of information. These two phenomena have helped change how we think about the ability to access information and the ability of millions, even billions, of people to make contributions. The Internet has also led to a new understanding of how to obtain value through the sharing of information.

It was commonly believed that value from information or innovation came through controlling it. By monetizing through licensing or other means, it was critical to control access to the information to obtain value from it. That notion of control underlies current intellectual property laws and is the basis for proprietary models.

Openness may be thought of as a continuum, running from closed to open. At one end are those things completely closed and controlled, such as writing a song, but never sharing the score. A point toward greater openness on the continuum is occupied by proprietary software that one can license under generally restrictive conditions. At the other end, the open end, is the sharing of information with anyone and everyone as seen through virtual posting of nearly anything on the Internet. Granted, someone might request that the material be removed because of intellectual property rules or some other legal restriction, but there is no central control agent that prevents individuals from posting. There are now licensing schemes, such as those that have evolved through the Creative Commons (http://www.creativecommons.org), by which the poster can inform everyone about the restrictions that exist on the use of the posting. These licenses also allow the poster to announce that there are no restrictions—that the information is completely "open." However, if no restrictions are placed on the use of the information, does it still have value? With the increasing use of the Internet, we are finding that we can obtain great value from sharing information. Wikipedia is an example of the creation of great value via sharing.

Wikipedia is not completely open, but it is much closer to completely open than the proprietary *Encyclopedia Britannica*. Since its founding several years ago, Wikipedia participants have created five times the number of entries found in *Encyclopedia Britannica*, which has been under development for a hundred years. A study in *Nature* found that scientific entries in Wikipedia were substantially equal to those in *Britannica*. We do not need to equate the two or to argue that Wikipedia should be regarded as a definitive source, but indisputably it has provided great value to millions of people, all based on contributions without any expectation of monetary reward. On the openness continuum, Wikipedia is not completely open. Postings can be taken down under certain circumstances by people trained and empowered to make judgments about, for example, the "neutrality" of a posting.

Similarly, open-source software, which is usually thought of as open because its underlying source code is available without restriction, is not entirely open. A software application like LINUX cannot be entirely open because no one would use an application that would change every time someone suggested an improvement. At the same time, open-source software is licensed in such a way that the source code will be seen by as many people as possible, which is the key factor in its success. It aims for continuous improvement through widespread sharing because such sharing makes it more likely that someone, somewhere, will have the inclination and the expertise to review and improve the code. The more open it is, the more likely it is to get better; however, it is not completely open because any new version of LINUX will not be released until a group of experienced coders exercise their judgment and determine that that version is ready for prime time.

The success of Wikipedia and open-source software demonstrate the power of the Internet and how value can be added by sharing rather than by exercising strict control. This success also reveals how openness allows value to be obtained from unexpected sources. At the same time, openness invites negative contributions as well as positive. Contributions are welcomed from experts, but also from a broad range of people because of the assumptions that many people can add value, but how many cannot be determined ahead of time. Another fundamental assumption is that the value of contributions from unexpected sources outweighs the cost of screening out contributions that do not add value.

This orientation toward facilitating contributions from a broad range of individuals and organizations underlies the development of what some have called Web 2.0. In the early days of the World Wide Web, it was equated with the great library at Alexandria—it would provide access to a vast store of information to anyone with an Internet connection. That was great—but passive in that little thought was given to what those people could contribute. How could they modify it, what could they do to improve it based on their own expertise and experience? The degree to which people can modify the information or the process determines how responsive the information or process is. Openness is about both accessibility and responsiveness.

As noted earlier, the greatest degree of openness is not always the best answer to the question of what degree of openness is best suited for a particular purpose. This is where we can begin to ask salient questions regarding health care. What kind of information do I need for this purpose, whether it is research or treatment or the detection of emergent diseases? Who should have access to it? Under what terms and conditions? Should I allow other people to contribute to it? Should everyone be able to contribute, or only those prequalified in some way? Can they modify, repurpose, or redistribute it, and if so, under what terms and conditions, if any?

The report entitled *Harnessing Openness to Transform American Health Care,* recently released by the Committee for Economic Development (CED) examines the terms and conditions under which greater openness might improve health care across its entire production function, from research to treatment (CED, 2008). The report also considers cases in which greater openness can be harmful—such as unauthorized access to medical records or unauthorized disclosure of genetic information about an individual—and destabilizing such as in the relationship between patients and their caregivers. The report is not exhaustive, but it is a first attempt to use the lens of openness in an area rich with opportunities for improvement.

The report is about openness rather than the use of information and communications technology (ICT) in health care because, ultimately, openness is about an attitude; openness is about a predisposition for giving more people more access to information and more opportunities to contribute based on their own expertise and insight. There are scores of very good reports on the value of increasing the use of ICT in health care, but that should not be equated with increasing openness.

A wonderful example of openness in health care has nothing to do with information and communications technology. A researcher at the University of California–Los Angeles has bush hunters in Cameroon send him samples of diseased animals they have caught. He then examines them for evidence of emerging diseases. That is about as far from HIT as you can get. But it is based on the recognition that great value can be obtained from unexpected sources—in this case the hunters add great value because they are operating on the front line of emergent diseases.

There are many other examples of openness, some of which are based on the use of information technology, while also demonstrating the importance of an attitude of welcoming contributions from others. Some of the most interesting examples come from interactions between caregivers and patients.

About half of primary care physicians report that their patients have arrived with research from the Internet. What does that tell us about openness? Obviously information technology has provided patients with greater access to information—some of it valuable, much of it wrong or inapplicable. What should caregivers do? They can elect to dismiss the Internet research out of hand or imply that valid information can come only from the doctor. They can treat such circumstances as a learning opportunity, educating patients to separate good research from bad. They might even find research that is new to them. In this context, openness is destabilizing the traditional doctor-patient relationship, but the end results may be more informed patients who can take more responsibility for their own

health, and new and rewarding partnerships between caregivers and their patients.

As the report touches on many areas, this will focus on the openness issues surrounding clinical data and electronic health records. Many kinds of information would be valuable for patients that are not available now. For example, it would be useful for patients whose caregivers recruit them for clinical trials to know whether the caregiver is being paid to recruit; similarly, it would be valuable for patients to know whether the caregiver has a financial interest in the treatment being recommended or is receiving gifts from a pharmaceutical firm whose product is being prescribed. In such cases more openness and more information would allow patients to make more informed decisions.

Congress acted recently to increase openness regarding clinical trials and posttrial surveillance. For drug trials that lead to an application for approval by the FDA, there is no compelling argument for disclosure before the application. On the other hand, data indicating safety issues that lead to a trial's termination should be made available immediately so that others do not repeat the trial and put trial participants at risk unnecessarily.

Currently, when an application for approval is filed, the information associated with the application can be protected as a "trade secret." It is not available at the time of the application, nor when the application is approved; it can be withheld for an extended period of time beyond approval based on its trade secret characterization. Moreover, once it is submitted it falls into a kind of regulatory black hole where the FDA makes no affirmative effort to make it available to researchers who might benefit from access to it. Even if the data are in an electronic form, they may not be arrayed in a manner that would allow other researchers to aggregate and manipulate them and to use them to develop more comprehensive databases.

There does not appear to be any compelling reason to withhold these data after a drug or other intervention has been approved. Companies have argued that the data should not be made available at all because doing so would provide a shortcut for competitors, but the company that submitted the data has had a large head start over any competitor because it has had years to scrutinize the data. The FDA approval has provided the company with a substantial legal benefit. On the other hand, withholding the data prevents academic researchers interested in the efficacy and safety of the intervention from benefiting from the data. Access to data underlying clinical trials does raise important questions of openness, including the value of the data to the company that submits the data and to competitors. But in making a decision about disclosure, the most important criterion should not be the impact on competition between drug-producing companies, but on the societal value of providing the information to researchers in general.

The specific recommendations made in the CED report regarding clinical trials and postapproval surveillance include the following:

- The FDA should review existing requirements on patient consent to participate in clinical trials and make changes as appropriate. The bifurcated authority in this area should be ended.
- Those recruiting participants for clinical trials should be required to disclose any financial interest in the recruitment.
- The FDA should require electronic filing for all drug and device approvals.
- The FDA should set standards for and require the filing of underlying clinical data, on approval, in a form that allows subsequent machine aggregation, search, and manipulation.
- The FDA should require the filing of all studies that an applicant has commissioned on a drug or device that is being submitted for approval, whether or not the study was commissioned as part of the application.
- The FDA should consider making public any studies that it conducts in the course of a drug or device approval.
- Those conducting clinical trials should be required to report to the FDA, on detection, any instances that would reasonably suggest the use of fraudulent data.
- The FDA should require disclosure of any limitations on researchers' ability to comment on clinical trials with which they are involved.
- The FDA should broaden the means by which postapproval adverse events can be reported and should make the reports more widely available.
- The FDA should encourage the disclosure of postapproval data indicating the efficacy of interventions for nonapproved purposes.
- The federal government should dramatically increase its efforts to directly compare the safety and efficacy of similar drugs and devices.

The report also deals with openness issues involving electronic health records. As with the data underlying clinical trials, data from EHRs are likely to be critical components of large databases that will serve as the breeding grounds for development of evidence-based medicine. The active mining of these large databases, whose development has been encouraged by the latest amendments regarding the FDA, should expand the number of medical practices that can claim an empirical base. At the same time, the CED report noted the need for disclosure of conflicts of interest by anyone participating in developing recommendations for clinical practice regimens.

The countervailing interests of privacy and security are evident in any consideration of openness and EHRs. Patients want their EHRs to be open enough to be accessible to anyone treating them, and open enough to receive data regarding appointments, prescriptions, treatments, tests results, etc., but not so open as to allow anyone to have access to them without authorization. It will be critical to resolve these tensions in a way that enlists patient support for the development of EHRs. Development of an interoperable system of EHRs is stymied now by a lack of standards, a lack of incentives for the predominantly small medical practices to adopt them, and a lack of demand from patients due to concerns about privacy and security.

Under today's HIPAA rules, there is a presumption toward openness in that a patient has a right to a copy of his or her records. But, as with filings regarding clinical trials to the FDA for drug approvals, there is no requirement that these records be in a digital form.

The CED made several recommendations regarding EHRs:

- Individuals and groups providing and funding health care should institute appropriate incentives for the adoption of information and communications technologies (including EHRs) to reduce health care's burdensome administrative costs.
- Federal research agencies should increase their support for the development of the large databases necessary for progress toward evidence-based medicine, including developing the necessary data standards.
- Strict requirements on the disclosure of conflicts of interest should be applied to those participating in the development of recommended clinical practice regimens.
- HIPAA should be amended to require that those parties who hold a patient's medical records must provide the patient with the opportunity to receive copies of those records pursuant to HIPAA in digital form.

Concerns about privacy and security are among the principal impediments for the development of an interoperable system of EHRs. But although there is much debate about how to deal with privacy and security, both technological and marketplace forces are racing ahead, rendering HIPAA's privacy regime increasingly problematic. For example, there are more than 200 different systems of personal health records now in the marketplace, including Microsoft's HealthVault. Yet because Microsoft is not a medical provider, the company is arguably not covered by HIPAA's privacy requirements.

Clearly some HIPAA rules should apply. Perhaps they will be covered

by an authority based on the Federal Trade Commission's jurisdiction over advertising of privacy, protections, or some other regulation that makes any entity that touches personally identifiable health data a steward of such data, with some enforceable responsibilities. Ensuring that entities that have this sort of information are covered and that the rules governing their responsibilities and obligations are clear will be of ongoing importance.

The need for clarity in the rules is often overlooked. One of the principal lessons from the failure to share information regarding the Virginia Tech shooter was that the individuals and institutions that had relevant information did not understand what they could do based on existing privacy rules—so they too often chose inaction as the safest response. That should not be the case. If the future of health care involves a far richer data environment, as I believe it must, we will need clarity in the rules regarding privacy and continuing educational efforts about what is and is not allowed.

Another lesson from EHRs is that there is a high level of anxiety and disquiet about privacy because there are no generalized privacy protections in the United States. Many individuals do not believe the environment is structured to protect their privacy. In this country we have tended to address privacy issues in "silos." We tend to identify particular information or particular technologies and address their privacy implications, rather than looking more broadly at privacy interests and how to promote them. Individuals are actually expected to understand the different privacy regimes of different domains.

There is yet another issue as to whether the respective rules are being enforced. One of the reasons for the log jam about EHRs is the belief that enforcement of the HIPAA Privacy Rule is nearly nonexistent. The people who are supposed to be protected must believe that they are, and that bad actors will face consequences. In yet another example of the power of openness, one needs access to the protected records to determine if there has been wrongdoing, such as unauthorized distribution of protected information. Ensuring access to one's records has been a fundamental part of privacy law for the past 25 years. The Freedom of Information Act practices require that you have the right to know what information is held about you and what has been done to it, and that you have a right to correct mistakes in the records. We do not have that today in a meaningful way with respect to medical records—and we should.

How to break the log jam regarding privacy and security is unclear, but here is a modest proposal. It will be challenging to develop the kinds of databases needed to provide evidence-based medicine unless there is societal agreement about the level of required protection for privacy and security. One part of the protection is the deidentification of patient records. The questions of how to deidentify the records and what level of protection is

required are not something that seems particularly amenable to congressional resolution. Perhaps Congress should commission The National Academies to formulate recommendations for the rules regarding deidentification within 18 months. The Academies would be told to use their judgment to make the best recommendations technologically, economically, and ethically. Such recommendations would, on their own, be useful, but we could take it a step further and have Congress treat these recommendations as they did recommendations from the military base-closing commission by making them subject to an up or a down vote. This is not the most elegant solution or one that is consistent with what we learned about civics in high school, but we need to resolve these issues in order to obtain the benefits of greater openness, particularly those related to the use of clinical data to develop more evidence-based medicine. We need to cut through this Gordian knot. This may not be the right way, but at least it is a way of dealing with these issues.

Without transparency, without clear rules, without some reasonable expectation of enforcement, there will continue to be great reluctance on the part of many people of good will to allow clinical data to be used to improve the general provision of health care. The benefits seem too far away, the threats too real and immediate. We will need to address both benefits and risks in order to foster a more open system. The CED report is an attempt to show the benefits, but it is only a beginning.

INSTITUTIONAL AND TECHNICAL APPROACHES TO ENSURING PRIVACY AND SECURITY OF CLINICAL DATA

Alexander D. Eremia, J.D., LL.M.
Associate General Counsel and Corporate Privacy Officer,
MedStar Health, Inc.

Healthcare providers have a duty to protect and secure the health information they receive or generate relating to their patients. Including their professional and ethical obligations, healthcare providers are now subject to a wide range of state and federal laws that impose various requirements and standards for the protection of health information. In addition, evidence suggests that most patients do not want their private health information to be: (1) *accessed* by people who do not need to see it; (2) *used* for purposes that will not benefit the patient; or (3) *disclosed* to someone who is not required to protect it or who might use it in a harmful manner (Westin, 2008). As a result, many healthcare providers view the protection and security of their patient's health information as essential to maintaining the trust and confidence of their patients and an

important element of patient satisfaction. At the same time, healthcare providers are rich sources of data, which when properly used in research, have the potential to greatly enhance the quality of clinical care and may result in better clinical outcomes, improved efficiencies, cost savings, or other medical advances.

Healthcare providers have an interest in each of these goals, but perceived and actual privacy or security hurdles, patient trust considerations, potential legal consequences, and actual costs associated with retrieval of data pose barriers to releasing data for research purposes. In particular, healthcare providers often find the privacy and security requirements of HIPAA confusing, and health information data custodians and researchers sometimes have limited awareness of HIPAA's data access and disclosure requirements.

Furthermore, even when access and disclosure are permitted under HIPAA, minimum necessary standards, accounting for disclosure obligations, and other patient considerations may impede the willingness to make certain disclosures of identifiable information. In addition, it is often costly for healthcare providers to divert resources and personnel away from their primary clinical care activities to attend to administering system and records access/disclosure activities for research purposes. Although technological solutions have the potential to mitigate some of these costs and resource burdens, at the current time, few such tools adequately address all of a healthcare provider's privacy requirements. In fact, often the implementation of new information technology brings with it additional complexities with respect to the ability to properly control research-related access.

As a result, healthcare providers are often more motivated to protect patient privacy, to respect physician-patient relationships, to minimize the administrative impact on data retrieval, and to minimize legal risks and customer complaints than they are to accommodate the needs of researchers. Absent adequate financial or strategic incentives, regulatory amendment, and greater appreciation of the public benefits of research, access to identifiable data for research will remain a challenge.

MedStar Health is the largest provider of healthcare services in the mid-Atlantic area, composed of eight hospitals, including community-based hospitals and academic medical centers, as well as numerous satellite clinics and outpatient facilities. In the District of Columbia, we own and operate Georgetown University Hospital, the National Rehabilitation Hospital, and Washington Hospital Center. Collectively, our system has about 25,000 employees and at least 5,000 affiliated physicians. System wide, we annually serve some 158,000 individual inpatients, have 787,000 inpatient days, treat 1,561,000 individuals on an outpatient basis, and make 208,000 home health visits. Therefore, the MedStar Health community is a rich source of diverse data that are potentially of great use to research. In

that context, this paper will reflect on some of the institutional challenges that we have balancing patient privacy interests with providing access for research purposes.

At MedStar Health we have a vision of being the "trusted leader in caring for people and advancing health," and we have long had a commitment and philosophy of putting the "patient first." As a result, our leadership feels strongly that beyond what the law says, we are devoted to protecting the interests of our patients and their information, and we are committed to promoting the trust of our patients by protecting their privacy. At the same time, we have a strong commitment to innovation, the promotion of research, and a shared vision of "advancing health" though our education, technology, and research capabilities.

As a part of MedStar's operations, we regularly create and maintain a number of databases and record sets into which patient information is placed, processed, and stored. Given the wide range of services provided by MedStar Health and the diverse patient base we serve, both the volume and the variety of data within these resources are large. We therefore are approached on a regular basis by researchers outside of our covered entity who request both large data sets as well as ongoing open access to patient information.

One area of significant concern, therefore, is how to most appropriately release and provide access to information to researchers who are not members of our workforce. Because these databases, repositories, and record sets are usually created primarily for treatment, healthcare operational purposes, or billing and financial purposes—not for research purposes—they often lack a built-in framework for addressing the needs and requirements associated with research-related access as well as the obligations we have for research-related disclosures. Furthermore, even when record sets are created in anticipation of potential research, they are often not designed to adequately facilitate compliance with privacy and security requirements.

Consistent with trends across the healthcare industry, MedStar is in the process of transitioning from being a largely paper-based organization to one with electronic records. We actually have four or five separate unique, stand-alone, traditional electronic medical records. In addition, MedStar has also developed a product, which was ultimately bought by Microsoft, that aggregates data from disparate systems, and which has led to an ongoing development relationship between MedStar Health and Microsoft. Although these systems have greatly facilitated our healthcare activities in many ways, one of the largest challenges we have with respect to research interests is getting information out of these systems in a cost-effective format that is useful to researchers.

Data that we collect, capture, and hold—whether in electronic or paper format—are generally meant for our own internal operational and clinical

purposes, and often are not easily retrievable in a format that is usable for research purposes. Even when the data are in electronic format, in many cases the way in which information is accessed and used for normal operations purposes (e.g., the types of queries made, the specific data points that need to be viewed, and the actual ways in which the data will be employed after retrieval) differs greatly from the manner in which researchers wish to interact with it. Because these systems have generally been designed and implemented with operational needs in mind, as opposed to the needs of researchers, the sort of retrieval, aggregation, and analysis tools necessary to researchers are often not readily available within these systems. As a result, often the extraction of data from our electronic systems requires a fairly manual and laborious manipulation process. To optimally meet researchers' needs and to contain costs for covered entities would, for some systems, require the development of new software interfaces and tools, all of which require investments of time and resources.

Furthermore, some of the information that MedStar collects or creates is proprietary information that we are unwilling to share in unfiltered/unredacted format, if at all. For example, we often get research requests for our billing and coding information. Although it may be possible to remove such confidential, proprietary information from a dataset intended for research use, it can be difficult to dissociate this information from what we are willing to share. In many cases this removal would require intelligent software capable of making fine-grained discriminations and would be very costly.

Moreover, concerns are sometimes raised by healthcare administrators that the goals of research are incompatible with the goals of being a leading community-based healthcare provider. Some of the goals of research—such as furthering scientific progress, translating research into improved clinical care, improving society, maintaining scientific integrity, perhaps pursuing technology transfer opportunities—are obviously all valuable, and no one denies that having appropriate information available to use for research is a public good. Nonetheless, such goals can sometimes run counter to the immediate goals of healthcare providers, which are fundamentally to provide quality health care to patients that results in high levels of satisfaction, trust, and confidence and to do this all on increasingly slim operational margins. For many healthcare providers, these goals (or at least the processes involved in achieving these goals) appear incompatible.

Beyond logistical barriers and differences in goals, moreover, HIPAA poses further obstacles to sharing information to outside parties for research purposes. The Privacy Rule continues to be confusing to many healthcare providers, who often view its requirements as arbitrary and overly complex. Healthcare administrators often face the burden of too many forms and policies that are generated as a result of our responsibilities to protect

patient privacy. Our administrators complain that they have inadequate resources to review requests and to assist in providing requested information, and this potentially results in reduced access to records and data. Furthermore, as with any large workforce, we experience frequent staff turnover, which results in a continual challenge of adequately educating our administrators and record custodians about how and when they can appropriately release health information for research-related activities. Similarly, often researchers and their staff do not understand or fully appreciate the requirements that we must fulfill with respect to the control of health information or the complex documentation requirements relating to the release of health information.

As an example, although researchers may understand that in order to review record sets for the purpose of identifying prospective participants, they must obtain a waiver of the requirement for authorization from an IRB, but they are rarely aware (or may not be concerned) that any access to PHI by non-MedStar research personnel that occurs under this waiver triggers an accounting of disclosure requirement on the part of the records custodian. As a result, they may not take proper steps to ensure that they limit the scope of their requests, limit which other persons receive the screening information, or adequately notify the records custodian of the involvement of external personnel and take steps to facilitate our accounting requirements.

Although IRBs can play an important role in ensuring that researchers properly address not just their own privacy requirements, but those of the information provider(s), IRBs need not be affiliated with the Covered Entity to grant a waiver of the authorization requirement and may not be entirely concerned with the Covered Entity's obligations. In addition, many IRBs also have regular turnover and have many members, including unaffiliated community representatives, who sometimes do not understand the requirements for protecting patient privacy. These issues of affiliation and education—whether it is of staff members, researchers, or IRB members—add to overall concerns in maintaining the trust placed in us by our patients.

To provide a few concrete examples of the challenges posed by the intersection of privacy concerns and research interests, I would like to focus briefly on a few specific issues that MedStar has encountered: (1) the need to adhere to different standards depending on who is requesting the information; (2) the potential for needing to honor patient restrictions; and (3) issues related to the relationship between an individual physician and his or her patient and the hospital in which the physician practices. These cases illustrate some of the difficulties inherent in trying to bridge the tensions between these two interests.

As touched on briefly above, depending on the specific relationship

between an individual researcher and the Covered Entity whose health information he or she wishes to access, HIPAA requirements associated with access differ. Although researchers are permitted to access PHI for research purposes without an authorization under the Privacy Rule, any time a researcher who is not a member of that entity's workforce does so, it is considered to be a disclosure that the entity must track and be able to account for on request. This is extremely burdensome for healthcare providers, particularly in the paper world, and often necessitates physically placing a marker or informational sheet in each record accessed. One might think this would be easier in an electronic world, but in reality it is not! Most of our electronic systems (especially our billing and other operational systems) do not include functionality that allows the adequate tracking of these disclosures with the level of associated information and detail required by the Privacy Rule.

HIPAA's alternative accounting mechanism, which provides for group or bulk accounting in cases where more than 50 disclosures are made for an individual study, is not really a viable alternative for a large decentralized and integrated healthcare organization. Without a central clearinghouse for evaluating data requests and/or registering the individual studies for which requests are made, it is difficult to confirm which studies may have had information released. For instance, for appropriate clinical efficiencies, some of our clinical systems allow physicians to access health information regardless of where the patients were seen in our system. As a result, it is possible that a researcher in Baltimore could request and access patient information from a system accessible at their location (i.e., in one of the Baltimore facilities) relating to information on a patient who was not seen in that Baltimore facility. Consequently, if a patient from the non-Baltimore facility requested an accounting of disclosures, it may be challenging to determine whether an accountable disclosure was made by the Baltimore facility. Dealing with this situation effectively would require the centralization of all research and other requests, so that all requests are handled by one central administrator. Unfortunately, this would be extremely burdensome and not currently a viable option for us because we have received potentially thousands of separate requests from thousands of different studies, resulting in hundreds of thousands of research-related disclosures over the course of the prior 6 years.

The issue of accounting for disclosures is one where researchers themselves could do much to help institutions in meeting the burdens associated with their privacy requirements and, in so doing, increase institutions' willingness to provide information for research purposes. Among the ways this can be done are: (1) developing or subsidizing the development of disclosure tracking software; (2) subsidizing staff positions dedicated to meeting accounting requirements (records custodians are often severely overworked

and unable to shoulder this); and (3) personally providing required information sheets or disclosure data where necessary. These strategies and "unforeseen" costs associated with data screening and recruitment should be considered by researchers when calculating the costs of conducting research at the time of grant application or protocol development.

Another difficulty associated with the research use of patient information involves the potential of "patient restrictions" placed on the use or disclosure of their own health information. Under the Privacy Rule, patients are permitted to request a restriction on how their health information may be used or disclosed. However, the Privacy Rule does not require a Covered Entity to accept that restriction request and, in fact, most healthcare providers try not to, because it is extremely burdensome to honor these requests. Even if such restrictions are accepted, healthcare providers are not necessarily culpable under HIPAA if the release of information is for research purposes.[8] Nonetheless, we believe that if we make a commitment to our patients, we are ethically obligated to try to fulfill it.

Though most Notices of Privacy Practices require that any request for restrictions be placed in writing and though most Covered Entities try to educate their staff to not accept a restriction unless it is in writing and clearly agreed to, it is possible that physicians or other staff members occasionally and informally make commitments and promises to their patients that their health information will not be used for any purposes except their own treatment unless the patients otherwise consent. In some cases, the physician or staff member may actually sequester a file or flag in an attempt to limit access to the information. Unfortunately, because billing systems, registration systems, and other clinical systems are often highly integrated, it is often difficult for healthcare providers to completely restrict who accesses and uses the patient's identifiable health information. When the patient is contacted by an outside researcher (even if the researcher legally and properly obtained the patient's information), the patient will obviously feel betrayed and lose confidence in his or her healthcare provider.

Given the number of employees that can potentially access any given patient's records, it is difficult to ensure that a pledged restriction made by one staff member or physician is known and adhered to by others. This issue, furthermore, is inherently resistant to a centralized solution because of the individual nature of the patient–provider relationship. Even with a centralized office for accepting and implementing patient restrictions in place, it would not prevent individual physicians from making personal agreements or commitments with patients that do not get propagated across the system. This challenge is, similarly, more difficult for researchers themselves to help mitigate than, for example, the accounting of disclosures

[8] 45 C.F.R. 164.522(a)(1)(v).

requirement because the researcher has little ability to discern where restrictions may be in place if they have not been adequately marked by those who accepted the restriction. As a result, completely confirming that healthcare providers are not violating any individualized commitments prior to making a research-related disclosure would literally require confirming such with each individual treating provider (obviously an insurmountably burdensome task).

Finally, another example of a problematic barrier has to do with the physician–patient relationship. A small but vocal community of our physicians has strongly objected to us allowing health information about "their" patients to be accessed by researchers. Some feel strongly that researchers are effectively trying to "cherry pick" their patients because some of these researchers are also clinicians. They have also argued that this violates the trust of their patients because patients may not understand why some outside researcher with whom they have no existing relationship is contacting them. It is argued that this may be perceived as similar to providing their contact information for "cold-calling" purposes. An additional concern is that these patients might get enrolled in trials that contraindicate the care their personal physician advocates. In fact, these objections run so deeply in some cases that some referring physicians have suggested, "If you do not protect my patients' information, I am not going to refer patients to your hospitals any longer."

This, again, is an area in which researchers can play a personal role in mitigating concerns. If researchers are prepared to engage in meaningful discussion with treating physicians about the value and benefit of proposed research and accept the expressed concerns, they can help to work around these potential barriers. For instance, rather than screening patient records without the knowledge of treating physicians and contacting patients themselves, researchers can work with physicians to identify potentially eligible patients and then ask the physicians to speak with them about the proposed research. This can alleviate both the potentially invasive feeling by patients of being contacted by a stranger for research as well as physicians' concerns that their patients may be recruited without the physicians' knowledge into research that they do not believe is commensurate with the care they provide.

Patient attitudes also play a key role in determining whether health information can or should be released for research purposes. Some patients are altruistic and have no difficulty sharing all their identifiable health information if it will better serve the community. Others are much more protective of their individual information because of fears over misuse, discrimination, or social stigma. Some patients are comfortable releasing some, but not all, of their health information for research purposes. However, although this could be a means of balancing privacy interests against

research interests, many researchers do not view this as an effective option because it potentially distorts the available data sources and could skew data results. Moreover, even in an electronic world, technical limitations can function as barriers to even this limited type of research access. As discussed above, many systems do not have built-in abilities to easily capture data in a format useful for research purposes. Additionally, many systems lack the functionality that would be necessary to allow a patient to partially opt out of disclosures for research purposes (e.g., portions related to mental health or substance abuse).

Unfortunately, most healthcare providers have no cost-effective way of protecting just limited portions of the patient record, even when individuals feels comfortable that the rest of their file could be used for research purposes. Eventually, we may get to a point where we can make such distinctions, but for now such requests put us in the untenable position as a Covered Entity of having to assume the burden and cost of basically pulling records or reports, reviewing eligibility criteria, and spending the necessary time to compile all this information to be used for research purposes. Under such circumstances, many administrators legitimately question what benefit these burdens provide to our patients and to our institutions. If researchers are willing to expend the time, efforts, and costs necessary to enhance these systems to better meet these needs, they can potentially go a long way toward increasing institutional support for research disclosures.

All of this is not to suggest that healthcare providers do not have any commitment to research at all. Healthcare providers recognize the value and the public good of research. They are committed to research, especially when it is consistent with their own mission or values or when there is a direct benefit to them. Obviously, however, they do not want it to interfere with patient care. They do not want it to be overly burdensome or costly, thereby detracting from the resources available for activities more directly related to patient care. They do not want it to interfere with their relationships with their physicians or the relationships between the patients and the physicians. In addition, of course, all healthcare providers have to be concerned about legal risks and compliance with applicable laws.

Recognizing the importance of research in furthering the practice of health care, and in improving society as a whole, MedStar has undertaken a number of different efforts to try to accommodate researchers in a fashion that balances our privacy concerns against the administrative burdens associated with research-related requests. One thing we have done is agree in some cases to effectively perform screening and recruiting activities on behalf of researchers. This includes screening participants—assuming there are no objections from patients or physicians—in order to obtain authorization on behalf of the researcher or to simply provide the subject with information about the research project and let him or her contact the researcher

directly. In theory, this avoids the accounting obligation, and could be more sensitive to some of our patients, but it requires time, training, and effort on our staff. In some cases, we have asked researchers for compensation to offset some of those costs. This is obviously not preferred by all researchers, but it is a step toward closer engagement between us, as a healthcare provider, and the research community as we work to foster coordinated EHR user organization evidence development work.

Another approach we have tried with limited success is to engage the researcher effectively as a business associate to handle all the screening, recruitment, and internal administrative processes that we have in place. This allows the researcher to recruit patients directly, but it avoids the accounting of disclosure obligations and shifts the burden to the researcher for cost. Depending on the HIPAA mechanism the researcher is using, the PHI may need to remain within our property as a Covered Entity. If the researcher does not obtain an authorization, the PHI would need to be returned or destroyed. This solution, unfortunately, is not appropriate or viable in all situations and, again, is not always palatable to researchers themselves.

Other data access options include Limited Dataset/Data Use Agreements. This option would generally permit researchers to have a limited set of identifiable health information, without a patient authorization and without the accounting of disclosures responsibility, but it still requires resources of the Covered Entity to create the Limited Dataset and to negotiate the Data Use Agreement with the researcher. Our experience has shown that this option currently has limited effectiveness for the majority of research conducted at our facilities because most of our research requests are for more complete, identifiable datasets. As a result, the Limited Dataset will not be a truly useful or viable option for us absent systems that can cost-effectively produce the data and absent an amendment to the Privacy Rule that greatly expands the number of identifiers available through this vehicle.

Going forward, we see several potential avenues for progress. We would like to see HIPAA amended to accommodate the needs of researchers while minimizing the burdens on Covered Entities. Eliminating the accounting disclosure obligations would go a long way toward reducing our costs and burdens. Expansion of the Limited Dataset concept could potentially assist both researchers and Covered Entities if the Covered Entity has systems that can cost-effectively produce data and the Limited Dataset vehicle is greatly expanded to include identifiers that would permit screening and recruitment activities.

In addition, as vendors and suppliers of our data systems and electronic medical records systems become more sophisticated in the potential applications of this information, the design of operational databases

and electronic records will allow us to more generally protect the patient information that needs to be protected due to applicable laws or commitments to our patients, while making available information that can and should be available for research purposes. With respect to the physician–patient relationship, continued work is necessary to build communication and trust between all parties, and opportunities exist to further educate treating physicians about research opportunities. With respect to technology, interoperable data exchange may ease some of the technological burdens we face and could result in greater access to health information by researchers, but the details and potential barriers associated with access to data exchanges remain uncertain and may require further legal clarifications. Perhaps most importantly, an increased awareness and sensitivity on the part of researchers to the requirements, burdens, and costs associated with healthcare providers' provision of information, and a willingness to share in those costs and burdens, can greatly aid in overcoming the obstacles that currently impede research efforts.

REFERENCES

CED (Committee for Economic Development). 2008. *Harnessing openness to transform American health care*. http://www.ced.org/docs/report/report_healthcare2007dcc.pdf (accessed August 21, 2008).

Westin, A. E. 2008. *How the public views privacy and health research: Results of a national survey commissioned by the Institute of Medicine Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule*. http://www.patientprivacyrights.org/site/DocServer/Westin_IOM_Srvy_Rept_2008_.doc?docID=3181 (accessed August 20, 2008).

6

Creating a Next-Generation Data Utility: Building Blocks and the Action Agenda

INTRODUCTION

The collective experience of presenters, workshop participants, planning committee, and Roundtable members offers many important insights on how the future architecture and policies of data systems can aid in leveraging health data to its fullest and best uses. Perspectives summarized in this chapter reflect on lessons from successes and failures across the healthcare industry to provide guidance for the development of next-generation applications and progress acceleration in the public health data agenda. Underlying the discussion summarized here is the principle that new data utilities will build on considerable past progress. Offered in this chapter are summaries of workshop presentations and a panel discussion that provide a mix of theoretical perspectives and specific ideas for practice to guide future work.

Building blocks for a next-generation public agenda were described by three presentations based on lessons learned about strategic priorities from past and ongoing work. Christopher Forrest, professor of pediatrics, senior vice president, and chief transformation officer at Children's Hospital of Philadelphia (CHOP), presents a CHOP initiative to transform the delivery of pediatric care and children's health through the power of data-driven decision making. Forrest describes CHOP's highly linked data system, which includes genomic, clinical, and environmental data used to support the organizational vision of transforming pediatric care, and discusses issues related to collaboration: developing cross-institutional relationships, providing the patient and family access to information, fostering provider–payer relation-

ships, working to reduce costs (both financial and nonfinancial), changing cultural assumptions, and improving communications. Brian Kelly, executive director of the Health & Sciences Division at Accenture, a global management consulting firm, details some challenges of managing and aggregating multiorganizational data and the associated influence of current privacy regulations on data activities, including practical challenges and the many entrenched and difficult-to-change systems for data aggregation. Guidance is needed on approaches to ensuring individual health data protection, questions of data ownership, on conveying the benefits of providing access to healthcare data through public advocacy initiatives. Finally, Eugene Steuerle, senior fellow at the Urban Institute, reviews current incentives to share health information that are at odds with positioning clinical data as a public good. Although the benefits of clinical data can be shared by all, the distribution of costs associated with collecting, storing, and analyzing the information are borne by few. In addition, the significant issues with privacy and confidentiality, bureaucratic policies, and providers and intermediaries further complicate structuring incentives, whether financial or otherwise, to share clinical data. Steuerle offers several suggestions on means of restructuring incentives associated with collecting and aggregating clinical data for healthcare improvement and suggests that consumers of health services may ultimately need to be the driving force behind changing current incentives to foster a more favorable approach to clinical data.

The chapter concludes with a summary discussion of six panelists, charged with moving the conversation around a next-generation data utility to an action agenda. The six panelists are Stephen Phurrough, director of the Coverage and Analysis Group at the Centers for Medicare & Medicaid Services (CMS); James Ostell, chief of the Engineering Branch of the National Center for Biotechnology Information (NCBI); John Lewin, chief executive officer of the American College of Cardiology (ACC); Evelyn Slater, senior vice president of Worldwide Policy at Pfizer; Janet Woodcock, deputy commissioner and chief medical officer of the Food and Drug Administration (FDA); Arthur Levin, cofounder of the Center for Medical Consumers; and session chair David Blumenthal, director of the Institute for Health Policy at Massachusetts General Hospital/Partners Health System. They discussed critical questions, including what decisions and actions are needed to advance access to and use of clinical data as a means of advancing learning and improving the value delivered in health care, and they offered perspectives on current activities and opportunity areas in the development of healthcare data resources. Current and emerging "what if?" opportunities to align policy through multiple stakeholder engagement are considered, along with implications of recent legislative initiatives. The identification of opportunities for enhanced coordination, investigation into aspects of policies on consent for data sharing and

research, and active translation of data insights to better engage the public emerged as important areas for future work.

BUILDING ON COLLABORATIVE MODELS

Christopher Forrest, M.D., Ph.D.
Senior Vice President, Chief Transformation Officer
Children's Hospital of Philadelphia

The Institute to Transform and Advance Children's Healthcare (iTACH) at Children's Hospital of Philadelphia is spearheading a novel effort to harness clinical, business, research, and public health information to improve children's health, make their health care more efficient, and transform the delivery system. The Institute has developed a data system that links the full spectrum of information about a child's health needs, from genomics to clinical to environmental data, in order to build out a vision of personalized pediatrics. This paper will provide an overview of this new approach to healthcare delivery and will assess issues related to collaborative relationships that are needed to realize a vision of personalized pediatrics, including forming linkages with multiple pediatric institutions, giving patients and families access to their data and obtaining information from them, and creating provider–payer collaborations.

In our organization I sit at the nexus of biomedical and applied informatics, genomics and other types of molecular diagnostics, healthcare management, quality and patient safety, and translational medicine—translating scientific discovery into clinical care using health information technology. One of the most urgent goals at Children's Hospital of Philadelphia is to transform ourselves into a data-driven healthcare organization. We view data as an essential asset that is just as important as the people in the organization, our financial reserves, and the buildings that we are building.

CHOP is the largest integrated pediatric network in the country, if not the world. We have 35 primary care practices and several specialty clinics located in 3 states, and have nearly 2 million encounters a year. We have an integrated specialty network, and staff both a 500-bed central hospital and a number of other community hospitals. We use electronic health records (EHRs), but we believe the EHR is just one piece of the health information technology needed for personalized pediatrics.

In the process of trying to become a data-driven organization, we have created a novel concept of personalized pediatrics, which we see as much broader than that of personalized medicine, which too often becomes conflated with pharmacogenomics. To build our concept of personalized pediatrics, we will need to create a number of collaborations that we are

either developing or need to develop—collaborations with other pediatric institutions, public institutions such as local government, payers, and patients and families.

Our concept of personalized pediatrics is one of CHOP's primary strategic initiatives. We think this is the future of health care—in fact, a new form of health care—and we are devoting substantial new investments to it. We believe all health care will become more data driven, much like other service industries. In so doing, we seek to go beyond conventional quality improvement, which tends to focus on process change and reliability of service provision. Personalized pediatrics is an approach to health care that customizes delivery of services to the individualized needs of children and adolescents. Conventional modes of treatment are based on caring for the "marginal patient" because so much of our medical evidence is based on average treatment response. Our model is predicated on giving care at the right time by the right person, in the right setting, minimizing waste, and shifting services from specialty to physician-focused and nurse-focused primary care, even at the home whenever possible. We believe that to be successful, we have to generate value for people, which translates into less time spent in health care, paying less out of pocket, and the ultimate goal, getting better more quickly.

To implement personalized pediatrics, we need to construct a biopsychoenvironmental profile for every patient (Figure 6-1). Data for this profile are

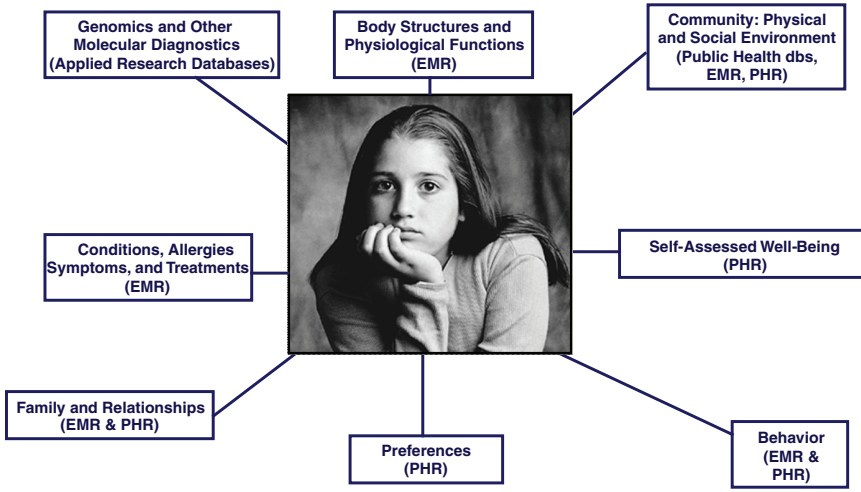


FIGURE 6-1 The biopsychoenvironmental profile and required data sources.
NOTE: dbs = database, EMR = electronic medical records, PHR = personal health records.
SOURCE: Reprinted with permission from the Children's Hospital of Philadelphia.

obtained from electronic medical records (EMRs), applied research databases, public health sources, and directly from patients and families using personal health records (PHRs). For now, genomics, gene expression, and other types of molecular diagnostics are collected as part of human subjects research. Within the next few years, however, we expect these types of data to be part of the portfolio of advanced diagnostic laboratory medicine and thus be available in the EHR. CHOP is building a patient portal into the electronic health record. It is not a full PHR, but at least it gives patients the ability to access their own data and input a limited set of information. The full biopsychoenvironmental profile, when fully available and used to improve care, will be a major advance in our ability to better personalize care, predict future health events, and ultimately prevent ill health.

Well-child care (i.e., preventive care services for children that are focused on optimizing health and development) has been the bedrock of pediatrics for years. Recommendations for the specific set of services are made according to age and sex, even though the risk for poor health and functional outcomes varies dramatically within age–sex groups. Where we are headed is toward more personalization of delivery of well-child care, according to factors such as the social complexity of the family and the risk of the child for early school failure. This type of information needs to be collected directly from parents in a uniform, structured way and incorporated into a biopsychoenvironmental profile to enable specific customization of preventive care. These measures of developmental and social risk will be used to produce scores that can be used to partition patients into "tiers" of need for preventive services based on their likelihood of poor future outcome.

The technology driver underpinning personalized pediatrics and improvement of outcomes for children is what we call the Pediatric Data Trust (PDT) (Figure 6-2). We use the word "trust" to convey two deliberate messages. One is the notion of a bank—your data are stored in a vault, they are going to stay there, they are not going to go to anybody else who may misuse the data, and the kinds of transfer we are engaged in are not to payers. The second connotation of trust is that patients can trust us with their information. Various types of data systems feed into this large data warehouse; electronic medical records are just one piece. In fact, to make personalized pediatrics work, data from the electronic medical record is insufficient; it needs to be greatly augmented with other data sources. We use a large vendor (Epic) to collect our EMR data, but to transform care we have to integrate multiple types of data and store them in a sophisticated relational database—the pediatric data trust. In addition, there is an applications layer integrated into the PDT that runs analytics and other kinds of algorithms to identify clinical decision support opportunities. The real technical challenge with which we struggle is this question: Once you have

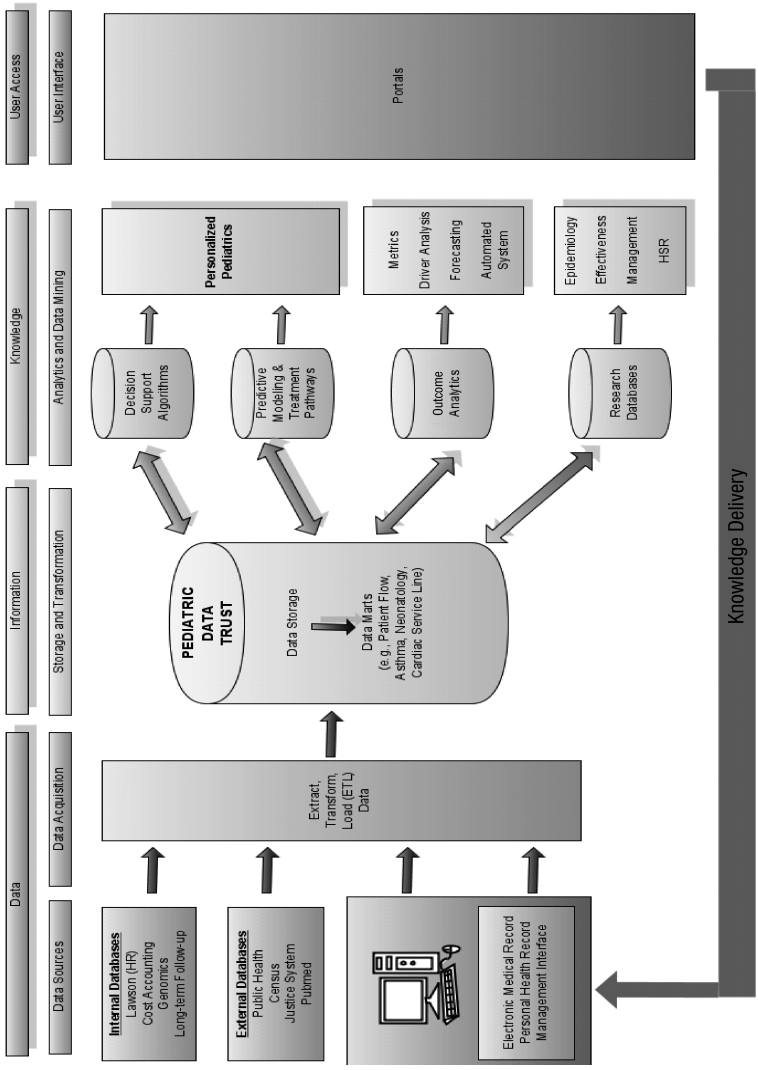


FIGURE 6-2 The Pediatric Data Trust.
SOURCE: Reprinted with permission from the Children's Hospital of Philadelphia.

identified a need and you want to personalize care, how do you take that information and transmit it in a timely way to the right decision maker, whether he or she is a patient, family member, clinician, or manager? It is one thing to build clinical decision support right into the electronic health record, but more advanced support—say, by linking genomics with patient preferences and clinical data—requires a kind of knowledge delivery system and creates technical hurdles that prove to be incredibly challenging.

CHOP recently established the Center for Applied Genomics, the largest genotyping facility of its type in the world devoted exclusively to understanding the genetics of pediatric disorders. The Center has conducted whole-genome analyses on more than 50,000 individuals, and in some cases their parents. When kids have a blood draw at a CHOP facility, our process calls for a research assistant to ask their parents whether they would like to have their child genotyped. The system is fully automated, using robots in the genetics lab to perform the assay on a platform capable of reading 500,000 single nucleotide polymorphisms (SNPs) (soon to be a million SNPs) for about $750 per test, a cost that is rapidly decreasing. Our goal is to identify new gene–disease associations, use that information to subclassify patients with a particular disorder, learn what treatments work best for each subclass, and then drive that information back into the clinical record. For example, let us assume that we have identified four genetic subclasses for patients with asthma. The Pediatric Data Trust is used to rapidly learn how alternative treatment pathways relate to those genetic subtypes and outcomes of care. We may learn that one type of inhaled corticosteroid is most effective for Subclass 1, while another is most effective for Subclass 2, and so on. Then if a patient with asthma has genetic Subclass 1, one set of order entry templates would be used, facilitating the usage of the "right" corticosteroid according to the genetic profile. In other words, we believe that the day when genomics and other molecular diagnostics information is embedded into the workflow of the physician—one of our ultimate goals with personalized pediatrics—will arrive in the near future.

A variety of types of collaborations will be important to the success of personalized pediatrics. Because so many conditions are uncommon in children, we will need to link databases across multiple pediatric healthcare organizations. Leadership for this type of clinical data linkage could be provided by the National Associations of Children's Hospitals and Related Institutions or the Child Health Corporation of America; both are professional organizations for children's hospitals and both currently collect administrative data from hospital discharge abstracts. Unfortunately, neither has yet embraced the collection of clinical data. Consequently, iTACH has been working with these partners and others to build the case for a national, even global, clinical database that would link the pediatric data trusts of multiple institutions. In the model we are developing, an

organization retains its own local pediatric data trust—not just the EMR, but also other kinds of biological and environmental data linked in a deidentified way. The data would be sent to a central repository, a national PDT. Standards will be critical here, some of which need to be developed (e.g., genomics). We believe an integrated PDT is going to be of incredible importance not only for benchmarking outcomes data, but also for mining information to identify the most effective treatments for patients with uncommon conditions, which is much of specialty pediatric practice. Our partner pediatric institutions are excited about this, and we are exploring ways to make this vision a reality. We know we are going to need better clinical data systems; administrative data from hospital discharge abstracts are not helpful for improving care for children. Large pediatric institutions will probably need to develop these systems themselves. We are not seeing leadership from the federal government, state governments, or any professional societies, at least in pediatrics.

Local government is the other partner that we need to make personalized pediatrics work. More than 250,000 kids are in CHOP's primary care network. With the data we have, we are able to geocode patients according to their place of residence. We can use this geographic information to identify and track epidemics, the leading edge of influenza, respiratory viruses, and infections. We also think we should start linking in pedestrian injury information, which, for example, we might obtain from police records. With such information in the EHR, a physician would know whether a child lived in a neighborhood where there is a high rate of pedestrian injuries and could customize preventive services and counseling around prevention of accidents for high-risk families.

We need information from the public sector, and we want to give information back to the public sector. There is a great deal of interest in Philadelphia on ways to address the childhood obesity epidemic. In our data trust, we have information that we would like to make available in some way to the local government. We can, for example, sort children in our database according to their body mass index and place of residence. We anticipate being able to look at these data over time and build predictive analytics as to which kids are more likely to be overweight. Right now the information is simply used for case finding; it needs to be linked with public health programs. Forming that linkage and getting the cooperation of local government to work with us has been very challenging, but it is definitely a direction of the future.

We also need to work with payers. We know about the care that children receive in our own network, but we have very limited information about the care they receive out of the network. Our local Blue Cross Blue Shield plan is interested in participating with us in a community-wide childhood chronic disease management program, say for asthma. With this type

of payer–provider collaboration, our organization will be able to receive from the payer healthcare, medication, and laboratory data on children who receive services outside of our network. All that information gives us a more complete picture to provide better quality care for asthma, improving children's health and keeping them out of the hospital. We are also exploring ways to work with other payer partners to develop a child-specific personal health record. If we really want to use data to improve kids' care, we believe there will be a need for a longitudinal personal health record, which will be the central repository of information for children. None of these programs will work without the support and participation of families. Our model of care is highly family centered. Everything we do is designed in partnership with families. We believe our privacy statement is quite good on *Health Insurance Portability and Accountability Act* (HIPAA) and Institutional Review Board issues. It does not, however, address issues about data mining, that gray zone between research and clinical care.

Our researchers strongly believe that any patient who walks through our doors is a research subject, that their data are the property of the institution and should be available to investigators for research. I am unsure whether our patients would agree with this position; we have never made this philosophy public nor have we had a public discussion on use of data for outcomes improvement and research. We also have not engaged in a discussion of personalized pediatrics with families. Although the notion of linking genomics all the way to the environment sounds good to us as providers, we do not have a sense about how this concept will be received by the public. That is a discussion we are going to need to have with our community. We will need to sort out whether we are going to require everybody who comes through our doors to approve the use of their data for quality improvement purposes, or whether we are going to make this optional.

In conclusion, personalized pediatrics is not only a critically important future for health care, it is where we are going to find real value for children and their families. We do not think better health achieved at lower costs is possible with process and quality improvement alone. Customizing care to the individual needs of a patient, tailored to a unique biopsychoenvironmental profile, will be necessary to truly transform child health. Personalized pediatrics focuses on outcomes, changes in health, and reductions in costs, both financial and nonfinancial. Three big challenges will be culture, communication, and collaborations. Changing the culture of our providers to collect data in a high-quality way is dramatically difficult. You can put a provider on the electronic health record, but getting the provider to enter accurate and valid data may take many more years. Communication about our intentions and having a dialogue about how to partner personalized pediatrics, particularly with families, is critical; it is something we must do

right now. To make this work, we have to make successful all the collaborations mentioned above.

TECHNICAL AND OPERATIONAL CHALLENGES

Brian Kelly, M.D.
Executive Director, Health and Life Sciences Division, Accenture

This paper provides an overview of many of the technical, operational, and organizational challenges faced when attempting to aggregate clinical data from multiple institutions to gain insights on clinical effectiveness or drug/device safety. Drawing on experiences from previous pilot projects and other work in this area, the paper will (1) provide an on-the-ground, real-life implementation perspective on the challenges of aggregating data from multiple sources for secondary use, and (2) discuss the impact of current privacy regulations. Issues of how current privacy rules inhibit the merging of data on individual patients for non-HIPAA-sanctioned public health use cases will be discussed.

Data sharing across institutions presents considerable technical and operational challenges. Part of this perspective comes from work done building a prototype for the Nationwide Health Information Network (NHIN), in which researchers tried to aggregate data from 15 different organizations in 4 states. The development of what Accenture had to do—from a technical architecture perspective and a business process perspective—to get people to agree to share that data is discussed. In addition, in just the past few months, we have worked with several organizations that are coming together to determine how we can begin to take data that have already been aggregated by some of the large data aggregators and to merge those data into very large datasets for purposes of monitoring drug safety adverse event signal detection. When you take data that have been aggregated and then try to aggregate them again, you have a volume of data sufficient to actually derive statistically significant findings, but such an avenue has its own set of challenges. The technical challenges and policy needs for this scale of data sharing are also explored. Topics discussed include critical features necessary to aggregate data for secondary use, challenges, opportunities, the limits on data sharing outside single care delivery systems, limits on the secondary use of non-HIPAA sanctioned data, and the need for education and advocacy for using data as a public utility.

Around the globe, payers, pharmabio companies, governments, and hospitals are working to improve health care. Data sharing is absolutely critical to all of these. The healthcare ecosystem is a $3 trillion business, but no one really shares data—for many reasons. One is that in many instances,

not sharing data gives them a competitive advantage. Sometimes, it is just too hard to share data because the standards to share that data are too different. Organizations around the world say they could be more effective if they had other organizations' data and could leverage that. Getting to that point is a wonderful vision, but major obstacles must be overcome.

Accenture's approach to facility data sharing for our NHIN prototype will illustrate what you really have to do to share and standardize data so that you can actually run secondary analysis on it and derive insight, which then can translate into improvements in care and outcomes. We had 15 separate hospitals. Each had its own registration system, lab system, and medication or pharmacy system, and in most instances those systems were not the same. Each hospital had completely unique platforms, with no common standards. Really, it is a data mess out there at the hospital level. Each of the 15 hospitals were aggregated into groups of 5, basically 1 group per state, and aggregated a subset of their data into a core data repository. The data aggregated were data critical to patient care, including demographic data, recent lab data, medications, allergies, and past medical conditions. Although we could have also pulled in other data, we elected to concentrate on these datasets because they were seen as core components to delivering clinical care.

We aggregated the data with an EMR view and a PHR view. Then we merged those data from those three regional areas into a central repository where we could perform secondary use. We made sure that the data were sent through messages that were standard HL7 version 3 messages. We spent probably most of our time mapping data elements to CPT4 codes, ICD9 codes, and to Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT) codes. We did a lot of heavy lifting on a small number of patients, showing that it is theoretically possible to do this.

This work is necessary to facilitate data sharing to produce aggregated data that would enable secondary use. The extensive data mapping we did is not necessary if one is focused on a single patient and does not intend to use the data as a tool to improve care. A doctor seeing a single patient, with lab results, medication charts, and so forth, can figure out how to take care of that patient. Although the physician cannot use EMR capabilities that reside in an EMR, such as decision support. if the data are not normalized, he or she could probably still do a fairly good job of taking care of that patient. To optimize care and secondary use, however, it is necessary to do the heavy lifting of putting the data into equivalent standards and terms.

It is important to recognize that systems that currently exist in hospitals are going to be there for many years. The average lifecycle of a typical hospital system is more than 20 years. People think these things are new systems and that they frequently change all over, but that is not the case. People make major capital investments in these systems over time, and so

their replacement is going to be incremental. It will be a generation before many hospital systems have turned over completely. For example, some of the best systems are in the Department of Veterans Affairs and Department of Defense. Their development started in the late 1980s and while they are clearly migrating forward, they are leveraging their legacy systems and have not changed that fast. To connect to these core systems and exchange data among institutions requires a very sophisticated approach to information governance because people do care about who can see their data and under what circumstances.

When you start doing these types of projects, one of the first discussions concerns who owns the data, and whether they will share those data. Even if we say the patient may own the data, in reality whoever puts the data into a database owns the data. In this case, hospitals own the data we are discussing, and if they do not believe there is a compelling reason for them to share their data, they will not do so. If their main EMR system has data on 100,000 patients, and some data requests pertain to only 50 of their patients, they are highly unlikely to share data on the 99,500 other patients in their database. You have to anticipate this in advance and map a strategy to pull only the 50 patients willing to participate in this data sharing. Thinking about this from a technology perspective, that is a much higher bar than to just take a flat file download of data tables and import it. This is the reality, however, of the technical components necessary for true data sharing to occur for the public good.

If data are not standardized, they cannot be used for secondary use in any meaningful way. That means you are not going to be able to do good public health surveillance, you are not going to be able to provide good care management and you are not going to be able to do clinical research. Whatever architecture you come up with had better be able to store the data, either where they sit or centrally, and should reflect ground rules set in advance about the manner in which people agreed to share data.

To obtain a perspective of what we need to think about, let us consider a model of three hospitals that wish to share data. The first challenge is that if you are going to pull data from each hospital system, it would be preferable to have a single, super-standardized interface, but it is not that way. Each hospital is likely to have a different lab system, a different radiology system, and a different demographic system, and you are going to have to pull all of them. Assume, too, that there is a central node where you are going to aggregate data, and then allow your patients to access PHR data, allow physicians to aggregate EMR data from that, or allow a researcher to look at data analytic tools for secondary use.

You will first need to develop some sort of filtering mechanism that pulls only data on people who have agreed to participate in this data sharing. First, therefore, you need to have a way of identifying a specific

person—you have to know, for example, that I am Brian Kelly and this is where I live, and that I am the same Brian Kelly who has data in hospitals A and B, but not the same Brian Kelly who has data in hospital C. You cannot pull that data across a hospital's firewall until you make that kind of identification—a hospital will not do random pulls of individual data based on the fact that a patient at the time of care signed a notification of privacy practices that said he will allow you to use his data for treatment, payment, and operations. That is the standard form. A key question, therefore, is what exactly is included in treatment, payment, and operations. That gets into a philosophical argument that most hospitals would rather not enter. Remember that health care is regulated at the state level, so while there is HIPAA, states can be much more restrictive than HIPAA, and in some instances they are. In our model for data sharing, therefore, the only way we could get this done—and we were doing this across four states—was to go out to patients and ask them whether they would allow us to use their data and participate in this prototype. When this becomes an outcomes and operational system and is being used for the care of patients, that restriction could potentially go away if a patient was going to rely on this tool as part of care, but for now you are still going to have to filter the patient data. For now, because patients are not getting care through this mechanism, you do not have a right to pull their data across, so you have to know whether the patient has opted to participate. We had to have each patient sign a consent form, which was essentially a notification of privacy practices allowing us to use their data for the purposes of this prototype. We then registered the patients centrally.

We then developed software tools that essentially had a small application that ran inside the hospital's firewall. Every day, when they went to send messages out, they would basically bounce the name of the patient up against this application, which would have a list of all of the patients who agreed to participate and would only filter through and then centrally store data for people with a signed consent. This is the kind of reality that will impede data sharing. Unless, for example, we change the standard notification of privacy practices to say the data can be used if they are deidentified for secondary use in clinical research, we will continue to have a lot of trouble aggregating data among various institutions.

Patient approvals are just one issue. Another problem is that each system usually has its own unique standards for data (or may have no standards). If a system could spit out an HL7 message, we could take that HL7 message, but if you know anything about messaging you know that if you have seen one HL7 message, you have seen one HL7 message. They are all different and you have to standardize them to a common type. We not only had to map the messages, but we had to map every term for data that we pulled over. We did that all in our customized application. Making this even

more complicated is the design of the security architecture—determining who can see what data, and making provisions, for example, for a patient who says that one doctor can see my data, but another cannot. There must also be robust auditing capabilities. We found it helpful to make an audit log available to patients to show them who had accessed their data.

Sharing data presents many challenges, and what we did in the pilot test is not scalable without policy change. We could not go to an area that has 300,000 patients and do one-by-one enrollment applications. The only way we as a nation are ever going to get to data sharing is for there to be a national policy discussion and for us to agree how it would be possible to modify things such as notification for privacy purposes to facilitate data sharing, and spell out the restrictions. To share data among different delivery organizations, there could be a different approach to notification for privacy purposes. That is one of the biggest policy areas that we as a nation have to grapple with and discuss.

However, in our model we saw that once we did the heavy lifting to get the data into the data warehouse, then beauty can really occur. If you have a set of completely normalized and standardized data on your population, you can do some really interesting analytics. That is how we can actually make things happen and, ultimately, transform health care.

Apart from the project described above, I have been involved in a project focused on aggregated datasets. The idea is that you could go to large organizations that have already done a good job of aggregating data and collect the 10 biggest datasets that exist in the world, put them into a database, and merge them successfully. Such datasets would be that much more powerful for drug signal detection, drug safety monitoring, and studying all kinds of related questions. The problem we encountered is similar to pooling data across different hospital databases. Data aggregators pooled their data with organizations with which they had business associate agreements about how the data were going to be used. The agreements did not allow for the data to be used for secondary use unless they were deidentified. So some issues are involved. The only way to aggregate datasets is when the data are totally deidentified, but then when you merge the data, you would not know how many times a particular patient was represented in the mix. So a key question is how we can develop a master patient index function that allows you to index so that you know that these patients are the same. This is a big issue. Quite honestly there is some benefit for the public good to aggregating these large data sources, if a way could be found to solve the deidentification questions. Possible solutions include initially registering patients with a unique identifier that could then be associated for research purposes. Currently, however, from an operational perspective, it is extremely difficult to get past that challenge.

Finally, there is a need for advocacy for using data as a public utility.

Particularly in the hospitals and research institutions I have been able to visit over the past few years, I have seen some that have been very innovative about proactively reaching out to their patients to educate them about how important clinical research is to patient care. Many of these organizations have actively gone out and started essentially marketing campaigns to educate patients and their families on how important it is to participate in clinical trials and related research endeavors. We need to do the same thing to begin to educate people on how important it is to be able to use data for secondary purposes. Such efforts would need to address all the security and privacy issues because these factors are currently the biggest barriers to using data to improve public health.

ECONOMIC INCENTIVES AND LEGAL ISSUES

Eugene Steuerle, M.A., M.S., Ph.D.
Senior Fellow, Urban Institute

If we wish to change behavior, then we must directly address incentives. The existing incentive structure in health care discourages information sharing, giving great weight to possible errors in protecting privacy relative to errors in failing to use existing information to improve public and individual health. In addition, incentives internal to the bureaucracy also discourage optimal use of information, even such items as merging already existing datasets. Because government now controls nearly three fifths of the health budget (if we count tax subsidies), it has a primary responsibility to improve these incentives. Some incentive changes are possible now, through reimbursement and payment systems. Others require examining the reward structure internal to the bureaucracy. In the end, however, the primary incentive needs to come from consumer demand—operating either directly on providers and insurers, or on the voters' elected representatives.

One does not have to be a genius to understand that incentives affect the extent of data sharing in health care. What are the incentives and barriers to providing this type of public good? In many cases the benefits from sharing clinical data and better use of clinical data are shared by everyone. Yet few individuals, insurers, doctors, or government workers gain by incurring more costs themselves. Accordingly, for solutions to data sharing, we have to examine and change the incentive structure.

The first incentive problem is that of the expected public good versus the potential private cost. One of the major barriers to moving forward here derives from the tension between privacy and confidentiality concerns and the goal of improving well-being through information sharing. From the privacy side, some people fear their data might be lost. Indeed, instances

of failure do occur—even on a grand scale, as when the Department of Veterans Affairs lost huge data files. So it is a real concern. Even if lost data were no problem, for their part, some individuals do not want to have their data shared, no matter what good sharing might do for the public. So there is this related fear of violating privacy as an individual right.

As one consequence, fairly significant threats of lawsuits often prevent some providers and vendors from taking actions that might serve the public good. But there is a large cost—the failure to improve the public good when in fact we know, we strongly know—that we can do it. We lack data sharing for individual care, we lack data sharing for an early warning system (e.g., through the Centers for Disease Control and Prevention [CDC] or other organizations), and we lack data sharing that would assist in finding cures or better treatments for various health problems and diseases. Researchers would probably agree that the costs are high, although I am not sure the public is totally convinced. In the end, we have to engage the public in these issues. That is the first incentive problem.

The second incentive problem is the lack of bureaucratic incentives to share datasets or for agencies to allow datasets under their purview to be shared. Imagine what we could do, for instance, if we could just merge some of our Medicare data, which are pretty significant, with other federal data. We have only barely begun to merge Medicare data with Social Security data, even though at one time those functions were housed in the same agency. Notwithstanding goodwill among many public servants who want to serve the public, there are strong disincentives in the bureaucracy to share and use data. One such disincentive is the possibility of bad publicity if something went wrong or if the information revealed failed policy. For many agencies avoiding bad publicity is a mark of success. You may never have a lot of friends if you are the agency privacy lawyer, for instance, and the bureaucratic instinct is to minimize your enemies.

To make matters worse, our democratic systems almost always work to put more demands on our public servants than they can possibly meet. If you are sitting in the bureaucracy, you cannot even get done what needs to be done—it is often difficult to accommodate someone who adds another task, such as merging datasets, even if it might provide some public good. Also, as anyone who worked in the government knows, there are calls every year to reduce spending. Research and statistics have been relatively easy budget functions to cut; from the politicians' viewpoint, costs are diffuse and in the distant future—long after the next election. The key question, therefore, is how we can add incentives into the bureaucracy to reward people for undertaking publicly beneficial actions such as enhancing data sharing and helping to ensure that researchers somehow have access to those data.

A third incentive problem revolves around providers and intermediaries. An insurance company really does not have an incentive to do

anything more than to serve as an intermediary. It has little incentive to provide for the public good. Although companies do have some incentives to reduce relative costs, many of the public good issues we are talking about do not provide any gain for an insurance company itself. For the insurance industry as a whole, moreover, there is a perverse disincentive at play; for example, if cancer is cured through use of shared administrative or clinical data, insurance payments for the industry as a whole might be reduced.

Incentives for doctors are mixed, too. Knowing that a certain percentage of any population will inherently be below average, for example, why would a physician want to encourage relative comparisons of his or her success versus those of others if that is one result of data sharing? Some doctors also voice the fear that data sharing via EHRs will even give more information to insurers, making it easier for them to decide not to pay for elements of care.

From the doctors' or hospitals' perspective, development of better data may enhance the power of lawyers to sue. Data from some studies suggest that autopsies find that a large percentage of patients who die have been misdiagnosed at least to some extent—although often not enough to cause the death. Nonetheless, you can see the potential for lawsuits if these data were openly available.

For solutions to some of these issues, we have to find ways to change the incentive structure. Several possibilities are listed here, with the hope that readers would add to this list.

Government has a lot of leverage, not just because it cares about the public good. Government spends a lot of money on health care—nearly $11,000 per household. Factoring in tax subsidies with the direct subsidies, it provides nearly 60 percent of the financing of health care. The point is that government is the primary payer and player.

Medicare could pay more for e-prescribed drugs, less for those not e-prescribed (expenditure neutral). It could pay differentially for lab tests put into electronic form for sharing with patients or the CDC. It could pay for electronic filing of information on diagnoses and treatment.

Consider a condition such as autism, where we really seem to be at a loss in terms of the information we need. Few families are unaffected by autism somewhere among their relatives, and it is safe to guess that friends or relatives of children and adults with autism would consider the privacy risks modest relative to potential gains from developing a better data system aimed at discovering solutions. Through Medicaid, and programs such as those that provide education for the disadvantaged, federal and state governments are primary players here.

Certainly government could provide more incentives for participation

in clinical trials; one of the Roundtable themes is to engage and inform the public in evidence development strategies.

What should we expect from government efforts to change the incentive structure behind its payments? Often it does not really want to set the ultimate standard. Consider EHRs. A key set of questions again revolves around how government can appropriately set incentives for the development and use of such records. Because we have not solved the privacy issue completely, and likely never will, any system would probably have to give people some ability to bow out. But that cost might very well be worth the gains from more electronic filing of information on diagnosis and treatment, for instance. No one change in incentives can be expected to be perfect or get us completely where we want to be, but we can ratchet up the benefits from this type of information development and sharing.

People working in the bureaucracy currently get little reward for fostering data sharing. However, we do know that bureaucracies respond to incentives. Outside the healthcare field, for instance, we saw that welfare reform took hold when governors changed incentives for their head welfare officers. These officers were told that their success was going to be measured not by how much money they brought into the state from a federal government match, but by the number of people they got off welfare. Whether or not that was the right thing to do, it dramatically changed the entire welfare debate. Thus, even before formal legislated welfare reform, governors dramatically changed the entire operation of the welfare bureaucracy. Another example was provided by the United Kingdom's recent efforts to reduce child poverty. The Prime Minister and Chancellor of the Exchequer (or head of the Finance Ministry) set as a target a zero poverty rate for children. Although that is probably not an attainable target, once they made the goal public, the whole government started changing the way it did things to try to come closer to the target.

Thus, the dynamics of a bureaucracy can be changed. Suppose as only a minor example that the head of a U.S. department decided to measure the number of research projects developed with each dataset. (That may sound like a trivial and fairly impure measure of success, but it is better than none.) As another example, he or she could ask for common standards and protocols to be developed for data sharing. As it now stands, often every time one wants to share data using some new dataset, the process requires running the gauntlet of a whole bureaucratic layer of decision making—including the confidentiality officer, typically a lawyer, who has every incentive to say "no." Changing incentives within the bureaucracy often has to come from the top.

Finally, it is not clear that we will ever get enough information development and data sharing by looking only to doctors, insurers, and hospitals to somehow "do the right thing." The good sought is for individuals, not

for those serving them. Ultimately, then, consumer demand probably has to drive the system. One need not be passive about it; such demand can be fostered. Imagine, for example, a Department of Health and Human Services Secretary who would go around the country and make the public case for e-prescriptions and for electronic reporting of lab tests, showing how such efforts could make individual health care more effective. The goal would be to encourage individuals, in turn, to demand improvements from their providers as a means of ultimately achieving better health care. Of course, the initial effort may not instantaneously result in datasets ideal for clinical research, but it would be a major step toward developing better data for those purposes, as well as better care for the individual.

There are other ways to foster demand, of course—in the example of families with an interest in autism, they would likely be quite willing to share data on the family.

In sum, there is no one-size-fits-all answer to the development of better shared clinical data. Incentives, however, can consistently be improved over time. I encourage all interested parties to add to the examples I presented here and to think rigorously about how to improve the incentive structure surrounding the development and use of data to achieve better health outcomes.

THE ACTION AGENDA

Engaging the spectrum of stakeholders working in healthcare data initiatives will be critical to the development of improved data generation, access, and evidence development. Opportunities to align policy developments or draw on synergies within organizations, initiatives, and advancements will serve to push the frontier of data collection and be used to drive the development of next-generation healthcare research and delivery. Workshop discussions included perspectives on the current and developing uses of healthcare data for insight, and presenters addressed opportunities to evaluate policies impacting the public good, security, and privacy aspects of data. Manuscripts in this chapter, as in previous ones, profile advancements and perspectives that might encourage the frame shifting associated with developing clinical data as a public good.

Summarized here are the discussions of a workshop panel featuring perspectives from government providers, researchers, and regulators; healthcare professional organizations; pharmaceutical manufacturers; patients; and consumers. Panelist comments focus on how the ideas and opportunities presented at the workshop might form the basis for an action agenda that will realize the full potential of developing initiatives and implications of new legislation on data initiatives. Approaches discussed aim to align

policy development with improved data access and evidence development, and to engage all stakeholders addressed.

Opportunities for Enhanced Coordination

Several workshop participants, both speakers and attendees, identified the need for a more coordinated approach to managing current data sources, ongoing research, and possible future synergies between the two and across the spectrum of stakeholders. Addressing the growing inefficiencies might yield greater health insights at less expense. There are needs for better coordination, a more standardized approach to healthcare data, and a more concerted effort to bridge research and data resource gaps through cross-organizational projects.

Standardize Approaches to Data Collection

Volumes of patient research and care data are generated every day, but the type of information stored can vary greatly between platforms. One participant provided an example of several research groups that collected different information about diabetes from various cohorts of patients and found correlations with different genetic markers. In addition to using different genetic typing platforms with different readouts, some correlations overlapped and some were different. The researchers were able to merge specific aspects of the disparate data and create a common dataset that both confirmed and disallowed some individual findings. Demonstrating the power of aggregating and standardizing data, this case also identified several additional genes of interest only after the data had been pooled.

Similar examples prompted the National Institutes of Health (NIH) to require researchers to pool data collected under NIH grants for the benefit of other investigators. The NIH created the database of Genotypes and Phenotypes (dbGaP) to archive and distribute the results of studies, including genomewide association studies, medical sequencing, and molecular diagnostic assays, that have investigated the interaction of genotype and phenotype. The advent of high-throughput, cost-effective methods for genotyping and sequencing has provided powerful tools that allow for the generation of the massive amount of genotypic data required to make these analyses possible. dbGaP incorporates phenotype data collected in different studies into a single common pool so that the data can be available to all researchers. Dozens of studies are now in the database, which by the end of 2008 was predicted to hold data from more than 100,000 individuals and tens of thousands of measured attributes. Hundreds of researchers have already begun using the resource. There is also a movement on the part of the major scientific and medical journals to require deposition accession

numbers when they publish the types of studies alluded to above, in the same way that they require for DNA sequence data. The publications recognize that investigators need to review the data that informed the paper in order to confirm or deny a paper's conclusions. Other accession numbers are also used when people take data out of a database, reanalyze them, and then publish their analysis.

Lewin provided additional comments on data standardization from the perspective of the American College of Cardiology. The new IC^3 Program (Improving Continuous Cardiac Care) is the first office-based registry designed to provide physicians with the most current, nationally recognized best practices for cardiac care. Approximately 2,400 U.S. hospitals participate in ACC registries, voluntarily contributing data and benchmarking their own performance against peer institutions. The ACC is working to standardize the data collected to be able to measure gaps in performance and adherence to guidelines, with an ultimate goal of being able to teach others how to fill those gaps and thus create a cycle of continuous quality improvement. Mandates from Medicare and the states have pushed hospitals to use the ACC registries, but there is room for wider adoption. ACC is working to eliminate barriers in the use of its registries, such as the need for standardization in the way data are collected, the expense of collecting needed data, and the lack of clinical decision support processes built into electronic health records (EHRs).

Lewin also provided insight into the ACC's National Cardiovascular Data Registry (NCDR), which is designed to improve the quality of cardiovascular patient care by providing information, knowledge, and tools; benchmarks for quality improvement; updated programs for quality assurance; platforms for outcomes research; and solutions for postmarket surveillance. The FDA's Critical Path Initiative is an attempt to combine research data from various clinical trials in different ways and to learn more than what was learned in a particular research program, Woodcock said. The FDA has also been working with the Clinical Data Interchange Standards Consortium to try to standardize as many data elements as possible. But Pfizer's Slater pointed out that significant roadblocks remain in the effective sharing of clinical data across multiple organizations and platforms. For example, among the trials posted on www.clinicaltrials.gov, for example, shared information can be incomplete, duplicative, and hard to search, and nomenclature is not always standardized. Slater suggested that addressing these issues might improve data resources.

HIV research is an example of how data sharing has worked effectively. Woodcock elaborated on data from multiple trials on CD4 count and viral load. The data were reviewed extensively by multiple bodies, and the FDA was able to advance the field by developing quantitative measures that could be used to guide therapy and drug development. Recently, too, the

FDA has done multiple analyses across drug classes, some of which have been publicized extensively. One set of analyses was of selective serotonin reuptake inhibitors (SSRIs). The FDA conducted analyses across all the SSRI clinical trials to look for suicidality and excess of suicidality, and found and publicized some evidence of those factors in various age groups. Similar work was conducted on epilepsy drugs. In effect, looking across multiple programs facilitates learning much more than in the past.

Facilitate Cross-Organizational Efforts

As session chair Blumenthal observed, one context for the panelists' remarks is that the environment for clinical data is much more distributive than ever. In terms of policy, that phenomenon overrides traditional instincts that policy makers bring to bear, which would be to assume that solutions would come by deciding what local, state, and federal governments could do. In a distributed environment, however, such an approach might be framed too narrowly. For example, if the conversation focuses on the personal health record and engages consumers directly, that policy environment is very different from one that would be relatively easy to address by a centralized authority. At the same time, the federal government is a big stakeholder and player in the collection of health-related data. However, the environment surrounding data differ across departments and agencies—the NIH, for example, can focus on promoting data sharing and has a broad mandate for data collection sharing, whereas the Centers for Medicare & Medicaid Services (CMS) operates in a much more restrictive environment.

Medicare collects data in each of the four parts of its program: A, B, C, and D. Collected data are used as the basis of paying claims. Data are collected to help improve healthcare quality, for payment purposes, to develop pay-for-performance qualitative information, noted CMS's Phurrough. Another set of data collection programs are in Medicare demonstration projects that explore a variety of issues and generally examine how different payment systems may affect outcomes versus clinical issues. Data are also collected to develop evidence.

Discoveries at the molecular level provide unprecedented insight into the mechanisms of human disease. Now that powerful genome-wide molecular methods are being applied to populations of individuals, the necessity of broad data sharing in molecular biology and molecular genetics is being brought to clinical and large cohort studies. This has resulted in the NIH Genome Wide Association Study policy for data sharing, and the new database at the NCBI called dbGaP, reported the NCBI's Ostell. In the course of collecting and distributing terabytes of data, the branch that Ostell oversees, the Information Engineering Branch, has wrestled with questions

concerning which data are worth centralizing versus which should continue to be distributed. For example, the commonality of molecular data might drive the desire to have all related information in one data pool so that a researcher could search all the data comprehensively—perhaps without a specific goal in mind—which in turn could lead to the kind of serendipitous connection that is fundamental to the nature of discovery. At the same time, however, efforts need to be tilted toward collecting only those pieces of data that make sense in a universal way.

The ACC supports the use of national patient identifiers that would enable the tracking of an individual's overall health continuum across organizations, while preserving patient privacy and bolstering longitudinal studies. Wider adoption of data sharing via registries is within reach and should be encouraged because ultimately it would result in better overall health care. However, strategies need to be developed and implemented that foster systems of care versus development of data collection mechanisms specific to a single hospital. Thus, the ACC is interested in collaborating with other medical specialties, EHR vendors, the government, insurers, employers, and other interested parties.

Providing additional perspective on opportunities for cross-organizational collaboration, Pfizer's Slater indicated that the pharmaceutical industry is interested in helping to ensure the widespread availability of data to support research at the point of patient care and care at the point of research. In the pursuit of that goal, the industry is interested in ensuring the alignment of data quality, accessibility, integrity, and comprehensiveness. Moreover, the *Food and Drug Administration Amendments Act* (FDAAA) of 2007 will ensure the posting of more clinical summary data. FDA Predicate Rules and FDA 21 C.F.R. Part 11, the *Health Insurance Portability and Accountability Act*, and other legislation that are part of FDA and European Medicines Agency standards also help to ensure public access to clinical data. In addition, many companies voluntarily share their postmarketing safety data, Periodic Safety Update Reports, and clinical trials summary data. Data are shared, for example, via the International Federation of Pharmaceutical Manufacturers and Associations, the Clinical Trials Disclosure Portal, the Clinical Research Information Exchange, and Sermo, a community of 60,000 physicians who exchange clinical insights, make observations, and review cases in real time. Woodcock pointed out that scientists have a strong interest in being able to combine data from various research studies, and thus the FDA is very active in promoting data sharing.

The Center for Medical Consumers' Levin offered general guidance on the subject. He discussed the implicit tensions, such as the tension between protecting private information and sharing it across organizations for specific reasons. Tensions also exist between the federal and state approaches; between the public and private sectors; and between individual and public

health. These tensions play out in the legislative process, which ultimately may facilitate cross-organization information sharing.

Policies on Data Sharing and Research

In addition to identifying opportunities for enhanced coordination to support evidence generation and application, many workshop attendees indicated a broad sense of need to examine the scope of data sharing and research policies, including those that encourage the development of data as a public good and that aim to engage the complexities of patient consent to enhance the research process.

Alignment of Incentives

One carrot that Medicare has developed is that it has required the delivery of clinical data beyond the typical claims data as a provision for payment for certain services. For example, a few years ago the system required additional clinical information for the insertion of implantable defibrillators. Such an approach can provide significant amounts of information if we can learn how to meet the challenge of what we can do with data that have been collected, and merge those data with other sources of data so that data collection can inform clinical practice. The ACC supports investing in rigorous measurement programs, advocating for government endorsements of a limited number of data collection programs, allowing professional societies to help providers meet mandated reporting requirements, and implementing systematic change designed to engage physicians and track meaningful measures.

An influx of regulations as well as an acknowledged need for transparency is prompting the appearance in the public domain of product development and testing data. Nonetheless, we must be careful to ensure data standards, integrity, and appropriate, individualized interpretation. The Center for Medical Consumers, a nonprofit advocacy organization, was founded in 1976 to provide access to accurate, science-based information so that consumers could participate more meaningfully in medical decisions that often have profound effects on their health. The Center's Levin believes government has a role to play in regulating health care. But as health information technology (HIT) and health information exchange (HIE) move forward, some are concerned that legislation might be pushing us backward, not forward.

Another issue is that data sharing is, in essence, a social contract between individuals and researchers who want to use their data. Patients are told that sharing data will eventually lead to better care. But perhaps

patients do not hear enough about how that is supposed to happen; they are interested in how researchers will use the data to improve care.

Evaluation of Consent Requirements

Establishing the appropriate consent mechanisms for using patient data for research purposes beyond the initial purpose or intention was another area of interest. The current authorities under which the Medicare system collects data are fairly narrow. Medicare has clear legislative authority to collect data for purposes of payment, but also works under legislative requirements to limit that information to the minimum amount necessary to pay that claim. A second authority has to do with quality; Medicare has authority from Congress to collect information necessary to improve healthcare outcomes. Additionally, there is a narrow authority to conduct demonstration projects, solely for the purpose of testing different payment systems. Given those limits, the agency has had to be somewhat innovative, for example, by linking some clinical data collections to coverage of particular technologies.

Woodcock also provided insight into a blood pressure study that had involved automated monitoring. Data from tens of thousands of patients were combined, creating a virtual control group that did not involve time from patients and healthcare systems. In such pooling of data, the FDA has had to address issues of going beyond the intent of the original trials and consent, an issue that will be addressed continuously as data resources grow.

Translation for Engaging the Public

One challenge identified in the workshop was the need to engage the public in data-driven health research through translation and interpretation of the individual and societal benefits of individual health data. Many organizations, including the IOM, pursue opportunities to engage patients and consumers on the issues surrounding healthcare data. Lewin described ongoing efforts to ensure that ACC guidelines, performance measures, and technology appropriateness criteria are adopted in clinical care, where they can benefit individual patients. Although most guidelines are currently available on paper, the vision is to have clinical decision support integrated into EHRs to stimulate conversations between providers and patients. The ACC's NCDR was designed to research solutions for postmarket surveillance. NCDR strives to provide standardized data that are relevant, credible, timely, and actionable and to represent real-life outcomes that help providers improve care and help meet consumer and patient demands for quality care.

Workshop attendees also indicated that information should be given in language that is more user friendly for patients. Once data are in the public domain, it becomes difficult to control quality assurance and the accuracy with which the information is translated to patients without an acceptable format for providing data summaries.

The FDA plans to build a distributed network for pharmacovigilance. The Sentinel Network would be established to integrate, collect, analyze, and disseminate medical product (e.g., human drugs, biologics, and medical devices) safety information to healthcare practitioners and patients at the point of care. Required under the FDAAA, the Network is currently the focus of discussions with many stakeholders about how best to proceed. One idea is to build a distributed network in which data stay with the data owners, but remain accessible to others.

From a consumer perspective, Levin said there should be a requirement that the data collector do something specific with the data collected, and an evaluation should follow. Such a mechanism can help researchers build trust with patients—by demonstrating the value, convenience, or pay-off of sharing data.

AREAS FOR FOLLOW-UP

Through session discussions and panelist commentaries, several possible opportunities were discussed to continue progress in the development of clinical data as a public good. The following areas were highlighted.

Data Sharing

A workgroup of EHR users and data-storing organizations should be convened to investigate the possibilities of gaining deeper insights through the use of clinical, research, and administrative data across multiple organizations. To start, the workgroup could address issues concerning elements of quality and outcomes, payment and payment system outcomes, and performance and accountability.

Develop Research Methodologies

Bench, clinical, and health services research data have the inherent possibility of being beneficial to multiple investigations beyond initial research intents. Methods are needed to "mix and match" datasets to uncover higher levels of information.

Evaluate Current Policy at All Levels

Because the costs of conducting studies continue to increase, it has become even more important to leverage limited resources to extract the greatest possible benefit. Similar efforts, such as those of dbGaP, have been successful, and lessons learned might be applied in other areas. Although the federal government is one of the biggest participants in the healthcare market, the level of data sharing varies by agency and/or department. Generating national identifier numbers that can be tracked through longitudinal studies from government databases may provide significant insight into the care and care processes provided.

Transparency in Communication

With the increase in data available to measure performance and outcomes at organizational and individual provider levels, many opportunities exist to teach the public how their health information helps themselves and others. Participants stressed the importance of encouraging transparency, while protecting confidentiality, across all levels of the healthcare system. Engaging health consumers in understanding the benefits of aggregated health data through public–private partnerships was suggested as a mechanism for facilitating the conversation with the public.

Increasing understanding of the complexities of the healthcare data environment will involve efforts from all stakeholders. This session of the workshop identified several areas for additional investigation in an effort to stimulate action. Evaluating the opportunities for enhanced coordination through data standardization and multiple-facility efforts was suggested as a key area of interest by workshop participants. Also highlighted was the possibility of examining current policies to align incentives that might encourage data sharing and evaluate patient and organizational consent requirements for participating in data-sharing initiatives. A final area for additional investigation was ensuring that benefits of healthcare data aggregation, sharing, and research are translated for and subsequently communicated to the public.

7

Engaging the Public

INTRODUCTION

Because the public is the key stakeholder in clinical data, efforts to improve data collection and use depend on public understanding and engagement. In addition, a key aspect of facilitating the dialogue between provider and patient is engaging the public in understanding how healthcare data can support individualized disease prevention and health promotion. Too often, however, patient knowledge or expectations are not aligned with the current reality and abilities of data sharing and analysis. This chapter summarizes discussion from the final workshop session, aimed at understanding public concerns and generating discussion on possible means through which the public and their concerns could be addressed. Participants provided insights into public perceptions on clinical care data for research and queried the public's interest in health data-supported guidance and information. Aspects of the public's extent of use of clinical data were also examined.

As noted in previous chapters, the public needs to be fully engaged in understanding, generating, and applying clinical data. While the public is aware of some issues and has basic knowledge of the arena, many participants noted that more needs to be done to bring the public fully into the conversation about clinical data, and the public needs to be better informed about what clinical data are and how those data can help them. Papers in this chapter explore how the public is now engaged, and what advances (technical, communication, demonstration of value) are needed to expand their participation in the next-generation public data utility.

Most members of the public do not have a sophisticated understanding of how their clinical data move within or outside an often fragmented system. But, notes Alison Rein, senior manager at AcademyHealth, they do want more and better access to health information about them, their family, and their peers; in particular, they have a strong interest in electronic access to personal health information. In part the public is not as fully engaged in the clinical data utility as it could be because much of the current activity takes place out of the public eye, with data residing largely in the private sector, where commercial interests and other factors inhibit sharing. Unless we overcome market obstacles related to sequestering data for proprietary interests and the technical obstacles related to individual identification authorization and related issues, Rein asserts that a full demonstration of the clinical data value proposition for consumers, both individually and collectively, is not possible. Efforts are also needed to help the public develop a deeper appreciation for research as a public good. As a strategy to build public understanding, Rein supports increased transparency of reporting data that are meaningful to the public and enhancing the coordination and development of registries. In addition, she suggests that policies geared toward a national network for researchers, clinicians, public health professionals—and the public—might generate additional information that could, in turn, inform the public.

Highlighting an opportunity for patients and the public to share individual information, Courtney Hudson, chief executive officer and founder of EmergingMed, describes how giving disease-specific populations increased access to information helps them access clinical trial information and supports a patient-centered search for treatment options based on a patient-generated profile. Although providing access to information is important, Hudson also discusses challenges associated with employing data for secondary uses such as services and systems improvement. Throughout EmergingMed's 8-year history, Hudson has found that patients and families are interested in informed decision making, and when they receive this benefit, they are generally willing to provide privacy-protected information in return. By contributing to a database and service that identifies similar patients, individuals can anonymously help the public in making decisions. Creating transparency and consequently building trust are also vital when working with individual health information, particularly when engaging the public in broad concepts such as evidence-based medicine. A key distinction in considering the patient's point of view might be to view clinical data utilities in terms of patient-driven solutions instead of system-driven solutions.

Involving patients in the learning process through personal records and portals is important to patient-centered health care. Today, more personal health data are created and analyzed than ever; however, the degree to

which the data are distributed across delivery, payer, research, and manufacturing organizations continues to increase. Jim Karkanias, partner and senior director of applied research and technology at Microsoft Corporation, notes that with the growth in patient-controlled information, stored in a variety of places, there is a growing opportunity associated with shifting control away from organizations and moving toward a more shared responsibility for information management. For all parties in the healthcare ecosystem, there are benefits associated with easily accessible data aggregated from multiple locations into one patient-controlled location. The connections built based on data aggregation have the power to transform healthcare decision making and improve clinical outcomes. Microsoft's HealthVault, one of several consumer-centered applications, provides a platform through which the patient determines the depth and breadth of information stored and shared about them.

GENERATING PUBLIC INTEREST IN A PUBLIC GOOD

Alison Rein, M.S.
Senior Manager, AcademyHealth

In many respects, the greatest challenge associated with establishing a medical care data system to serve the public is that such data largely reside in the private sector, where commercial interests and other factors limit sharing. This paradigm has benefited discrete entities, but fails to fully serve the public health interests of the broader U.S. population or promote awareness of how health information can improve clinical decision making for individual treatment. Ideally, the public would express considerable interest in understanding the importance of health information; however, the limitations of current data systems severely inhibit demonstration of the value proposition for consumers—both individually and collectively. This commentary aims to identify some issues for consideration in order to affect public awareness and perception of health data and their use for generating greater public good. The overview will provide what is known about this domain from the public's perspective and discuss some assumptions and attitudes that may impede progress. Finally, examples will highlight what we might learn from others, and some possible strategies for generating public interest and engagement.

A good place to begin is with public perception of the status quo. The public generally has a limited understanding of the extent to which their clinical data move and are shared—largely, at this point, for payment—within our fragmented healthcare system. The public assumes that clinical information is accessed and maintained by people and organizations who

need it to provide care, yet likely knows little about who (both within and outside the healthcare system) actually accesses the information and for what purpose. Although most covered health providers must give notice of privacy practices to patients at the first medical encounter (usually, a notice in which the content is specified by the *Health Insurance Portability and Accountability Act,* or HIPAA, waiver) (Health Information Privacy, 2009), patients likely have little or no understanding of the scope and requirements of the law or what it means for them. Even within the research and operations communities, there seems to be a lack of agreement on the extent of HIPAA coverage. If those in the field disagree or misunderstand elements of HIPAA, it is no wonder that the public should also be confused. Furthermore, there are regulations and statutory requirements beyond the scope of HIPAA (e.g., state laws) about which the public knows even less. Probably even state legislators could not enumerate the variety of different protections and state laws that exist.

Unfortunately, it often takes a chronic illness or significant health event for individuals to appreciate the paradox associated with accessing the volume of clinical data maintained on their behalf by various healthcare entities. Significant clinical, financial, and demographic information can be accessed by numerous parties for a variety of legitimate reasons, yet patients themselves have little or no access to the same information. Researchers wishing to use patient data for public health research are similarly challenged, as many of these data are considered proprietary and in the private domain. An important issue on multiple levels, it will require public education, outreach, and, more importantly, demonstration of value to reestablish the public's baseline understanding of current uses of health data and to encourage broader support for its application to public-sector research.

Though many point to other industries as examples of successful migrations from paper-based to electronic systems for record maintenance, health care is different in that there is a public expectation of trust and privacy between providers and patients. Unlike grocery store affinity cards, there is more substance to the provider–patient relationship than between the consumer and the supermarket. Additionally, with health information sharing, the potential for irrevocable harm is real. Some believe health care must simply learn from and catch up to other markets; however, this seems short sighted and insufficient given public expectations.

One need only to look at some recent examples of data-sharing practices from other markets to see how such behavior, albeit legal, would be undesirable in the healthcare context. These behaviors include the practices of private entities that sell granular-level cellular and landline phone records of individual consumers; websites that, unknown to individuals, coordinate information from multiple sources and display or sell the aggregated

personal and behavioral data. The extent to which the public knows or cares about such practices is not entirely clear. Some recent experiences suggest that once consumers are aware, tolerance for such practices is low. Facebook, for example, experimented with a model of open information disclosure and received immediate negative feedback from users. In response, the company changed its policy and reverted to a more data-protective approach. By and large, people believe that decisions about how and by whom their personal clinical information can be used (especially by non-healthcare providers) should somehow incorporate individual input or preferences.

Furthermore, there is a growing recognition that many consumers would like and need increased access to their own personal and family health information. Given the challenges inherent in navigating multiple provider settings and the inefficiencies of paper-based records, there is high interest in electronic access to personal health information. Although HIPAA provides for access to paper copies via formal request and a fee, this likely will not satisfy the on-demand expectation of today's consumers, who have access to other information in real time and free of charge. What is interesting and challenging about the current migration from paper to electronic health records is that no one wants to permit unrestricted access, least of all the individual patient. This type of situation prompts numerous conversations about appropriate policies and incentives both for instituting adequate consumer protections and for deriving optimal value from the emerging wealth of clinical data. This essentially describes a social contract in which both parties, having complete information, actually obtain a desired outcome.

Two major obstacles to realizing the goal of electronic health information access are (1) the current fragmented delivery system, and (2) the free market for data. One promise of electronic health information exchange is that it could help—virtually bind—the various data-holding entities, thereby creating a more integrated care delivery and data management system. However, unless obstacles related to sequestering of data for proprietary interests, as well as technical obstacles related to individual identification authorization and related issues, are addressed, we will not enjoy the electronic exchange of health information that could serve both the individual and public interests.

Until mechanisms are established through regulation or otherwise to compel provider and other healthcare institutions to share data appropriately, leveraging clinical data for the public good will be significantly constrained. The current free-market approach propagates business models that encourage continued segregation of valuable clinical information. What is needed is a more thoughtful approach for determining the types of business models that should be encouraged to flourish because they

could actually stimulate appropriate data sharing. Consider, for example, the tethered personal health record (PHR) system offered by an insurer or provider who refuses to provide data electronically to patients. The rationale supporting this decision is a perceived susceptibility to the competitive advantage derived from owning the data. Another example can be found in many health information exchanges, which sell aggregated patient data to commercial data vendors. Such models enable entities to charge a premium for access to valuable data, which ultimately precludes use by academic and public-sector research communities unable or unwilling to pay for access.

As we work to transition to a new health information paradigm, do we really want to extend models that perpetuate data silos, or do we want to envision another type of business model that promotes sharing and broader dissemination? In this respect, the public and research sectors share a common problem and could benefit from aligning forces and identifying common research areas of interest.

Exacerbating this state of affairs is the fact that patients have very limited access to or control over how their health information is used and shared, and most institutions have little motivation to share patient information for public research purposes. Ironically, patients likely would support use of their clinical information for public research. A recent Markle Foundation survey shows a fairly strong willingness among consumers to share their clinical information for noncommercial research studies (assuming appropriate controls and safeguards).[1] The challenge is that most institutions with access to this information do not readily share it for such purposes. Although individuals might be willing to share data in a manner that contributes to the public good, they do not generally have the means to ensure that their clinical data are leveraged for such purposes. In large part this is because most people do not have ready access to their clinical and other health information electronically on demand.

Another issue involves the extent to which the public perceives research and its products as a public good. There has been some discussion of how clinical discovery and science are valued by the public. An entity called Research America regularly conducts surveys through *PARADE* magazine, which has 74 million readers.[2] The survey consistently finds that people believe there is value in clinical research and in supporting research both to sustain our competitive advantage and to advance our own well-being. An entire market exists to promote all the products that emerge from this field (e.g., new treatments for cancer). We do not experience that at all on the public health level. Little, if any, discussion occurs on how registries could

[1] See http://www.markle.org/resources/press_center/press_releases/2006/press_release_12062006.php (accessed August 31, 2010).

[2] See http://www.researchamerica.org/parade_poll (accessed August 31, 2010).

identify disease trends and track outcomes among patients on alternative treatments or how clinical data collection at the community level could support public health and wellness promotion activities. These other types of research lag significantly in the level of marketing behind them; this highlights the need for more energy and focus to these areas.

Building Public Support

To build public support, the value of sharing clinical information must be demonstrated. Only limited public demonstrations of this kind are currently used, and even fewer that illustrate the potential impact that might be meaningful to individuals. For example, people might be more inclined to contribute clinical information to a county or other community registry if doing so would increase the ability of local health authorities to identify disease trends or anomalies that could indicate an environmental or other health concern (e.g., asthma related to air pollution, or cancer related to toxic waste). Such an effort could prove highly effective at galvanizing the public and public officials to make policy changes necessary for improving environmental health. The National Academy of Sciences estimates that 25 percent of developmental diseases, such as cerebral palsy, autism, and mental retardation, are caused by environmental factors. The American Cancer Society estimates that one-third of cancer deaths could be prevented through lifestyle and environmental changes, but we are not feeding enough information back to people to highlight existing opportunities and where these environmental factors influence different areas. We do not supply the public with adequate tools needed to make the case for change.

One possible approach to demonstrating the value of research as a public good could be to expand reporting of limited, but meaningful, clinical health data to public health entities. The New York City Department of Health's effort to have all labs report A1C[3] data to the public health department has experienced feedback from those with privacy concerns. The move does, however, serve as an interesting example because the program's goal is to engage the public health department actively in prevention measures. Furthermore, the program maintains a controlled approach, and only limited data from the public are required. Hopefully what they do with the information can end up having a meaningful clinical benefit for New York's diabetic community. One likely application for city officials will be to evaluate the impact of the city's recent ban on trans-fats on A1C data.

Another opportunity to demonstrate the value of research is through the enhancement and expansion of clinical data registries. Today, many registries offer limited accessibility and do not necessarily collect data in

[3] HbA1C is a measure of average glucose levels.

a way that can be useful and meaningful to the public. There is a clear opportunity as we construct registries and future databases to think about public health priorities that could be addressed, as well as ways to stimulate more interest and engagement from the public and produce more value for the consumer.

Multiple organizations have recommended the development of a nationwide health tracking network to help identify, track, and prevent known causes of death (environmental, occupational, and lifestyle/behavioral) and poor outcomes. It could yield information such as geographic and ethnic incidence and prevalence of diseases and inform the public health community, providers, policy makers, and consumers. Currently, the Department of Health and Human Services operates 200 separate data systems, but has little or no means of coordinating across the range of systems. Similar to registries, these systems are not designed to track major causes of death and disability in the United States, even though the data may be compelling and important to the public. In many of these cases, the public is required to contribute information, but may not benefit (or not be aware of the benefit) from these systems. Regardless of what form data efforts take, it is vital to commit more energy and resources to collection, integration, and interpretation of health data in order to better inform policy makers and the public.

Moving forward, we should not underestimate the power and potential of good information in the right hands—including the public. Ideally, we would work to ensure that information resources generated from use of patient clinical data are relevant, valuable, and appropriate. Finally, we should not pit public interests against those of the research community. There is more alignment of interest than there is divergence.

IMPLICATIONS OF "PATIENTS LIKE ME" DATABASES

Courtney Hudson, M.B.A.
Cofounder, EmergingMed

The longstanding tension between an individual's desire for personalized health information and the population's interest in healthcare research is exacerbated by scientific advances such as molecular profiling, information sharing on the web, and modern data management tools. Both the public and private sectors are struggling to navigate this logistically challenging landscape to gain medical insights (and sometimes to monetize these insights). Patient-focused clinical trial information services created in the past decade provide a unique perspective on a patient's view of healthcare research at both the individual and the population level. This paper provides an overview of more than 115,000 cancer patients' responses to a

paradigm that blends personalized information services with a shared public platform. It discusses how these services have addressed the intersection of an individual's need for information, access, and transparency with the U.S. healthcare system's desire for population-based research and data sharing in light of modern data management and data-sharing capabilities.

EmergingMed is an 8-year-old company founded to help cancer patients obtain access to clinical trials as part of their search for treatment options. It is a year-2000 paradigm that allows patients to remain in control: Internet-based searching supplemented with telephone-based support on request. We allow patients to create their own medical record or profile, and they compare their profile to a structured database of cancer clinical trials in the United States and Canada. Today we have a coded database of 10,000 cancer clinical trials, structured by eligibility criteria. Patients coming to our system can, in a matter of minutes, figure out on a preliminary basis whether they are a match for clinical trials in North America.

Our experience gives us a unique perspective to talk to the need, the absolute mandate, to have a patient focus going forward as we think about medical information and its uses for research. Patients in this country are supportive of mining clinical databases for the public good. Overwhelmingly, they believe that it is already happening and they would be alarmed to know that it does not happen.

More than 115,000 cancer patients have created profiles in our system. We have been on the phone with 35,000 patients over the past 7 years. On average, we engage in five follow-up calls with each person so we hear their stories. We hear the successes they have, the challenges they face, and the barriers they encounter. Most of these conversations start similarly—with a discussion that corrects patient misperceptions, allays patient fears, or corrects mistaken assumptions patients have about the healthcare system, what is going to come next, or their personal situation. All of that has been valuable.

When you start with a patient-focused system and then use derivative information so that you have a secondary database, you end up with a very exciting—but complex—situation. Your first priority is serving patients so they can make an informed treatment decision. Through this service, we gather information on the healthcare process, access to care, and outcomes. Day in and day out, our service has to be valuable to patients. The flip side is that you want to do secondary analysis of the information—data mining—not only to run the service better but also to gather insights to improve the system. It is challenging for a system to do two things at once—to process transactions and to run data analysis—and the two systems often collide.

We have had the benefit of being able to create our system from scratch. We are on an Oracle platform. Every module we have, including the call

center software, was newly developed to be integrated so it is a seamlessly integrated software platform. We did not have to contend with a legacy system. Nonetheless, 8 years into our program, the challenges are growing.

Our experience gives us perspectives that merit sharing. First, we find that overwhelmingly, patients are looking for data to make informed treatment decisions. Withholding information available in the public domain would be unconscionable. Patients expect us to share information that might affect their decisions. As we think about ways to use public datasets and aggregate them, it would be extremely difficult ethically to justify policies that withhold information from individuals when they could act on the information. Legacy systems can make it difficult to make information actionable to patients, but nonetheless we need to remember that there is something of a quid pro quo or social contract here. Generally, the patients and families that we work with seem to believe, in essence, that if you will help me with information to make my decisions, I am happy to have you use it to learn things in general, so long as you protect my privacy. As a small company and a private company, one violation of privacy and we are out of business. We do not have an option of violating privacy, nor does anybody else in the private sector.

We believe our patients' feedback is applicable across the healthcare spectrum. The basic set of assumptions that patients bring start with an assumption that "my doctor will automatically tell me everything that I need to know either about treatment options or what is going on in clinical research and the clinical trial world." With cancer, however, there are some 400 to 500 drugs in development at any day, and their design, use, and targets are constantly shifting. At any one time there are some 10,000 clinical trials for cancer in the United States alone, and trials are continuously opening and closing. It is simply not feasible for any one person, such as the doctor, to have full command of all of that information.

To take this one step further, if you are a physician who is not literate with computers, you and your patient may have very different expectations about what information you should have at your fingertips. You are also going to have a different set of expectations about what knowledge you should be able to acquire.

One reason we formed this company and this search mechanism was to create a single step (completing a single questionnaire) through which patients could determine which cancer clinical trials are relevant for a specific situation. Patients should be able to determine that in 5 minutes; it should not take 8 hours of random Internet searching through 10,000 clinical trials. Fortunately, finding such information quickly is possible now with modern data management tools.

Imagine physician committees, hospital providers, and healthcare providers making decisions about storing and mining health information if

they are not aware of tools that are available or how data can be stored or structured. Last year an oncology-based physician group was evaluating electronic health records systems for the oncology community. Unfortunately, many physicians and leaders of these committees did not know the difference between free text and structured data and had no concept of long-term implications for data mining and data sharing. The implications for the decision making are striking. The healthcare information technology companies invited to assist this consortium were struck by the relative computer illiteracy of the assembled leadership.

Today patients and the public in general have much more access to some modern, sophisticated data tools outside the healthcare system. Yet we continue to allow healthcare providers to be in the dark—or to be the patient's sole information provider. The conflict between the two, or the discrepancy between what patients expect and understand and what healthcare providers know and understand, only gets wider. That in turn creates a divide whereby data-savvy healthcare professionals become heroes to their patients, and physicians without those skills not only run the risk of weakening the doctor–patient relationship but also of risk providing less than state-of-the-art healthcare services.

One of the key factors in providing education about clinical trials and cancer research is a focus on educating patients about their diagnosis, their stage, and their treatment history. This paradigm or rubric is identical to the one used by the medical community to assess patients and make treatment recommendations. We find that it remains nearly impossible to educate a patient about a clinical trial if he or she assumes that all cancers and all treatments are the same. If a person does not understand that breast cancer is treated differently from colorectal cancer and uses different research, he or she is not going to understand the specific questions needed for a clinical trial search. If a breast cancer patient does not understand there is a difference between hormone receptor-positive and hormone receptor-negative cancer, she is not going to understand why tests are being done, why treatments are being recommended, or which clinical trials may be an option. We have found that the more one identifies key personal decision points and those key learning disconnects, the faster the patient learns.

Transparency and trust are an absolute requirement in a healthcare system that mines electronic data for health research purposes. The more transparent the system, the more quickly you get a patient's trust. One of the decisions we made in designing our service was to start with the patient's perspective. We looked at where patients search for information, knowing that patients search in many places. We created a single system that would serve many types of organizations, including advocacy groups and cancer centers. We operate services for the state of Florida as well as for pharmaceutical companies. We run services for each of them from the

same platform—one dataset, one process, one call center. We were guided by a clear view of the ethical mandate to provide patients with an unbiased, transparent system to help them make informed decisions. We knew we needed to have a conversation with each patient because clinical trial information was prone to misinterpretation and confusion. Hence, customizing the education for each patient was vital.

Virtually every client of ours has some misconception about clinical trials, but nobody misunderstands everything. Most patients have one or two points for which they need clarification, and then they continue their search knowing that the process is confidential and that their information will remain private. We operate with the same standards across all organizations and we never share a patient's information with third parties. We do share aggregate, deidentified information. We obtain explicit permission from patients up front. We have found that of those people who complete a profile and find a match, 90 percent give permission for us to call them. We are able to do long-term follow-up and tracking through outbound calling. We use each call to get permission for the next call, which also gives us a chance to obtain permission for another use of the data as we go along. It becomes a partnership over time. Patients can opt out at any time, but most do not. Some patients ask us to keep their information in our system, but not call them—they tell us they will call us back when they are ready to use the service again—and of course we respect that. We have stored their data in the meantime, and when we do reconnect with such patients, we pick up where we left off.

Running our program in scale has become cost-effective over time and enables us to track outcomes. We know who enrolls in a study. We know who does not. We know why they drop out of the process along the way. It becomes a rich collection of data. It also lets us provide feedback to stakeholders.

We operate as an independent intermediary. The system requires that the patients refer themselves to trial sites, to actually consider a clinical trial. We do not share patient information with clinical trial sites, but patients can come back to us throughout the process if they need help with logistical information or definitions as the process continues. We are navigators in that respect, but we do not become part of the informed consent process. Consent continues as it always has, at the institution, under the control of ethics committees and Institutional Review Boards.

The informed consent process remains somewhat problematic. It is designed to ensure that the patient is informed about a trial at a site. However, this process does not require that the patient be made aware of other clinical trial options. Competing trials might be available across the street or across town, but the informed consent process does not require this disclosure to the patient. In view of the pervasive fear of being treated like

a guinea pig, we remain concerned that withholding options from patients does, in fact, mean that we are treating them like guinea pigs.

Informed decision making requires full disclosure from the healthcare system. We remain convinced that patients have the right to specific and general medical insights that might impact their treatment decisions and their ability to make fully informed decisions.

IMPLICATIONS OF PERSONAL HEALTH RECORDS

Jim Karkanias, Partner and Senior Director
Applied Research and Technology, Microsoft

Health care is a complex and challenging environment. Massive increases in medical information, in part through the Internet, are making health care one of the most significant hot spots for technology innovation today. Clearly medicine suffers from an information management problem. Control of this information will eventually shift, from a top-down doctor-to-patient model to one in which there is mutual control. For physicians the information control issue is about aggregating data within and across provider organizations. For consumers it is about aggregating data across all of their sources of health data. Ultimately these views will connect for informed health decisions and better clinical outcomes. Today we have more personal health data than ever; however, these data are dispersed over a variety of facilities, providers, and even our own monitoring devices and home computers.

Microsoft is working with our partners to address gaps in the healthcare data management system, both from an enterprise and a consumer standpoint, to enable a more connected, informed, and collaborative healthcare ecosystem. A consumer health platform with specialized health search capabilities is the first application/service and the first step in our strategy that centers on delivering a platform that puts users in control of their information so they can access it, store it, and use it however and wherever they want it.

Microsoft recognizes that sensitive health information should be protected by strong policies and clear operating standards. In consultation with consumer privacy experts, we developed and implemented industry-leading Health Privacy Commitments for ourselves and stringent privacy principles for solution providers, developing on the Microsoft HealthVault platform. We are committed to making a difference in health care, and we firmly believe in software's ability to make a positive impact on the healthcare ecosystem worldwide. However, no one company can resolve health and data issues alone. Transforming health care is a complex problem with no

easy answer and certainly no quick-fix solution. We are taking the long view in our approach—it will be a marathon versus a sprint.

In terms of data sharing, we are all quite familiar with the Gordian knot of problems with data use, privacy, and the associated tensions. Meetings such as this one sponsored by the Institute of Medicine (IOM) help us begin to identify the elements of answers. That reminds me of a quote from William Gibson, the science-fiction author, who said: "The future is already here. It is just not evenly distributed."

Microsoft has engaged for the long haul in the process to improve health around the world and connect communities for positive health outcomes. The ecosystem in which Microsoft's HealthVault operates is meant to be a comprehensive platform in the center of this environment. It is named HealthVault because it serves as a vault for personal health data. It is meant to focus on the consumer.

We have an illustration that provides some detail about HealthVault (Figure 7-1). It covers every detail of what goes on in the hospital, what

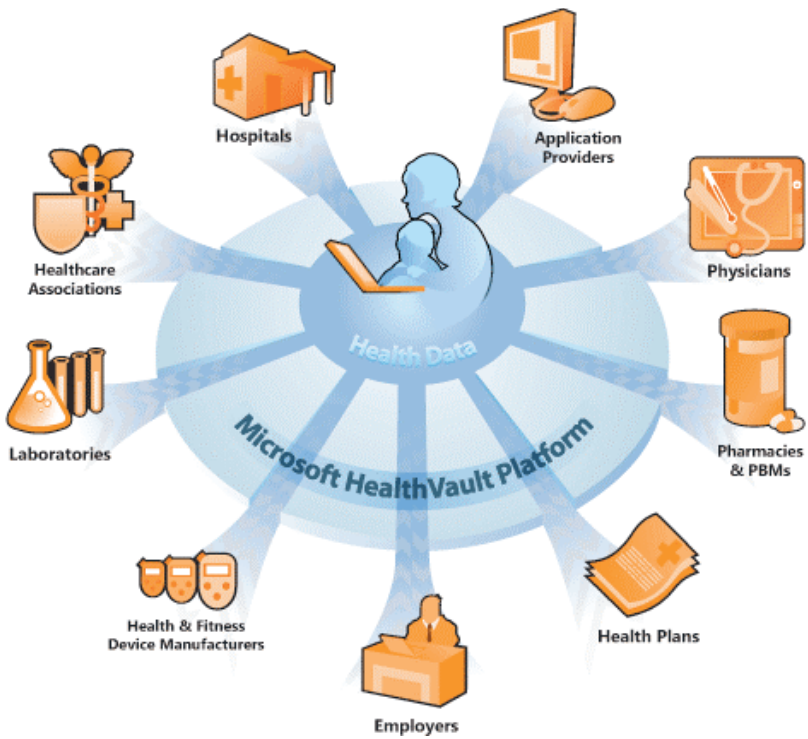


FIGURE 7-1 The HealthVault ecosystem.
NOTE: Reprinted with permission from Microsoft.

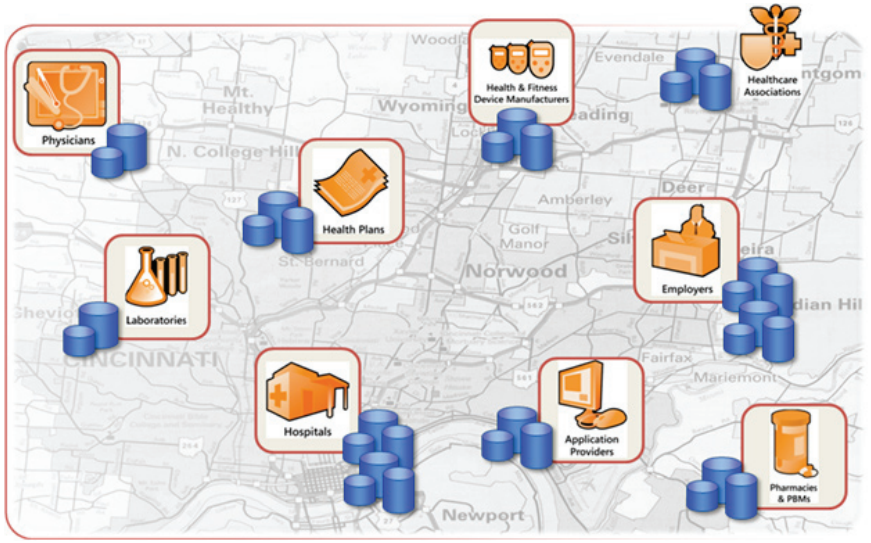


FIGURE 7-2 Silos of health information.
NOTE: Reprinted with permission from Microsoft.

might go on in the wellness side of the equation, what an employer might need to see, drug fulfillment requirements, and all the various medical devices in a patient's life. An important aspect of HealthVault is that it helps to collect the data that one might want to store and share with others. The notion here is that the consumer is in charge.

We first shipped HealthVault last year. The year before that we launched a product named Amalga, an enterprise hospital environment software that is meant to automate all the various activities in a hospital. (MedStar actually helped to develop that; we produced the software framework.) Obviously, many sources of information are in the system already. The problem is that too many sources of information in the system are silos that individually provide various important aspects (Figure 7-2). Integrating all of these silos is the key. There are many ways to achieve that—an overly centralized solution is just as difficult as not having a solution at all; a democratic-style solution is a different model; and so on. The question about addressing this as an information platform needs to be discussed.

The Consumer at the Center of the Equation

If we accept the premise that consumers are at the center of this equation, and that the information, activities, and processes that they are wor-

ried about in their healthcare journey all affect them anyway, they are a logical aggregator for this information, if only they could have a platform that would allow for that aggregation to occur. That is what HealthVault is. These various silos, through devices or through applications, allow data to flow into an environment that the consumer controls and owns. These data are the consumer's data.

HealthVault is an application platform. We do not propose to build these applications, but rather want to create an ecosystem of integrated software vendors to develop applications and devices in this environment. Many partners are already hard at work creating a set of services to support this activity. The folks who have put together these kinds of things in the past have often seen their efforts stymied by having to create all the vertical infrastructure to integrate across the various silos, creating firewalls or filters, or all of these things that our platform will help provide. The notion is that you can copy data from various sources and aggregate them for yourself in your HealthVault account as a consumer.

It is important to stress that HealthVault itself is not a PHR. It is the environment that allows a PHR to exist to be created from those pieces of data. Other providers of wellness data, such as body mass index calculations, or condition management devices such as glucometers that generate data and flow it into the system, will have their own applications that leverage the platform.

Other applications may also benefit the consumer. For example, a weight management application may benefit from a connection to an EMR or to cell phone data that measures location over time and correlates with pollen counts to which a person with asthma may have been exposed.

We do not intend to be the entity that stores this data, but rather see our role as facilitating the flow of information. PayPal, the online payment service owned by eBay, is an interesting analogy in the way it enables transactions to occur and accelerates all the convergence that we recognize must happen in this environment, such as accessing funds without revealing credit card numbers. So just as PayPal accelerates the storing and sharing of financial data, HealthVault accelerates the storing and sharing of health data. All of the complexity of dealing with that data securely and privately acknowledges the regulations that are in place, or connects to pharma, or the clinical trial provider, or the recruiters, or those who provide advice about a condition. All of that difficulty is handled by the platform, and only a relatively small amount of effort compared today's development activities is necessary to create that functionality.

Built into this platform is the notion of privacy and security. You cannot add that later; it has to be part of the environment. It needs to be adhere to industry standards. We are not charging for it—it is a free platform meant to enable and activate the paradigm shift we think can occur

when placing the consumer at the center of the equation. The consumer is highly motivated to focus on these pieces of information and aggregate them. We want to enable that. Appropriately, therefore, privacy and security are integral parts. We took a lot of time to determine how to make these factors as solid as possible given state-of-the-art technology. Again, the consumer is in complete control of the data. There is no way that that cannot be the case.

We think there is a notion of a family health manager, typically the mom who is worried about keeping her family safe. A lot of this activity goes on already; it is just done manually. The scenarios can be segmented further along a continuum. The complexity of the application increases, as does the user's engagement. Consumer applications fall into three primary segments: primary prevention, secondary prevention, and acute care. Patients who are in the primary prevention wellness side of the equation, where the applications are fairly straightforward, are engaged according to their own self-motivated needs. They are maintaining their health and wellness. Then we move to the more complex patient, who might be hospitalized or be living with a chronic disease. Finally, we see uses in acute care, among hospitals, group practices, and physicians.

Today we are in a reactive model. In terms of the economics of a health-related event, showing up at a physician's office has been the first step. A cost is associated with that event, and the cost spikes each time things get worse. We believe patients would prefer to be proactive in managing their health, and that is the model we are trying to create. They would avoid repeated costly visits and, instead, have daily in-home monitoring to allow for proactive measures to be taken to the extent they can be detected. This might play out in the use of a spirometer to encourage patients to inflate the lungs and thus avoid pneumonia, or some other device to manage a condition, and making that data available to the clinicians the patient authorizes. If that kind of approach can drive proactive management of care, how much better has the system become by definition? You can extend these to other scenarios. What if you had devices that were much more advanced and could detect a condition or other event and issue an alert? Again, that is all possible with this new technology.

REFERENCE

HHS (Department of Health and Human Services). *Health Information Privacy*. http://www.hhs.gov/ocr/privacy/hipaa/understanding/consumers/noticepp.html (accessed October 12 2009).

8

Clinical Data as the Basic Staple of Health Learning: Ideas for Action

INTRODUCTION

The availability of timely and reliable evidence to guide healthcare decisions depends substantially on the quality and accessibility of the data on which to base findings and conclusions. Information about the results of diagnostic and treatment interventions is collected in multiple forms by many institutions for different reasons and audiences—providers, patients, insurers, manufacturers, health researchers, and public agencies. These medical care data represent a vital resource for improving insight and action for more effective treatment. With the increasing technical potential for aggregation and sharing of data while ensuring confidentiality, the prospects are at hand for powerful and unprecedented tools for data analysis and determination of the circumstances under which medical interventions work best, and for whom. Although many challenges exist to the use of such data—coding discrepancies, platform incompatibilities, patient protection tools—as evident throughout this publication, practical approaches can be developed for most. The most significant challenge may be the barrier to data access and the restrictions in treating clinical outcome data as a proprietary commodity. This chapter summarizes the workshop discussions on principles, opportunities, strategies, and projects as possible follow-up actions for Roundtable engagement.

COMMON THEMES

Across the 2 days of presentations and discussion, several common themes—listed in Box 8-1 and elaborated below—emerged as issues for particular attention in accelerating progress in the availability and use of clinical data for new insights.

- ***Clarity on the basic principles of clinical data stewardship.*** The starting point for expanded access and use of clinical data for knowledge development is agreement on some of the fundamental notions to guide the activities for all individuals and organizations with responsibility for managing clinical data. Workshop participants repeatedly mentioned the need for consensus on approaches to such issues as data structure, standards, reporting requirements, quality assurance, timeliness, deidentification or security measures, access, and use procedures—all of which will determine the pace and nature of evidence development.
- ***Incentives for real-time use of clinical data in evidence development.*** Current barriers to the real-time use of clinical data for new knowledge discussed at the workshop ranged from regulatory and commercial issues to cost and quality issues. Participants suggested the need for a dedicated program of activities, incentives, and strategies to improve the methods and approaches, their testing and demonstration, the cooperative decision making on

BOX 8-1
Workshop Common Themes

- Clarity on the basic principles of clinical data stewardship.
- Incentives for real-time use of clinical data in evidence development.
- Transparency to the patient when data are applied for research.
- Addressing the market failure for expanding electronic health records (EHR).
- Personal records and portals that center patients in the learning process.
- Coordinated EHR user organization evidence development work.
- The business case for expanded data sharing in a distributed network.
- Assuring publicly funded data are used for the public benefit.
- Broader semantic strategies for data mining.
- Public engagement in evidence development strategies.

priorities and programs, and the collective approach to regulatory barriers.

- ***Transparency to the patient when data are applied for research.*** Patient acceptance is key to use of clinical data for knowledge development, and patient engagement and control are key to acceptance. In this respect, clarity to individual patients on the structure, risks, and benefits of access to data for knowledge development was noted by participants as particularly important. Patient confidence and system accountability may be enhanced through transparent notification and audit processes in which patients are informed of when and by whom their information has been accessed for knowledge development.
- ***Addressing the market failure for expanding electronic health records.*** Currently, market incentives are inadequate to bring about the expansion of use of electronic health records necessary to make the point of care a locus for the development, sharing, and application of knowledge about what works best for individual patients. Shortfalls noted by participants included demand by providers or patients that is not sufficient to counter the expense to small organizations; competing platforms and asynchronous reporting requirements that work against their utility for broad quality and outcome determinations; and the reality that even the larger payers—apart from government—do not possess the critical mass necessary to drive broader scale applicability and complementarity. It will likely take a deeper, more directed, and better coordinated strategy involving Medicare leadership to foster such changes.
- ***Personal records and portals that center patients in the learning process.*** Patient demand could be instrumental in spreading the availability of electronic health records for improving patient care and knowledge development. Such demand will depend upon much greater patient access to, comfort with, and regular use of programs that allow either the maintenance of personal electronic health records or access through a dedicated portal to their provider-maintained electronic medical record. As noted during the workshop, many consumer-oriented products currently under development give patients and consumers more active roles in managing personal clinical information, and they may help to demonstrate value in the speed and ease of personal access to the information, to better accommodate patient preference in care, and to foster a partnership spirit conducive to the broader EHR application.
- ***Coordinated EHR user organization evidence development work.*** The development of a vehicle to enhance collaboration among larger

EHR users of different vendors was raised during the workshop as a means to accelerate the emergence of more standardized agreements and approaches to integrating and sharing data across multiple platforms, common query strategies, virtual data warehousing rules and strategies, relational standards, and engagement of ways to reduce misperceptions on regulatory compliance issues.

- ***The business case for expanded data sharing in a distributed network.*** Demonstrating the net benefits of data sharing could promote its use. Benefits suggested by participants included cost savings or avoidance from facilitated feedback to providers on quality and outcomes; quick, continuous improvement information; and improved management, coordination, and assessment of patient care.
- ***Using publicly funded data for the public benefit.*** Federal and state funds that support medical care, as well as support insights into medical care through clinical research grant funding, are the source of substantial clinical data; yet, many participants observed that these resources are not yet effectively applied to the generation of new knowledge.
- ***Broader semantic strategies for data mining.*** Platform incompatibilities for clinical data substantially limit the spread of electronic health records and their use for knowledge development. Yet discussion identified strategies using alternative semantic approaches for mining clinical data for health insights, which may warrant dedicated cooperative efforts to develop and apply them.
- ***Public engagement in evidence development strategies.*** Generating a base of support for and shared emphasis on developing a healthcare ecosystem in which all stakeholders play a contributory role was noted by many participants as important for progress. Ultimately, the public will determine the broad acceptance and applicability of clinical data for knowledge development, underscoring the importance of keeping the public closely involved and informed on all relevant activities to use clinical data to generate new knowledge.

ISSUES AND OPPORTUNITIES

Workshop discussions touch on various issues and opportunities to improve the clinical data utility. These related to notions summarized below on clinical data stewardship, clinical data infrastructure, incentives for data sharing, creating the next generation data utilities and models, creating next generation data policies, and engaging the public.

Frequently expressed throughout the workshop were the beliefs that

new clarity, new thinking, new practices, and perhaps new regulations are needed broadly across a range of issues related to data stewardship. Expanded access and use of clinical data for knowledge development should begin with efforts to build consensus on guiding principles for the individuals and organizations who manage clinical data. Issues that were raised as particularly important to progress included data structure, standards, reporting requirements, quality assurance, timeliness, deidentification and other security measures, and access and use procedures.

Positioning Data as a Public Good

Emphasized in several presentations was the notion that understanding data ownership is central to considering data as a public good. Although data ownership is complex and highly nuanced, one consideration advanced focused on whether data are rightfully owned by those who provide the data—that is, patients—or those who collect and maintain the data. Similarly, a question repeatedly raised was whether data, if collected with public funds, should be considered part of the public domain. Data ownership is further complicated when data are deidentified and recollated for secondary use. The pressing needs expressed during discussion to clarify data ownership and to conceptualize new models for data stewardship present a compelling set of challenges. If an overarching goal is to make data more readily available and accessible for informing medical decision making, new solutions need to be identified and implemented to ease the tensions inherent in data ownership.

A significant challenge noted by participants is the need to break down barriers to data access based in assumptions and practices that treat clinical outcome data as a proprietary commodity. A paradigm shift might change assumptions that data are a commodity to be traded in a competitive marketplace into more open and supportive thinking about practices and policies that make clinical data more portable, transparent, and contributory to better health care. In discussions on the nature of clinical data, participants suggested that these data seem to exist in a gray area between a public and private good—and it is precisely in that gray area that considerable energies are needed to establish new definitions concerning the use of data for today's healthcare market.

Data for Improvements at the Point of Care

Ultimately, the utility of data is dependent on its application. As one workshop speaker said, it is important to "be able to take information drawn from actual experience in care delivery to be able to shape that process." For example, although strong progress has been made in understand-

ing disease and prevention, apart from filling the gaps in the evidence base, considerable delays persist in applying the knowledge to improve actual patient care. Part of the delay may be related to current approaches to using clinical data, which historically involved collecting, cleaning, and then—to some extent—hoarding data. Participants noted several key shortfalls of this approach. Separate repositories tend to be created for every specific purpose at great cost in both money and time. Privacy and security vulnerabilities are created due to multiple redundancies of large datasets and an absence of connectivity. Restrictions on data have the cumulative effect of precluding the ability of many people to use these data to make better decisions. Notably, exclusive and proprietary treatment of data effectively keeps the consumer and patient out of the equation. Furthermore, as many participants noted, in the current paradigm, by the time one collects and cleans data, the original research question or topic has often changed.

A new action agenda, one that is open to resetting some definitions and assumptions of health data and research approaches, was suggested as a first step toward the next generation of data translation and to positioning clinical data resources so that they contribute more readily and directly to effective health care. Progress might also depend on a fundamental shift in perspective. For example, instead of looking at how to achieve public health objectives from a data perspective, progress may require approaching issues and opportunities in evidence collection from the perspectives of medical decision makers, defined broadly to include providers, consumers, payers, and policy makers. Such a model for the clinical data utility would start in a climate of trust, with a policy framework that enables information liquidity. It would engage stakeholders in a constructive forward-looking process that prioritizes creating value for the participants and involves and rewards consumers for participating. It would focus on the infrastructure requirements to push more questions to the data as opposed to trying to bring all the data to every question. The vision put forth by many at the workshop was one in which research is a normative part of health care and in which every intervention with a patient presents an opportunity for learning.

Incentives for Data Sharing

Clarifying the business case for data sharing is one significant dimension of the economics of clinical data. Many participants noted the challenge and need for articulating a value proposition that clarifies the potential economic and health returns for the investment—for individual patients, for individual organizations that hold data, and for society as a whole. Unless overcome, the market obstacles related to sequestering data for proprietary interests and the technical obstacles related to individual identification

authorization and related issues will continue to compromise efforts to improve clinical data. Workshop presentations highlighted several opportunities for expanded data sharing to increase the value of healthcare delivered. More feedback to providers on quality and outcomes could provide incentives for those who collect data. The pay-off for patients could be measured through continuous improvement data, improved management and coordination of patient care, and possible cost savings. Similarly, the broad ability to share results from clinical and research data may reduce or avoid costs associated with clinical trials, as well as care delivery and patient management. Collaborative efforts to find new incentive aproaches that meet the multiple, often conflicting needs of diverse stakeholders will be important to facilitate data sharing.

Access to Data

Throughout the workshop, the need for increased access to a wider sample of clinical data called for new models that promote data sharing in both the public and private sectors. For the private sector, new incentives might encourage broader sharing of data in the marketplace. One area highlighted in the workshop focuses on data that are collected with public funds. Federal and state funds that support medical care as well as clinical research represent a source of substantial clinical data, yet participants repeatedly observed that these data are not effectively applied toward the generation of new knowledge. Government agencies that distribute funds to support such data collection have established guidelines and regulations designed to foster more data sharing, but an ethos committed to data sharing remains elusive. More aggressive enforcement of government regulations on privacy and funding was offered as a possible next step to encourage broader data sharing.

The model of the Human Genome Project was offered as an example of the benefit of open data sharing, as were the public registration of clinical trials and the growth of new models of disclosure/publication of research results in open-access journals and digital repositories. Expanded access and use, though, go hand in hand with clarity, assurance, and transparency as to security safeguards and nature of health research processes.

Creating Next-Generation Data Utilities and Models

Potential solutions to the tensions around data ownership and stewardship come in many forms, including ideas and models for next-generation data utilities. Broadly speaking, the goals of the approaches suggested by participants are to find utilities that align data quality, standards, integrity, accessibility, and comprehensiveness, and that offer quick, open sharing

of data. Pathways are also needed that enable appropriate, individualized interpretation of data, and that appropriately translate information for patients.

Models of Note

Federal agencies are developing a number of utilities that can help make data—often large pools of data—more broadly and readily accessible, and that serve as potentially replicable models for further development. The National Institutes of Health (NIH), for example, requires researchers to pool data collected under NIH grants so that other investigators might benefit from those data. The NIH created the database of Genotypes and Phenotypes (dbGaP) to archive and distribute the results, for example, of genome-wide association studies. Significant amounts of product development data are required by law to be in the public domain. The more than 50,000 trials registered at www.clinicaltrials.gov also provide rich data resources. The Food and Drug Administration (FDA) registration approval process deriving from the *Food and Drug Administration Amendments Act of 2007* will ensure the posting of more clinical summary data. The Cancer Biomedical Informatics Grid (caBIG) program was launched by the National Cancer Institute to connect the cancer research community to more easily share information and to build or adapt tools for collecting, analyzing, integrating, and disseminating data. The Centers for Medicare & Medicaid Services (CMS) has developed The Coverage with Evidence Development program to require the delivery of clinical data over and above the typical claims data as a provision for payment for certain services. Such an approach has the potential to provide significant amounts of information if we can learn how to meet the challenge of what we can do with data that have been collected, and merge those data with other sources of data so that data collection can inform clinical practice. The development of such approaches, which seek to make data in the government's domain more readily and openly accessible, could be expanded across all of the government—and serve as models for the private sector.

Efforts to develop next-generation data utilities are occurring in multiple loci. The model presented at the meeting of registries developed by medical disciplinary societies hold promise for the future, and their further development should be encouraged and supported. The American College of Cardiology's (ACC's) National Cardiovascular Data Registry (NCDR), for example, collects data for measuring quality in the catheterization laboratory and on acute coronary syndrome, percutaneous coronary interventions, implantable cardioverter defibrillators, and carotid artery revascularizations. NCDR registries were designed to improve the qual-

ity of cardiovascular patient care by giving cardiologists wide access to information, knowledge, and tools; benchmarks for quality improvement; updated programs for quality assurance; platforms for outcomes research; and solutions for postmarket surveillance. Mandates from Medicare and states have pushed hospitals to use the ACC registries, but there is room for wider adoption of such approaches.

Electronic Health Records

A central emphasis in presentations was the need for much more developmental work to address the issues of basic technology, interoperability, and standardization of terminology that currently impede the sharing of knowledge via EHRs. A vehicle to enhance collaboration among larger EHR users of different vendors may accelerate the emergence of the following: (1) more standardized agreements and approaches to integrating and sharing data across multiple platforms; (2) common query strategies; (3) virtual data warehousing rules and strategies; (4) relational standards; and (5) engagement of ways to reduce misperceptions on regulatory compliance issues. A potential area of action would be to convene an affinity group of EHR stakeholders to consider approaches to cooperative work on knowledge development, including issues related to standards and rules for governed data query.

There was also the sense that correction is needed to what is perceived as a market failure for expanding electronic health records. As evidenced by participant observations such as the insufficient demand by providers or patients in the face of the considerable expense of EHR adoption to small organizations, and challenges related to the diversity of platforms and reporting requirements, market incentives have not resulted in the large-scale EHR adoption needed. A particular emphasis in discussions was the notion that apart from government, no stakeholder group has the critical mass necessary to drive the broad-scale adoption, application and complementarity of EHR systems and a more focused strategy, perhaps involving Medicare leadership, is needed for progress.

One impediment to the use of EHRs for data aggregation was illustrated through the experience of one disciplinary association, which found that early adoption of EHRs among physicians may not have as much to do with quality, e-prescribing, or population management as with day-to-day business management concerns. Broader thinking about the potential and use of EHRs is needed to achieve widespread data aggregation at any level other than administrative data.

Patient-Focused Approaches to Clinical Data

Patient demand was viewed by many participants as an important strategic consideration for efforts to accelerate broad adoption and use of EHRs to improve patient care and knowledge development. Such demand may be boosted by greater patient use of programs that facilitate access to or maintenance of personal electronic health records. Several consumer-oriented products—currently deployed or under development—were presented at the workshop as emerging services that seek to enable patient and consumer engagement in clinical information management. Participants noted several potential impacts of these tools on patient care, including thedemonstration of the value in the speed and ease of personal access to the information, better accommodation of patient preference in care, and fostering a partnership spirit conducive to broader EHR application.

Patient engagement and control is key to enabling broader use of clinical data for knowledge development. One suggested approach to investing patients in the process is to guarantee transparency when data are provided for research. Clarity should extend to issues such as the structure, risks, and benefits of access to data. A model being developed by vendors in the private sector—notably, Microsoft and Google—is for a health platform that puts users in control of their information so they can access, store, and call on it however and wherever they wish, including for the purpose of making better informed health decisions. Importantly, though, strong policies and clear operating standards are needed when anyone, including patients, access and use sensitive health information.

Interoperability, Data Aggregation, and Data Mining

Across the $2.5 trillion healthcare system few resources are devoted to data sharing, despite its centrality to patient care and improvement. Many barriers to data sharing exist, ranging from competitive concerns to technological challenges, and overall progress in this area has been frustrating. As one workshop participant said, "we have spent the past decade wondering why we cannot collect the data we need to answer our questions." Stakeholders across the healthcare system currently struggle with poor data quality and formats and lack of the data needed. Nonetheless, despite the considerable barriers, many voices recognize that combining large sets of data offers distinct advantages, underscoring the need for practices that will encourage such aggregation.

In practice, there are considerable technical and operational challenges in data sharing across institutions. Health information technology (HIT) can effectively support quality improvement only to the extent that concerted efforts are made to ensure interoperability. Workshop discussion highlighted

the barriers to broad adoption and use of EHRs for knowledge development posed by platform incompatibilities. For example, the Department of Health and Human Services (HHS) operates 200 separate data systems in a range of areas with little or no means of coordination. Options proposed to help address these issues included establishing standards for data sharing, resolving issues of patient privacy that sometimes preclude data sharing, and revising the architecture of platforms to expedite the sharing of data so they can inform agreed-on goals. Workshop presentations illustrated that while some progress has been made in all of these areas, additional efforts are needed, including collaborative initiatives involving many stakeholders and a commitment of more energy and resources to the ongoing collection, integration, and interpretation of health data in order to better inform policy makers and the public.

In an examination of the trade-offs between pulling data from different data aggregators and the concept of a pooled mega-database, it was suggested that the need is growing for such a database that would pull data from different health plans, the Department of Veterans Affairs, state Medicaid programs, Medicare, and other sources. Such data could then be standardized, creating a public good that would be available for research and multiple other purposes, such as cost-effectiveness studies and explorations of drug safety. An alternative model would be a distributive research network, such as one in which it would be possible to conduct research in different settings with a standard research protocol, and then to pool the results afterward; specific data could conceivably continue to reside with the individual partners who collected them.

Standards

To improve interoperability, access, transferability, and translation, improved standards are needed across the clinical data space. Data standardization is partially an issue of messages and transporting data; however, many workshop participants emphasized the need to address standardization issues related to terminologies and classification—noting that a common lexicon is needed to help ensure that data entered into one place in one system can be useful not only elsewhere in that system, but also in other systems. Common diagnosis codes, for example, are needed to help push the translation of data into healthcare improvements at the point of care—to say nothing of accelerating efficiencies in billing processes.

The ACC is working to standardize data that are collected for its registries. The short- term goal is to measure gaps in performance and guideline compliance; the long-term goal is to teach others how to fill those gaps, thus creating a cycle of continuous quality improvement. The American Academy of Family Physicians has been working to establish and promote HIT

standards that are focused on clinical data, such as the American Society for Testing and Materials' Continuity of Care Record Standard.

Another standardization issue discussed at the workshop underscores the ways in which current technologies lag behind current needs. Although large repositories of controlled clinical trial data exist, including primary data, much of that information exists on paper in various archives, not in any electronic form that would enable sharing. Efforts to digitize health records continue, but further work is needed to effectively retire the paper generation and fully migrate data to electronic forms. Without the complete migration to digital health records, these data will not be immediately accessible, useful, and analyzable.

Creating Next-Generation Data Policy

In one sense, limitations imposed on clinical data by law create what one workshop speaker described as a health information chess board. Often, legislation and regulation impede rather than support our ability to ensure that health information is a public good. In the sense that many of these restrictions—such as those that govern paper-based data collection—were born in a different age and may not be as relevant and useful today, policy does not keep up with practice. Some of today's barriers are unanticipated byproducts of earlier regulatory decisions. To reach goals based on wider access and utility of clinical data, therefore, some of those regulatory underpinnings may need to be reevaluated.

Privacy

New thinking regarding patient privacy in the context of data emerged from discussion as an important opportunity for progress. One participant noted that in some ways we still have a privacy paradigm rooted in the paper age. Important implications of this lag behind technology include the absence of a framework for privacy that recognizes that health information is no longer a static good, but increasingly is a portable, moving, compounding, growing asset, and privacy practices that impede the sharing of data for point-of-service health care.

Deidentification of data raises privacy issues. Several participants stressed that unless, for example, the standard notification of privacy practices is changed so that data can be used if they are deidentified for secondary use in clinical research, the challenges to aggregating data among various institutions will likely persist. Possible solutions proposed include initially registering patients with a unique identifier that at some point could be associated for research purposes.

As an advisory committee to HHS, the National Committee on Vital

and Health Statistics (NCVHS) has developed some principles that might inform further thinking. For example, NCVHS has recommended that covered entities strengthen the terms of their business contracts to be more specific about what data will be used, how, and by whom. Recognizing that transparency is important to consumers, another NCVHS recommendation was that notices of privacy practices need to be more meaningful, and that individuals should be able to request and be given additional information about the specific uses and users of their data. New models of personal health records, such as Microsoft's HealthVault, are being designed in tandem with new models of voluntary patient control privacy policies. Such avenues deserve closer attention as potential guides for new thinking.

Revisiting HIPAA

A view expressed throughout the workshop is that many elements of the Health Insurance Portability and Accountability Act of 1996 (HIPAA) are outdated. HIPAA's Privacy Rules were characterized as confusing and, among other drawbacks, HIPAA limits the portability of data and overall data sharing, does not encourage standardization, and is inadequate in ensuring timely use of data. Moreover, enforcement of HIPAA was seen as nearly nonexistent. Several participants suggested that modifications are needed to strike the proper balance between protecting patient privacy and making data available for research necessary to improve healthcare quality and lower costs; and that a revised legislative code of privacy, confidentiality, and security would help to support and promote the next generation of clinical data.

Raised during the meeting as an important issue to address was the elimination of accounting disclosure obligations—which could help to reduce the cost of sharing data. Because of HIT and other developments in technology, the deidentification standard discussion might be best framed in the context of whether the deidentification safe harbor is too narrow. Another question raised by participants as important for discussion is whether liability burdens are properly distributed. Identifying the most significant barriers that remain, including those related to future unspecified research and data deidentification, and clearly defining policy alternatives will be helpful in promoting the research enterprise. Bills currently proposed in support of HIT and EHR systems, with strong built-in privacy protections, may offer necessary remedies, but continued work and vigilance in this area is needed. One specific suggestion was that at the conclusion of the Institute of Medicine's recent study on HIPAA and privacy protection regulations a follow-up meeting could be convened to explore what can be done within the existing structure to clarify definitions and reduce the

tendencies for unnecessarily restrictive interpretations, in particular related to secondary use of data. [1]

Legal Considerations

In addition to regulatory concerns, participants also raised concerns regarding clinical data within the construct of the legal system. The legal system enters the "public good" debate because it reflects and so perpetuates the current "excludability" state of clinical data with property and intellectual property models. Furthermore, market exchanges or shifts to public good "nonexcludability" face legal barriers (e.g., privacy, confidentiality, and security) that are designed to reduce or eliminate negative externalities suffered by data subjects. Presentations illustrated how certain legal rules create barriers to clinical data as a public good, such as property or inalienability rules, federal–state disconnects, and evolving data protection models. Resolving the current excludability rules was noted as an area in need of greater attention, and it was suggested that a more rigorous data protection model may be required as a predicate for greater access to patient data.

For clinically rich information, one suggested approach was to start from scratch to build a new, much stronger model, as has been recommended several times by NCVHS. Alternatively, NCVHS's secondary stewardship model could be combined and melded with a slimmed-down version of the European data directive. This would potentially create a more robust data protection model that would impose stronger obligations on data stewards through chain-of-trust data processes. Other attributes of this model would be strict limitations on data being processed for a purpose other than the original purpose of collection, a more relative approach

[1] In 2009, the *American Recovery and Reinvestment Act* (ARRA), which included several provisions aimed at improving privacy and security standards for health information technology (HIT), was signed into law. As noted by the Markle Foundation, "The Department for Health and Human Services is charged with developing regulations and/or guidance for ARRA's new health information privacy provisions and enhanced enforcement, including the following: HIPAA security and privacy rules extended to business associates of HIPAA-covered entities; new provisions for notification to consumers of information breaches; limitations on sales of protected health information; new guidance on 'minimum necessary' (i.e., the notion that no more than the necessary information should be disclosed); guidance on implementation specification to deidentify protected health information; individual right to access personal information in electronic format; annual guidance on the most effective technical safeguards for carrying out the HIPAA Security Rule; recommendations on technologies that protect the privacy of health information and promote security; restrictions on use of protected health information for marketing; and consumer access required to an accounting of disclosures of information maintained in EHRs" (Markle Connecting for Health Collaborative. *Achieving the Health Objectives of the American Recovery and Reinvestment Act* [April 2009]).

to data than HIPAA, and an overall design aimed at gaining the trust of patients and providers by making sure they understand secondary uses of data, adjustments possible when relative levels of deidentification are encountered, lack of reidentification, and so on.

Engaging the Public

The meeting emphasized the fact that moving to the next generation of clinical data depends in large part on an ability to engage and inform the public in shaping evidence development strategies. As illustrated by the broad representation at the meeting, all stakeholders can play a contributory role in better orienting data collection and use efforts to improve health. Noted in particular was the importance of keeping the public closely involved in and informed about all activities related to using clinical data to generate new knowledge—as progress will likely hinge on public acceptance. Discussion focused in particular on efforts that encouraged direct engagement of consumers with their personal health data.

Participants underscored the need for new approaches that engage the public in a deeper conversation about data utility—to better educate the public about the importance and intricacies of the data debate, to help the public become more fully vested in the discussion, and to help the research community to better understand the public's point of view and stake in the data utility. Such efforts simultaneously present the opportunity to reinstate the patient as the focal point of the data utility, in contrast to a role somewhere on the periphery, and to engage the public as a key driver of change to improve data utility overall. Finding genuine ways to respond to the public's desires regarding key issues, including access to data in language they can understand, portability of individual data throughout the healthcare system, and quality data that informs health care at the point of care was viewed as an important means to fostering broader public engagement in a healthcare system.

Clarifying the Social Contract

Data sharing is, in essence, a social contract between individuals and researchers who want to use their data. The current approach to obtaining data access; however, was described as falling far short of a clear illustration of the potential returns on this contract. Patients are told there will be some pay-off from sharing data—that they will have better, safer quality care, and that researchers will learn more about the disease so the healthcare system may be able to take better care of the patient or the patient's children. Yet many participants suggested that perhaps patients do not hear enough about how that is supposed to happen. Where does the pay-off come? How does

the other side of that contract deliver? What are the deliverables? Is there a timeline for those deliverables? Is there accountability for those deliverables? Those questions need to be addressed, probably by a broad coalition of stakeholders in clinical data, with patients at the center of the discussion. The experience of the VA with *My Healthy Vet* showed that patients want to be asked about the use of their personal data and to engage in a process to give consent about how those data are used; but, overall, patients do not demonstrate a need to control their data at a microscopic level.

As part of the social contract, it was suggested that perhaps more specific requirements are needed of those collecting the data. Fundamentally, policies could be developed that require those collecting data to do something specific with the information they collect, and that further require reporting back on what was done with the data. At the same time, perhaps policies could be set that inform and guide quality assurance about the data, and set specific expectations about the timeliness of their use.

Participants also noted the need for more attention to mechanisms and practices that ensure and protect patient privacy. For example, the suggestion was made that an independent health privacy audit and verification process be established. One presenter suggested that a program of empirical studies on the impacts of EHRs may be warranted—to assess privacy, access, and security programs in selected research settings—as well as the development of patient satisfaction and trust surveys to chart patients' experiences and attitudes in evidence-based research programs.[2]

Public Information and Education

Educating patients and the public about the value of clinical data, and promoting privacy-compliant, evidence-based health research might require national educational campaigns. To date, there has not been a public demonstration of the utility of data sharing and of the potential impact of data on personal lifestyle, bottom line, or other meaningful endpoints—an area viewed by some participants as ripe for further exploration. National public information campaigns and programming could go a long way to help convince consumers and patients of the public good aspect of sharing their medical data.

One suggested option to demonstrate the value of research as a public good is to consider expanded reporting of meaningful clinical health data. The New York Department of Health's effort to have all labs report A1C data from diabetics to the public health department, while not a perfect

[2] As of 2009, several relevant federal and private programs were under consideration for auditing privacy and security operations in EHR systems (e.g., HITECH, Patient Privacy Rights proposals).

model, shows potential in part because it actively engages the department in prevention measures. Toward different but similar goals, the Massachusetts Health Quality Partners draws on health plan claims data to report trusted, reliable information on physician performance as a means of developing valid, comparable measures that could lead to quality improvement. The information is shared both with physicians, to help them improve the quality of care they provide, and to consumers, to help them make informed decisions about their health care. Information posted for the public has a patient information component in the sense that it is accompanied by consumer information about specific medical conditions, and about what patients should expect of their physician in that regard. Such models may have potential for duplication in other venues.

Another possible approach to demonstrating the value of research is the enhancement and expansion of clinical data registries. Many registries today remain siloed, and do not collect data in a way that can be useful and meaningful to the public. Current systems are not designed, for example, to track major causes of U.S. death and disability that may be compelling and ultimately important to the public. Discussion about the use of registries highlighted their development as a key opportunity to think about data elements needed and information and outcomes desired so that the use of registry data can generate more public interest and consumer value.

Another possibility suggested is to develop a nationwide health tracking network. Such an approach could help to identify, track, and prevent health-related causes of death, whether they are environmental, occupational, or lifestyle/behavioral. It could inform the public health community, providers, policy makers, and consumers about disease rates by geography, ethnicity, and other relevant criteria.

The public and the patient must drive change and advocate for using data as a public utility. As noted by several participants, some hospitals and research institutions have shown innovation in educating their patients about how important clinical research is to patient care; many have started what are essentially marketing campaigns to educate patients and their families on the importance of participation in clinical trials and related research endeavors. Learnings from these initial efforts might help shape broader efforts to educate people on the importance and potential benefits of using data for secondary purposes.

Ultimately, the disparate voices engaged in efforts and thinking designed to improve clinical data utilities are looking for clarity on how they should go about, in the words of one workshop presenter, "compiling, analyzing, structuring, artifacting, [and] accountable-izing." As demonstrated throughout workshop discussions, stakeholders across health care are interested in designing an agenda that will move this work forward, and engaging with leaders who can advance the agenda. Such work was viewed by many par-

ticipants as essential to making the use of clinical data more efficient and perhaps equitable, and, ultimately ensuring that evidence-based medicine can draw more effectively and productively on the rich intelligence inherent in the data.

AREAS FOR INNOVATION AND COLLABORATIVE ACTION

Workshop presentations and discussions highlighted multiple opportunities to accelerate collaborative efforts to advance clinical data frontiers. Potential opportunities for follow-up attention by the members of the IOM Roundtable on Value & Science-Driven Health Care include those noted below—with Roundtable Innovation Collaboratives already engaged in related follow-on work indicated in parentheses.

1. *Principles:* Foster the development, review, and implementation of basic principles for data stewardship.
2. *Use of electronic health records for knowledge development:* Convene an affinity group of EHR users and vendors to consider approaches to cooperative work on knowledge development, including issues related to standards and rules for governed data query and application (EHR Innovation Collaborative).
3. *Collaborative data mining:* Organize exploratory efforts to investigate cutting-edge data-mining techniques for generating evidence on care practices and research (EHR Innovation Collaborative).
4. *Incentives:* Convene an employer–payer workgroup to explore approaches for the use of economic incentives to reward providers/groups working to improve knowledge generation and application in the care process.
5. *Privacy and security:* Follow the IOM study on HIPAA and privacy protection regulations with a series of meetings to explore and clarify definitions and reduce the tendency toward unnecessarily restrictive interpretations, in particular as they relate to data sharing and secondary uses.
6. *Transparency and access to federal data:* Explore the marketplace for data, opportunities to enhance data sharing, governance/stewardship issues, and ways to make federally sponsored clinical data widely available for secondary analysis. This includes not only data from federally supported research, but Medicare-related data, including from Part D (pharmaceutical) use.
7. *Public involvement in the evidence process:* Engage the public through communication efforts aimed at increasing public understanding and involvement in evidence-based medicine (Evidence Communication Innovation Collaborative).

These issues, and other related issues, will be further explored by the members of the IOM Roundtable on Value & Science-Driven Health Care, in collaboration with their colleagues in the field.

Appendixes

Appendix A

Workshop Agenda

Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good

A Learning Healthcare System Workshop
IOM Roundtable on Evidence-Based Medicine
February 28–29, 2008
The Keck Center of The National Academies
Washington, DC 20001

Issues Motivating the Discussion

1. Discovering what works best in medical care—including for whom and under what circumstances—requires that clinical data be carefully nurtured as a resource for continuous learning.
2. Transformational opportunities are presented by evolving large and potentially interoperable clinical and administrative datasets.
3. Clinical data are recorded and held in multiple activities and many institutions, including medical records, administrative and claims records, and research studies.
4. Public policy and public awareness lag behind the technical, organizational, and legal capacity for reliable safeguarding of individual privacy and data security in mining clinical data for new knowledge.
5. A significant challenge to progress resides in the barriers and restrictions that derive from the treatment of medical care data as a proprietary commodity by the organizations involved.

6. Even clinical research and medical care data developed with public funds are often not available for broader analysis and insights.
7. Broader access and use of healthcare data for new insights requires not only fostering data system reliability and interoperability, but addressing the matter of individual data ownership and the extent to which data central to progress in health and health care should constitute a public good.

Goal: To explore these issues, identify potential approaches, and discuss possible strategies for their engagement.

DAY ONE

8:30 **Welcome and Opening Remarks**
J. Michael McGinnis, Institute of Medicine

8:45 **Clinical Data as the Basic Staple of the Learning Healthcare System**
What is the current profile of our clinical data "utility"? What might be possible if all data sources could be readily and reliably drawn upon for new insights into healthcare effectiveness? What specific key steps would foster achieving this vision?
David J. Brailer, Health Evolution Partners

9:30 **Session 1: U.S. Healthcare Data Today: Current State of Play**
What purposes drive the collection of healthcare data in the United States and what is the system's current profile? How accessible are clinical data for new clinical insights, how well are they used, and what are the barriers? How might clinical data from all sources—publicly funded and privately funded—be made more useful to monitor clinical effectiveness?
Chair: Cato T. Laurencin, University of Virginia and IOM Roundtable on Value & Science-Driven Health Care

- **Current healthcare data profile**
  Simon P. Cohn, Kaiser Permanente and National Committee on Vital and Health Statistics
- **Data used as indicators for assessing, managing, and improving health care**
  Barbra G. Rabson, Massachusetts Health Quality Partners

[10:10–10:30 Break]

- **Data primarily collected for new insights**
  Michael S. Lauer, National Heart, Lung, and Blood Institute
- **Health product marketing data**
  William D. Marder, Thomson Healthcare

Panel discussion to follow

11:45 LUNCH

12:45 SESSION 2: CHANGING THE TERMS: DATA SYSTEM TRANSFORMATION IN PROGRESS

How is the national data utility changing now in arenas ranging from large linked sets to aggregated data and registries? What notable existing efforts are making medical care data more readily available and usable? What are the incentives and drivers for these activities? What are the shortfalls, limitations, and challenges highlighted by different categories of approaches to organizing and aggregating data? What dynamics are pushing integration?
Chair: Peter M. Neupert, Microsoft and IOM Roundtable on Value & Science-Driven Health Care

- **Emerging large-scale linked data systems and tools**
  Peter A. Covitz, National Cancer Institute
- **Networked data-sharing and standardized reporting initiatives**
  Pierre-André La Chance, Kaiser Permanente
- **Large health database aggregation**
  Steven E. Waldren, American Academy of Family Physicians
- **Registries and care with evidence development**
  Peter K. Smith, Duke University

Panel discussion to follow

2:15 SESSION 3: HEALTHCARE DATA: PUBLIC GOOD OR PRIVATE PROPERTY?

How does the structure of the medical care data marketplace affect research priorities, gaps, and possibilities? What are the characteristics of a public good or a public utility? On what dimensions do healthcare data compare? Can important distinctions be made within the spectrum of data types or sources? How might the case be made for improved access and sharing of medical data? What types of conceptual advances, guidance, or policy are needed?
Chair: Carmen Hooker Odom, Milbank Memorial Fund and IOM Roundtable on Value & Science-Driven Health Care

- **Characteristics of a public good and how applied to healthcare data**
  David Blumenthal, Massachusetts General Hospital
- **Characteristics of the marketplace for medical care data**
  William H. Crown, i3 Innovus
- **Legal issues related to data access, pooling, and use**
  Nicolas P. Terry, St. Louis University Law School

Panel discussion to follow

[3:45–4:00 BREAK]

4:00 SESSION 4: HEALTHCARE DATA AS A PUBLIC GOOD: PRIVACY AND SECURITY

Where is public opinion on these issues? What are the current legal and social challenges? What is the experience in other large data collection and management arenas?

Chair: Lynn Etheredge, George Washington University

- **Public views**
  Alan F. Westin, Privacy Consulting Group
- **HIPAA implications and issues**
  Marcy J. Wilder, Hogan & Hartson
- **Examples from other sectors**
  Elliot E. Maxwell, Johns Hopkins University
- **Institutional and technical approaches to ensuring privacy and security of clinical data**
  Alexander D. Eremia, MedStar Health

Panel discussion to follow

5:30 WRAP-UP COMMENTS FOR THE DAY, FOLLOWED BY RECEPTION

DAY TWO

9:00 WELCOME AND SHORT RECAP OF DAY ONE

J. Michael McGinnis, Institute of Medicine

9:15 VISION FOR THE FUTURE—CREATING A PUBLIC GOOD FOR THE PUBLIC'S HEALTH

What might be achieved if clinical data could be positioned as a public good? How would such a system work, and what are the

technical and policy issues to engage in fostering its evolution? Do we want to define integrated data as a public good?
Carol C. Diamond, The Markle Foundation

9:45 **Session 5: Creating the Next-Generation Data Utility—The Action Agenda**
What are some current or emerging opportunities to align policy developments with improved data access and evidence development? How might all stakeholders be engaged and what strategies or incentives are necessary given different vantage points? What are the implications of recent legislative initiatives: FDA safety legislation, national CER entity, SCHIP bill, and others? A panel of key decision makers and policy leaders will offer brief reflections, followed by an interactive discussion.
Chair: David Blumenthal, Massachusetts General Hospital

- **Government-sponsored clinical and claims data**
  Steve E. Phurrough, Centers for Medicare & Medicaid Services
- **Government-sponsored research data**
  James M. Ostell, National Center for Biotechnology Information
- **Professional organization-sponsored data**
  John (Jack) C. Lewin, American College of Cardiology

[10:30–10:45 Break]

- **Product development and testing data**
  Eve E. Slater, Pfizer
- **Regulatory policies to promote sharing**
  Janet Woodcock, Food and Drug Administration
- **Legislative change to allow sharing**
  Arthur A. Levin, Center for Medical Consumers

Panel discussion to follow

12:00 **Lunch**

1:00 **Session 6: Building Blocks for the Next-Generation Public Agenda**
What are the lessons from past efforts and signal features for future systems? Where are the greatest opportunities to take better advantage of existing data? What are some key strategic priorities in the architecture of the next-generation data utility? What specific actions will help to accelerate progress?
Chair: Peter I. Juhn, Johnson & Johnson

- **Organizational models**
  Mark B. McClellan, The Brookings Institution
- **Building on collaborative models**
  Chris B. Forrest, Children's Hospital of Philadelphia
- **Technical and operational challenges**
  Brian J. Kelly, Accenture
- **Economic incentives**
  C. Eugene Steuerle, Urban Institute

Panel discussion to follow

2:45 SESSION 7: ENGAGING THE PUBLIC

What are current public views on using clinical care data for research? In what types of information are patients interested, and how might this influence how they respond to potential uses of health information? What are the needed advances (technical, communication, demonstration of value) that might help address the concerns of healthcare consumers?

Chair: Donald M. Steinwachs, Johns Hopkins University and IOM Roundtable on Value & Science-Driven Health Care

- **Generating public interest in a public good**
  Alison Rein, AcademyHealth
- **Implications of "patients like me" databases**
  Courtney Hudson, EmergingMed
- **Implications of personal health records**
  Jim Karkanias, Microsoft

Panel discussion to follow

4:15 CONCLUDING SUMMARY REMARKS AND ADJOURNMENT

J. Michael McGinnis, Institute of Medicine

PLANNING COMMITTEE:

David Blumenthal, Massachusetts General Hospital, Harvard University
Marc Boutin, National Health Council
Mary Durham, Kaiser Foundation Hospitals
Lynn Etheredge, George Washington University
George Isham, HealthPartners Inc.
Peter Juhn, Johnson & Johnson
Alexander Walker, Harvard School of Public Health, Harvard University

Appendix B

Biographical Sketches of Workshop Participants

David Blumenthal, M.D., M.P.P., is director of the Harvard University Interfaculty Program for the Improvement of Health Care Policy and Systems, director of the Institute for Health Policy at Massachusetts General Hospital/Partners HealthCare System, and the Samuel O. Thier Professor of Medicine and Health Care Policy at Harvard Medical School. He has served as senior vice president at Boston's Brigham and Women's Hospital, a 720-bed Harvard teaching hospital, from 1987 to 1991, as well as executive director of the Center for Health Policy and Management and lecturer on public policy at the John F. Kennedy School of Government at Harvard (1981–1987). During the late 1970s, he served as a professional staff member on Senator Edward Kennedy's Senate Subcommittee on Health and Scientific Research. He is a member of the Institute of Medicine (IOM) and serves on several editorial boards, including *The New England Journal of Medicine*, *American Journal of Medicine*, and the *Journal of Health Politics, Policy and Law.* He is the founding chairman of the Academy for Health Services Research and Health Policy, the national organization of health services researchers.

Marc M. Boutin, J.D., is the executive vice president and chief operating officer of the National Health Council. He previously served as the executive vice president and prior to that as the vice president, policy development and advocacy. In addition to overseeing financial management and operations at the National Health Council, Mr. Boutin builds consensus among member patient advocacy organizations enabling them to speak with one voice on systemic policy initiatives resulting in legislation and

regulations that address the collective needs of patients and their family caregivers. Throughout Mr. Boutin's career, he has been highly involved in health advocacy, policy and legislation. He has designed and directed advocacy strategies for legislative initiatives, which have included issues ranging from access to health care to cancer prevention. Before joining the Council, Mr. Boutin most recently served as the Vice President of Government Relations and Advocacy at the American Cancer Society for New England. In addition, he was a faculty member at Tufts University Medical School, where he lectured on health care policy.

David J. Brailer, M.D., Ph.D., holds doctoral degrees in both medicine and economics. He earned his M.D. from West Virginia University and his Ph.D. in economics from The Wharton School. He became board certified in internal medicine after an internship and a residency at the Hospital of the University of Pennsylvania, and practiced in HIV medicine and immune deficiency until 2002. Dr. Brailer was appointed a Charles A. Dana Fellow and a Robert Wood Johnson Clinical Scholar at the University of Pennsylvania. He founded and was chair and chief executive officer (CEO) of CareScience, Inc., which set a new standard for healthcare quality and accountability by developing the nation's first online physician and hospital quality reports, the first healthcare Internet-based application service provider, and the first health information exchange. Under his leadership, the company built a network of hospitals and physicians that is still improving quality of care today. In May 2004, President George W. Bush tapped him to be the nation's change agent and chief evangelist for health information technology (HIT). Dr. Brailer led federal and private-sector efforts to improve healthcare quality, accountability, and efficiency through widespread deployment of HIT. Dr. Brailer was voted the Most Powerful Person in Health Care by the readers of *Modern Healthcare* in 2004. In just 2 years, Dr. Brailer set the nation's healthcare industry on a course toward modernized health information standards, certification of health information tools, state-of-the-art information-sharing architectures, and new policies for protection of consumer privacy. He pushed information technology (IT) solutions for adverse drug events, bioterrorism, pandemic flu, and other public health threats. Having set the foundation for the nation's digital era of medicine in place, Dr. Brailer left the federal government to return to the private sector. In May 2007, Dr. Brailer founded and became chair of Health Evolution Partners, a private equity fund focused on transforming the healthcare industry. Health Evolution Partners finances innovative ways for health care to be financed, organized, and delivered.

Simon P. Cohn, M.D., M.P.H., is the associate executive director of Health Information Policy for The Permanente Federation, Kaiser Permanente.

Dr. Cohn has been a leader in Kaiser Permanente's efforts to develop and implement comprehensive health information systems to support both the delivery of health care and health research. He is a nationally recognized expert on health information policy, including the *Health Insurance Portability and Accountability Act* (HIPAA) Administrative Simplification, healthcare data management, clinical and administrative classifications, and the electronic transmission of healthcare data. Dr. Cohn is chair of the National Committee on Vital and Health Statistics (NCVHS), the statutory public advisory committee to the Department of Health and Human Services (HHS) on health information policy, HIPAA, and the national health information infrastructure. He was also a member of the IOM's Committee on Data Standards for Patient Safety and was a member of the American Medical Association (AMA) Common Procedural Terminology Editorial Panel from 1997 to 2005. In 2002, Dr. Cohn was a recipient of the President's Award from the American Medical Informatics Association for his contributions to the field and was also elected a Fellow of the American College of Medical Informatics. In 2005, he was the recipient of the Leadership in Technology Award from the Workgroup on Electronic Data Interchange in recognition of his national leadership related to healthcare electronic data interchange and e-commerce. Dr. Cohn's medical specialty is emergency medicine, and he is a Fellow of the American College of Emergency Physicians.

Peter A. Covitz, Ph.D., is chief operating officer of the National Cancer Institute (NCI) Center for Bioinformatics in Rockville, Maryland. He previously led the core infrastructure group at the NCI Center for Bioinformatics, and was responsible for developing the NCI's platform for interoperable information systems, a model-driven, service-oriented architecture called caCORE. Prior to joining the NCI, Dr. Covitz was vice president of professional services at InforMax, Inc., where he ran the bioinformatics service and support division of the company. Earlier in his career, Dr. Covitz worked as a research scientist and manager at Incyte Pharmaceuticals and Molecular Applications Group. Dr. Covitz did his graduate work on transcriptional regulation at Columbia University, and postdoctoral training in genomics and bioinformatics at Stanford University.

William H. Crown, Ph.D., is president of i3 Innovus, the Health Economics and Outcomes Research division of Ingenix. From 1982 to 1995, Dr. Crown was a faculty member at the Florence Heller Graduate School, Brandeis University, where he taught graduate courses in statistics and conducted research on the economics of aging and long-term care policy. Prior to joining Ingenix in 2004, Dr. Crown was vice president of Outcomes Research and Econometrics at Medstat, where he conducted numerous

retrospective database analyses of the burden of illness associated with various diseases—particularly respiratory and mental health conditions. Dr. Crown's work in the area of depression was one of the first applications of econometric techniques in outcomes research to control for the effects of selection bias when using retrospective data to evaluate drug technologies. He has 25 years of experience conducting health policy and income maintenance research for private- and public-sector clients. Dr. Crown is author or coauthor of 4 books and more than 90 refereed journal articles, book chapters, and other publications.

Carol C. Diamond, M.D., M.P.H., is the managing director of the Health Program of the Markle Foundation and chairs Connecting for Health, a public-private collaborative working to realize the full potential of information technology in health and health care in the United States. Connecting for Health engages more than 100 diverse organizations and institutions in an approach rooted in core values, including achieving medical excellence, fostering patient participation, and protecting personal privacy. Before joining the Markle Foundation, Dr. Diamond was president of U.S. Quality Algorithms® (USQA®), Aetna U.S. Healthcare's performance measurement affiliate. Prior to joining USQA, Dr. Diamond was a consultant for Johnson & Johnson (J&J) Health Care Systems and The Robert Wood Johnson Foundation (RWJF). Dr. Diamond sits on the American Academy of Family Physicians (AAFP) Public Advisory Board and the Electronic Health Record Safety Institute Advisory Board of the Geisinger Center for Health Research, and is a member of the IPRO Advisory Board for the Centers for Medicare & Medicaid Services' Doctor's Office Quality-Information Technology (DOQ-IT) project in New York. Dr. Diamond earned her dual B.A./M.D. at the Medical School of the State University of New York at Brooklyn and her master's degree in public health at the University of Medicine and Dentistry of New Jersey, part of Rutgers University.

Mary L. Durham, Ph.D., is the director of the Center for Health Research (CHR) and Vice President/Research for Kaiser Permanente. In addition to her leadership roles, Dr. Durham conducts her own research on workplace health and translational research. She has designed and conducted research studies with employers as partners in the research, such as the Work, Family Health Network, and has over 20 years of hands-on content knowledge and experience in workplace policies and practices. Dr. Durham is also the Associate Director of the Oregon Clinical and Translational Research Institute, (OCTRI), one of the first 12 NIH Clinical and Translational Science Awards. In a unique partnership with Oregon Health & Science University, CHR role in OCTRI contributes to the science of translating research to

practice and developing strong interactions between discovery and dissemination. In tandem with her distinguished health research career, Dr. Durham has worked with state and federal lawmakers in crafting policy-level decisions across a wide range of topics, such as privacy, mental health law, genetics, research, and human subjects protection. She was a Commissioner for the State of Oregon's Senate Commission for Health Care Access and Affordability. She has had diverse roles, from providing expert testimony to the President's National Bioethics Advisory Commission to consulting to the World Health Organization. Dr. Durham has served on the boards of the Association for Health Services Research (Academy Health), Group Health Community Foundation, and Kaiser Permanente Health Care Alternatives. A professor in the Department of Public Health and Preventive Medicine at the Oregon Health & Science University, Dr. Durham is also an adjunct professor in the Department of Sociology at Portland State University. She is a former commissioner of the American Bar Association, where she served on the ABA's Commission on Mental and Physical Disability Law.

Dr. Durham received her Ph.D. in sociology from the University of Oklahoma in 1978, specializing in medical sociology. Before moving to Portland, she was deputy director of Group Health Cooperative of Puget Sound Center for Health Studies and on the faculty at the University of Washington School of Public Health and Community Medicine, in the Department of Health Services.

Alexander D. Eremia, J.D., LL.M., is associate general counsel and chief privacy officer for MedStar Health, Inc., one of the largest providers of healthcare services in the Mid-Atlantic region. MedStar Health has more than 25,000 employees and 5,000 affiliated physicians. Mr. Eremia provides legal guidance to the MedStar Health system on a wide range of regulatory and compliance matters, including implementation of Office of Inspector General (OIG) workplan initiatives, the regulatory requirements for clinical research, laboratory activities, Nuclear Regulatory Commission regulations, and the health information privacy and security requirements under HIPAA. He also serves as legal counsel to the various corporate committees charged with oversight of these functions. In addition, Mr. Eremia provides general legal counsel to the physicians and staff on many other legal issues, including contracting, employment, litigation/subpoenas, mental health law, and patient care issues, among others. Mr. Eremia came to MedStar Health from the Civil Recoveries Branch in the Office of Counsel to the Inspector General, OIG, HHS, where he was an associate counsel. During this period, he was also appointed as a special assistant U.S. attorney for the Eastern District of Virginia to prosecute healthcare fraud matters. He received his J.D. from DePaul College of Law with certification as a health lawyer; he was a

staff writer on the *Journal of Health Law* at DePaul. He received his LL.M. (Master of Laws, Health Law) from Loyola University–Chicago School of Law, where he was awarded the LL.M. Fellowship and was senior editor of the *Annals of Health Law*.

Lynn M. Etheredge is an independent consultant working on healthcare and social policy issues. His career started at the White House Office of Management and Budget (OMB). During the Nixon and Ford administrations, he was OMB's principal analyst for Medicare & Medicaid and led its staff work on national health insurance proposals. He returned to OMB as a senior career executive and headed its professional health staff in the Carter and Reagan administrations. He was a coauthor of the Jackson Hole Group's proposals for healthcare reform and a cofounder of the Health Insurance Reform Project at George Washington University. During the past several years, Mr. Etheredge has authored policy studies about Medicare reform, Medicaid, evidence-based medicine (EBM), and expanding health insurance coverage. His current projects include: (1) developing a "tax credits + Medicaid" model for health insurance coverage; (2) assessing the role of technology in health sector changes; and (3) developing a national rapid learning system for new technologies, including the use of electronic health records. He is author of more than 70 publications and is a graduate of Swarthmore College.

Christopher B. Forrest, M.D., Ph.D., serves as senior vice president and chief transformation officer at the Children's Hospital of Philadelphia (CHOP). He is leading the development of the Institute to Transform and Advance Children's Healthcare (iTACH). The goals of iTACH are to harness clinical and business information to improve children's health, make their health care more efficient, and transform the delivery system. Dr. Forrest is a general pediatrician and professor of pediatrics in the Division of General Pediatrics at the University of Pennsylvania School of Medicine. He lectures on transforming health care, health information technology innovations, and child health. Dr. Forrest has methodological expertise in health services and outcomes research and evaluation, health status assessment of children and adolescents, primary care, and use of health informatics to improve child health. He has authored numerous scientific manuscripts and reviews, and his research is supported by a broad mix of public, foundation, and private funders. Dr. Forrest received his B.A. and M.D. at Boston University as part of a dual-degree program. He trained in pediatrics at CHOP, where he also served as chief resident. He completed a Ph.D. in health services research at Johns Hopkins University.

Carmen Hooker Odom, M.R.P., is president of the Milbank Memorial Fund. She was appointed as secretary of the North Carolina Department of Health and Human Services by Governor Mike Easley in 2001. Ms. Hooker Odom, a former Massachusetts lawmaker and healthcare lobbyist, has spent her professional life working in health and human services. Prior to her appointment, she served as vice president of government relations for Quintiles Transnational Corporation in Research Triangle Park and as the group vice president for Carolinas HealthCare System. She is also an adjunct professor at the University of North Carolina School of Public Health. From 1995 to 1996, Ms. Hooker Odom worked as a project officer for the Milbank Memorial Fund, a New York–based foundation that conducts nonpartisan analysis, study, and research on significant issues in health policy. Prior to moving to North Carolina in 1995, Ms. Hooker Odom served as a member of the Massachusetts House of Representatives for nearly 11 years. She was the primary legislative author of both the 1991 Massachusetts comprehensive health reform legislation and the Children's Medical Security Plan, which targeted young children not covered by medical insurance. Ms. Hooker Odom cochaired the North Carolina Health Care Reform Commission and is a member of the North Carolina Institute of Medicine. She received a bachelor's degree in sociology and political science from Springfield College and a master's degree in regional planning from the University of Massachusetts at Amherst.

Courtney Hudson, M.B.A., cofounded EmergingMed in 2000 after spending 3 years as a health services analyst and institutional sales representative with CIBC Oppenheimer. Prior to joining Oppenheimer, Ms. Hudson was cofounder and vice president of business development and information systems with a multistate, Medicaid-focused health maintenance organization (HMO) start-up. In addition to two other HMO start-ups, Ms. Hudson has worked for a variety of academic medical centers and county public health agencies and clinics, where she cultivated expertise in the design and operation of efficient healthcare delivery systems for vulnerable populations. Ms. Hudson holds an M.B.A. (supplemented with extensive training through the School of Public Health) from the University of Washington as well as a B.S. in biology from Yale University.

George Isham, M.D., M.S., is the Chief Health Officer and Plan Medical Director at HealthPartners where he is responsible for quality, utilization management, health promotion and disease prevention, research, and health professionals' education at HealthPartners. He is active in strategic planning and policy issues. He is a founding board member of the Institute for Clinical Systems Improvement, a collaborative of Twin Cities medical groups and health plans that is implementing clinical practice guidelines

in Minnesota. Isham is a past member of the board of directors of the American's Health Insurance Plans and he is currently on the board of directors of the Alliance of Community Health Plans. He is past co-chair and current member of the National Committee for Quality Assurance's (NCQA's) Committee on Performance Measurement which oversees health plan quality measurement standards. He has served on the Center for Disease Control's (CDC's) Task Force on Community Preventive Services and on the Agency for Healthcare Research and Quality's (AHRQ's) Advisory Board for the National Guideline Clearinghouse. He has served on the Institute of Medicine's Board on Population Health and Public Health Services and chaired the committee that authored the report Priority Areas for National Action, Transforming Health Care Quality. In 2003, Isham was appointed as a lifetime National Associate of the National Academies of Science in recognition of his contributions to the work of the Institute of Medicine. Epidemic of Care, published in April 2003, with co-author George Halvorson, is Isham's examination of the impending healthcare crisis with suggestions on ways to solve it. Prior to his current position, Isham was medical director for MedCenters Health Plan in Minneapolis and executive director for University Health Care, Inc., in Madison, Wisconsin. His practice experience as a primary care physician includes eight years at the Freeport Clinic in Freeport, Illinois, and 3 and one-half years as clinical assistant professor in Medicine at the University of Wisconsin. Dr. Isham has a B.A. in Zoology and completed his doctoral training at the University of Illinois followed by an internship and residency at the University of Wisconsin in Madison.

Peter Juhn, M.D., Ph.D., is responsible for shaping EBM policies at the J&J corporate level, especially as payers use EBM as a basis for decisions on reimbursement and coverage of pharmaceuticals and medical devices. He works with the various J&J operating companies on a global basis to anticipate the methods and types of evidence needed in this evolving payer environment. He also provides policy coverage for developments in new regulations for advanced therapies, HIT initiatives, and "pay-for-performance" programs. He is a member of the Agency for Healthcare Research and Quality (AHRQ) Effective Healthcare Program Stakeholder Group and is the industry representative on the Medicare Evidence Development and Coverage Advisory Committee. Most recently, he was vice president, Health Improvement Resources, at WellPoint Health Networks, where he managed the disease management programs for all the operating units. He also held senior positions at Kaiser Permanente, including founding executive director, Care Management Institute, Kaiser's corporate disease management and clinical policy entity, and president and CEO of CareTouch, Inc., an e-health start-up venture. He has a B.A. from the University of Chicago,

an M.D. from Harvard, and an M.P.H. from the University of Washington, where he was a Robert Wood Johnson Clinical Scholar. He completed his internal medicine residency at the University of Pennsylvania.

Jim Karkanias is a partner and the senior director of Applied Research and Technology at Microsoft Corporation. His background is in science, engineering, technology, and business. Formally trained as a researcher, Mr. Karkanias began his career in the labs as a bench scientist conducting neuroscience research for McNeil Pharmaceutical, a subsidiary of J&J. His research continued across several labs and companies, including Merck Research Laboratories. His preclinical research career expanded to encompass clinical research in the neurosciences and graduate work in bioengineering, spanning neuroplasticity, sensorium integration, pain modulation, cognition, memory, and network theory. Mr. Karkanias also developed significant technical skills in computer hardware and software to help streamline his activities. This led to a position with Merck's Worldwide Clinical Information Systems, where he managed a group that created the systems infrastructure necessary to conduct research. The group's most notable innovation was a distributed system that allowed for the remote collection of data from physicians, which in 1993 predated both distributed system environments and the widespread use of the Internet as a business tool. He is now applying his skills in the Health Solutions Group, which aims to revolutionize health care through paradigm-shifting approaches that integrate next-generation business, process, and technology. He has a B.S. in neuroscience from Rutgers University and did graduate work in bioengineering, with a specialization in artificial intelligence, at the University of Pennsylvania and Drexel University.

Brian J. Kelly, M.D., M.B.A., M.S., is executive director in Accenture's Health & Life Sciences division, focusing on the federal health market. He also serves as the lead for Accenture's Global Electronic Health Record practice. He joined Accenture in 2003 after retiring from the Navy Medical Corps and concentrates on helping governments and commercial organizations optimize health care through effective business process transformation and the appropriate use of information technology. He has done work for HHS, Medicare, and a variety of commercial health plans and pharmaceutical companies. During his 20-year career in the Navy, he held a variety of clinical (hospital-based and operational) and teaching positions, and he provided clinical leadership within the Department of Defense in the fields of medical informatics and HIPAA. Dr. Kelly has an M.D. from New York Medical College, an M.B.A. from George Washington University, an M.S. in bioengineering from Clemson University, and a B.A. in Russian/pre-Med from Holy Cross College. He received his residency training in neurology at

Bethesda Naval Hospital and fellowship training in critical care medicine at the University of California–San Francisco. He is an associate professor of neurology and informatics at the Uniformed Services University of the Health Sciences in Bethesda, Maryland, and an associate professional lecturer in the School of Business and Public Management at George Washington University.

Pierre-André La Chance is the chief information officer and the research privacy officer for the Kaiser Permanente Center for Health Research. He also serves as a compliance consultant to the Kaiser Permanente National Research Council. Additionally, Mr. La Chance is the codirector of biomedical informatics for the Oregon Clinical and Translational Research Institute, a partnership between TCHR and Oregon Health & Science University. Across his various roles, one of Mr. La Chance's primary accountabilities is data warehousing as it applies to sustainable, rapid, cheap, safe, and high-quality data sharing across numerous research institutions.

Michael Lauer, M.D., joined the National Heart, Lung, and Blood Institute (NHLBI) in 2007 as director of the Division of Prevention and Population Science. A board-certified cardiologist, he received his M.D. from Albany Medical College in 1985 and underwent postgraduate training within the Harvard University system at Massachusetts General Hospital, Boston Beth Israel Hospital, and the Harvard School of Public Health. After completing specialized research training in cardiovascular epidemiology at the Framingham Heart Study, he joined the staff at the Cleveland Clinic in 1993. During 14 years at the clinic, he established a world-renowned clinical epidemiology research program with primary focus on diagnostic testing and comparative effectiveness. His research led to more than 150 publications in top medical journals, grant support from the American Heart Association and the National Institutes of Health (NIH), and election to the American Society of Clinical Investigation. Dr. Lauer has served as contributing editor for the *Journal of the American Medical Association*, co-director of the Cleveland Clinic Coronary Care Unit, director of cardiac clinical research, and first vice chair of the Cleveland Clinic Institutional Review Board. He achieved distinction in medical education, leading the development of a clinical research curriculum at the newly founded Cleveland Clinic Lerner Medical College at Case Western Reserve University, where he was professor of medicine, epidemiology, and biostatistics. In his current position at NHLBI, Dr. Lauer is leading a $300 million per year research division that oversees major programs in cardiovascular epidemiology and prevention.

Cato T. Laurencin, M.D., Ph.D., is the Lillian T. Pratt Professor and Chair of Orthopedic Surgery, university professor, and professor of biomedical engineering and chemical engineering at the University of Virginia (UVa). Prior to his appointment at UVa's Department of Biomedical Engineering, he was at Drexel University as the Helen I. Moorehead Professor of Chemical Engineering, and clinical associate professor of Orthopedic Surgery at The Medical College of Pennsylvania and Hahnemann University School of Medicine. Dr. Laurencin attended Princeton University, where he received his B.S.E. in chemical engineering, pursuing a topical program in polymer science and engineering. On completion of his undergraduate program, Dr. Laurencin began pursuing research at Massachusetts Institute of Technology (MIT), earning a Ph.D. in biochemical engineering/biotechnology. In parallel with his research training, Dr. Laurencin attended the Harvard Medical School, graduating magna cum laude. While directing his MIT laboratory, Dr. Laurencin undertook clinical residency training in orthopedic surgery at Harvard, and served as chief resident in orthopedic surgery at the Beth Israel Hospital, Harvard Medical School. Dr. Laurencin subsequently completed fellowship training in shoulder surgery and sports medicine at the Hospital for Special Surgery in New York, Cornell University, working with the team physicians for the New York Mets, and St. John's University. The focus of Dr. Laurencin's research is novel methods for bone and musculoskeletal tissue engineering and polymeric systems for drug delivery.

Arthur A. Levin, M.P.H., is cofounder and director of the Center for Medical Consumers, a New York City–based nonprofit organization committed to informed consumer and patient healthcare decision making; patient safety; evidence-based, high-quality medicine; and healthcare system transparency. Mr. Levin was a member of the IOM Committee on the Quality of Health Care that published *To Err Is Human* and *Crossing the Quality Chasm*. He has served on several other IOM committees and is a member of the IOM Board on Health Care Services. Mr. Levin serves on the National Committee for Quality Assurance's (NCQA's) Committee on Performance Measures and is a member of the National Quality Forum Consensus Standards Approval Committee. He has just completed a 4-year term as the consumer representative member of the Food and Drug Administration's (FDA's) Drug Safety and Risk Management Advisory Committee and continues to serve on select FDA advisory committees as a consultant expert in drug safety and risk management, representing consumers.

John (Jack) C. Lewin, M.D., has been CEO of the American College of Cardiology (ACC) since 2006. Under Dr. Lewin's leadership, ACC has aspired to contribute significantly to national leadership in advocacy related to expanding access to care for uninsured persons, and in reform-

ing Medicare, Medicaid, and the financing and delivery of quality health care. These efforts are part of ACC's mission to promote "heart health" and reductions in cardiovascular morbidity and mortality worldwide. Before coming to ACC, Dr. Lewin was CEO of the California Medical Association and its subsidiaries. He was also Hawaii's director of health from 1986 to 1994. In this role, he helped Hawaii achieve near-universal access to health care and revitalize statewide public health systems. In Hawaii, he was also CEO of the statewide 13-facility Community Hospital System. As a commissioned officer in the U.S. Public Health Service, he was the founder and first director of the Navajo Nation Department of Health, serving the needs of America's largest Indian tribe, straddling the three states of Arizona, New Mexico, and Utah. Trained in internal medicine, Dr. Lewin has also enjoyed many years of practicing primary care medicine during his career in Arizona, Hawaii, and California. He serves on numerous national boards and advisory bodies, including being founder and president of the Physicians' Foundations. He received his B.A. in biological sciences from the University of California–Irvine, and his M.D. from the University of Southern California.

William D. Marder, Ph.D., is senior vice president and general manager of the research and pharmaceutical units of Thomson Healthcare. Dr. Marder is a health economist whose work has focused on issues pertaining to physician behavior and medical markets. He is responsible for directing Thomson Healthcare's research, and database development services for the federal government, private research, and policy development organizations, and the pharmaceutical/medical technology industry. For the past 12 years, Dr. Marder has been responsible for developing and enhancing Thomson Healthcare's MarketScan® databases. As a professional economist and health services researcher, Dr. Marder has written on a variety of topics. His articles have been published in numerous scholarly journals, including the *American Journal of Public Health*, *Inquiry*, *Journal of Health Economics*, *Journal of Human Resources*, and *Medical Care*. Recently, his work has examined trends in spending and use for the privately insured U.S. population, with special focus on the interaction of clinical and economic incentives in healthcare delivery. He holds a Ph.D. in economics from the University of Chicago. He is a past president of the Illinois Economics Association and chair of the Health Economics Committee of the American Public Health Association.

Elliot E. Maxwell, J.D., advises public- and private-sector clients on strategic issues involving the intersection of business, technology, and public policy in the Internet and e-commerce domains. He is a Fellow of the communications program at Johns Hopkins University and Distinguished

Research Fellow at the eBusiness Research Center of Pennsylvania State University. From 1998 until 2001, Mr. Maxwell served as special adviser for the digital economy, including the Internet and e-commerce, to U.S. Secretary of Commerce William Daley and U.S. Secretary of Commerce Norm Mineta. After leaving the government in 2001, he was senior fellow for the Digital Economy and director of the Internet Policy Project for the Aspen Institute. Previously Mr. Maxwell worked for a number of years as a consultant and as assistant vice president for corporate strategy of Pacific Telesis Group. He has served at the Federal Communications Commission as special assistant to the chair, deputy chief of the Office of Plans and Policy, and deputy chief of the Office of Science and Technology. He also worked for the U.S. Senate as senior counsel to the U.S. Senate Select Committee on Intelligence Activities. Mr. Maxwell graduated from Brown University and Yale University Law School.

Mark B. McClellan, M.D., Ph.D., became director of the Engelberg Center for Healthcare Reform at the Brookings Institution in 2007. The center studies ways to provide practical solutions for access, quality, and financing challenges facing the U.S. healthcare system. In addition, Dr. McClellan is the Leonard D. Schaeffer Chair in Health Policy Studies. Dr. McClellan has a highly distinguished record in public service and in academic research. He is the former administrator for the Centers for Medicare & Medicaid Services (CMS) (2004–2006) and the former commissioner of the FDA (2002–2004). He also served as a member of the President's Council of Economic Advisors and senior director for healthcare policy at the White House (2001–2002). In these positions, he developed and implemented major reforms in health policy. Dr. McClellan was also an associate professor of economics and associate professor of medicine (with tenure) at Stanford University, from which he was on leave during his government service. He directed Stanford's Program on Health Outcomes Research and was also associate editor of the *Journal of Health Economics*, and coprincipal investigator of the Health and Retirement Study, a longitudinal study of the health and economic status of older Americans. His academic research has been concerned with the effectiveness of medical treatments in improving health; the economic and policy factors influencing medical treatment decisions and health outcomes; the impact of new technologies on public health and medical expenditures; and the relationship between health status and economic well-being. Dr. McClellan is a member of the IOM and a research associate at the National Bureau of Economic Research. A graduate of the University of Texas–Austin, Dr. McClellan earned his M.P.A. from Harvard's Kennedy School of Government in 1991, his M.D. from the Harvard–MIT Division of Health Sciences and Technology in 1992, and his Ph.D. in economics from MIT in 1993.

J. Michael McGinnis, M.D., M.P.P., a longtime contributor to national and international health policy leadership, is now senior scholar at the IOM, and executive director of the IOM Roundtable on Evidence-Based Medicine. He is also an elected member of the IOM. He previously was senior vice president at RWJ, and, unusual for political posts, held continuous appointment through the Carter, Reagan, (George H.W.) Bush, and Clinton administrations, with responsibility for coordinating activities and policies in disease prevention and health promotion. Programs and policies created and launched at his initiative include: the *Healthy People* process setting national health objectives, the U.S. Preventive Services Task Force, the *Dietary Guidelines for Americans* (with the U.S. Department of Agriculture), the *Ten Essential Services of Public Health*, the RWJF *Health and Society Scholars Program*, the RWJF *Young Epidemiology Scholars Program*, and the RWJF *Active Living* family of programs. Internationally, he chaired The World Bank/European Commission Task Force on postwar reconstruction of the health sector in Bosnia, and worked both as field epidemiologist and state coordinator for the World Health Organization's successful smallpox eradication program in India.

Peter M. Neupert, M.B.A., is corporate vice president for health strategy at Microsoft Corporation and is responsible for Microsoft's collaboration with the healthcare ecosystem to address global infrastructure issues of significant scale. Before rejoining Microsoft, Mr. Neupert served as president and CEO of Drugstore.com, Inc., from 1998 to 2001, and then as chair of the board of directors from 1999 to 2004. He led Drugstore.com to become a top online retail store and information site for health, wellness, beauty, and pharmacy products. Mr. Neupert served in various capacities at Microsoft from 1987 to 1998. He started at Microsoft as the director of operating systems responsible for shipping OS/2, and later was responsible for MSNBC as vice president of news and publishing for the interactive media group. He served on President George W. Bush's Information Technology Advisory Committee (PITAC) from 2003 to 2005. On that committee, he cochaired the Health Information Technology subcommittee and helped drive the report *Revolutionizing Health Care Through Information Technology*, published in 2004 by PITAC. In 2000, Mr. Neupert received an Ernst & Young Entrepreneur of the Year award for his work at Drugstore.com. He holds a bachelor's degree from Colorado College and an M.B.A. from the Tuck School of Business at Dartmouth College.

James M. Ostell, Ph.D., is chief of the Information Engineering Branch (IEB) of the National Center for Biotechnology Information (NCBI) at the NIH. The IEB is responsible for designing, building, and deploying nearly all the production resources at the NCBI, including PubMed, PubMed Central,

GenBank, BLAST, RefSeq, OMIM, dbSNP, dbGaP, and many others. The NCBI is one of the most heavily used biomedical information websites in the world, supporting millions of users a day at rates up to 5,000 hits a second. Dr. Ostell received his Ph.D. in cellular and developmental biology from Harvard University. He then wrote a commercially successful molecular biology software package. He joined the NIH and created the IEB when the NCBI was founded in 1988 and has run it ever since. Dr. Ostell was recently elected to the IOM.

Steve E. Phurrough, M.D., M.P.A., is director of the Coverage and Analysis Group at CMS. Using EBM principles, Dr. Phurrough assists in developing national policy on the appropriate devices, diagnostics, and procedures that should be provided by the Medicare program. Dr. Phurrough joined CMS in 2001 as the director of the Division of Medical and Surgical Services in the Coverage and Analysis Group after completing a long, distinguished career in the U.S. Army. In addition to being a practicing family practitioner, his military career also included managing Department of Defense regional healthcare delivery systems, creating national and international healthcare policy for the Army, and developing practice guidelines. Dr. Phurrough received his M.D. from the University of Alabama in Birmingham and an M.P.A. from the University of Colorado in Colorado Springs. He is board certified by the American Board of Family Practice and is a certified physician executive by the American College of Physician Executives.

Barbra G. Rabson, M.P.H., has been the executive director of the Massachusetts Health Quality Partners (MHQP) since 1998. Under Ms. Rabson's leadership, MHQP has become a trusted source of physician performance information in Massachusetts, and MHQP is nationally recognized for its collaborative approach to collecting and reporting performance information to improve care. MHQP is one of six quality coalitions selected nationwide to be an AQA/BQI pilot site for aggregating commercial and Medicare claims data. Ms. Rabson was the principal investigator for the RWJF Rewarding Results grant awarded to MHQP to evaluate how financial and nonfinancial incentives impact the quality of care. Ms. Rabson is a founding member of the Network for Regional Healthcare Improvement, a national network of regional quality collaboratives. She serves on the Board of the Massachusetts eHealthCollaborative, and is a member of Health Care for All's Advisory Committee on Quality Initiatives. Ms. Rabson brings broad-based experience from the managed-care, hospital, and healthcare arenas to her collaborative role at MHQP. She received her undergraduate degree from Brandeis University and her M.P.H. from Yale University. She was selected to participate in the Executive Leadership Development Program at the Hauser Center for

Nonprofit Organizations, where she studied strategy and nonprofit leadership at the Kennedy School of Government at Harvard University.

Alison Rein, M.S., is a senior manager at AcademyHealth. She works on several projects related to quality improvement and consumer engagement, primarily within AHRQ's Knowledge Transfer program. Ms. Rein also serves as the AcademyHealth project manager for the AHRQ Quality-Based Purchasing Knowledge Transfer project, and is working to expand AcademyHealth's involvement in state and national efforts to integrate health information exchange systems into the U.S. healthcare system. She serves on a number of advisory bodies related to health information exchange, including the Confidentiality, Privacy, & Security workgroup of the American Health Information Community, and the Health Information Protection Taskforce of the State Alliance for e-Health. Prior to joining AcademyHealth, Ms. Rein was assistant director of food and health policy at the National Consumers League, where she represented consumer interests in a variety of issues. Ms. Rein served as a healthcare consultant to a number of private and nonprofit organizations, for which she conducted strategic evaluations, market studies, and research efforts aimed at evaluating the relative effectiveness of numerous drug, biologic, and device interventions. She holds a master's in public policy analysis from the University of Rochester, and has coauthored several articles published in peer-reviewed medical journals.

Eve E. Slater, M.D., joined Pfizer as senior vice president for worldwide policy in 2007. Dr. Slater is a member of the Worldwide Public Affairs and Policy Leadership Team. She spent 19 years with Merck in a number of critical scientific, strategy, and public policy positions. In 1983 she came to Merck Research Laboratories as senior director of biochemical endocrinology, and in 1988 she was promoted to vice president, regulatory affairs. She advanced to vice president of clinical and regulatory development for Merck Research Laboratories in 1990, and senior vice president in 1994. In 2001 she was named senior vice president of external policy for Merck Research Laboratories, and vice president of Merck Corporate Public Affairs. An expert on HIV treatment research, she served as a member of the U.S. Keystone National Policy Dialogue on HIV, as well as the NIH Office of AIDS Research Advisory Council. She was appointed by the president in 2001 as assistant secretary of health. In this role, she served HHS Secretary Tommy Thompson as chief health policy adviser, with special emphasis on e-health and innovation, biosecurity, clinical trial protocols, women's health, eldercare, and HIV/AIDS. She received her B.A. from Vassar and her M.D. from Columbia University, and she is board certified in internal medicine and cardiology. She was the first woman named as chief resident at Massachusetts General Hospital, where she later led the hypertension

unit. She also served as assistant professor of medicine at Harvard Medical School. She continues to serve patients as an associate attending physician at New York Presbyterian Hospital and the medical profession as an associate professor of medicine at Columbia.

Peter K. Smith, M.D., is professor and chief of cardiothoracic surgery at Duke University. He is a graduate of Princeton University (Phi Beta Kappa), Duke Medical School (AOA), and the Duke general surgery and thoracic surgery residencies. He has received the AHA Clinician Scientist Award and an NIH Research Career Development Award, and is currently the Duke principal investigator in the NHLBI Cardiac Surgery Research Network. He has authored or coauthored 140 peer-reviewed publications. He has a long-standing interest in clinical databases and has managed the Duke Cardiac Surgery clinical database since 1987, collaborating with the STS National Cardiac Database since its inception. His most recent work has focused on comparative effectiveness of PCI and CABG and he is the surgeon member of the AHA/ACC Appropriateness Criteria writing committee. He pioneered the use of clinical databases to improve the accuracy of the Medicare Physician Fee Schedule as the thoracic surgery member of the AMA Relative Value Update Committee. In 2006 he received the Distinguished Service Award of the Society of Thoracic Surgeons.

Donald M. Steinwachs, Ph.D., is a professor in the Department of Health Policy and Management and director of the Health Services Research and Development Center at the Johns Hopkins University Bloomberg School of Public Health. Dr. Steinwachs's research seeks to identify opportunities to improve quality of health care and patient outcomes and, when feasible, evaluate promising quality improvement interventions. His previous research includes studies of medical effectiveness and patient outcomes for individuals with specific medical (e.g., asthma), surgical (e.g., cataract surgery), and psychiatric (e.g., schizophrenia) conditions. Dr. Steinwachs has contributed to the literature on the impact of managed-care and payments systems on access to care, quality, use, and cost. He was a codeveloper of the widely used Adjusted Clinical Groups (ACG) case mix adjustment. He has developed methods for measuring provider continuity, needs and unmet needs for care, and measures of the timeliness of care. He has a particular interest in the role of the routine management information system (MIS) as a source of data for evaluating the effectiveness and cost of health care. This includes work on the integration of outcomes management systems with existing MIS in managed-care settings. He is a member of the IOM Board on Health Care Services. An NCVHS member since 2002, he chairs its Subcommittee on populations and serves on its executive committee. He also

serves on the Board of Mathematica Policy Research, Inc. Dr. Steinwachs holds a B.S. in engineering mathematics, a M.S. in systems engineering from the University of Arizona, and a Ph.D. in operations research from Johns Hopkins University.

Eugene Steuerle, Ph.D., is a senior fellow at the Urban Institute and codirector of the Urban-Brookings Tax Policy Center. He is also the author, coauthor, editor, or coeditor of 13 books and nearly 1,000 reports, articles, columns, testimonies, and reports. His latest book is *Contemporary U.S. Tax Policy* (2nd ed.). In the area of health, Dr. Steuerle serves on the NCVHS and has published articles on issues such as the financing of health care, the use of mandates, and the economic effect of health insurance subsidies. He has provided Congress with testimony and served as "faculty" at health reform retreats by both the Senate Finance Committee and the House Ways and Means Committee. He has made proposals to focus on children, as well as to make better combined use of both incentives and mandates as a way to approach coverage goals that neither alone can attain. His proposals on the use of tax penalties as a way to enforce individual mandates recently have been adopted by a number of states as one mechanism for enforcement.

Nicolas P. Terry, LL.M., is the Chester A. Myers Professor of Law and codirector of the Center for Health Law Studies at Saint Louis University School of Law. He is an internationally recognized scholar on health law whose research interests lie primarily at the intersection of medicine, law, and technology. His recent scholarship has concentrated on technologically mediated health care (including telemedicine), privacy of medical information, electronic medical records, and the use of information and other technologies to reduce medical error. Educated at Kingston University and the University of Cambridge, Mr. Terry began his academic career as a member of the law faculty of the University of Exeter in England. In 1980 he joined Saint Louis University School of Law, where he has taught torts, products liability, healthcare law, eHealth, Internet law, and insurance law. Mr. Terry has served as a visiting professor at the University of Missouri–Columbia School of Law, Washington University School of Law, Santa Clara University School of Law, and most recently at the University of Iowa College of Law. During the 1996–1997 academic year, he was on leave from the law school and served as d19irector of legal education for LEXIS-NEXIS. He is a senior fellow at Melbourne Law School and holds the secondary appointment of professor of health management and policy at the Saint Louis University School of Public Health. Since 2000, Mr. Terry has been codirector of the Center for Health Law Studies at Saint Louis University—consistently ranked by *U.S. News & World Report* as the finest

health law program in the nation. The Center publishes the *Saint Louis University Journal of Health Law & Policy*.

Steven E. Waldren, M.D., is director of the American Academy of Family Physicians' (AAFP's) Center for Health Information Technology. Dr. Waldren joined the AAFP in 2004 because of his knowledge of design and management of health information systems, computer science, and medical informatics. Prior to joining the AAFP, Dr. Waldren was a National Library of Medicine Medical Informatics Fellow at the University of Missouri–Columbia. Dr. Waldren is also a residency-trained and board-certified family physician. He also participates in many healthcare informatics initiatives, including vice chair of the ASTM International E31 Health Information Standards Committee; cochair of the Ambulatory Functionality Working Group of the Certification Commission for Health-IT; and cochair of the AQA Alliance's Data Aggregation and HIT Subcommittee. Dr. Waldren combines his strong technical informatics knowledge and his clinical knowledge to further the adoption of standards-based HIT by physicians and consumers to improve the quality and safety of health care.

Alexander Walker, M.D., Dr.P.H., is Adjunct Professor of Epidemiology at Harvard School of Public Health, where he was formerly a professor and Chair of the Department of Epidemiology. He is also a principal of World Health Information Science Consultants, LLC. His research encompasses the safety of drugs, devices, vaccines, and medical procedures. Current studies include post-marketing safety studies for recently approved drugs, natural history of disease studies to provide context for Phase III clinical trials, studies of the impact of drug labeling and warnings on prescribing behavior, and determinants of drug uptake and discontinuation. Additional areas of research and expertise include health effects of chemicals used in the workplace and statistical methods in epidemiology. Dr. Walker received an M.D. degree from Harvard Medical School in 1974, and a doctorate of Public Health in Epidemiology from the Harvard School of Public Health in 1981. Dr. Walker is associate editor of Pharmacoepidemiology and Drug Safety and is on the Board of Directors of the International Society for Pharmacoepidemiology, which he also served as President in 1995–1996. He was a statistical consultant for the *New England Journal of Medicine* from 1992 through 1996 and a Contributing Editor of *The Lancet* from 1999 through 2001. From 2000 through 2007, he served as Senior Vice President for Epidemiology at Ingenix. Dr. Walker has written or contributed to more than 250 peer-reviewed articles in drug safety, epidemiology, and occupational health, and is the author of a book of essays, *Observation and Inference: An Introduction to the Methods of Epidemiology*.

Alan F. Westin, Ph.D., LL.B., is professor emeritus of public law and government, Columbia University, author of *Privacy & Freedom* (1967) and *Databanks in a Free Society* (1972), and recipient in 2005 of the Privacy Leadership Award of the International Association of Privacy Professionals. One of his main areas of activity has been the impact of information technology applications in health care. In 1976 he led the first field study of computer applications and privacy issues in U.S. health care, for the National Bureau of Standards, and served in the 1980s as research director for the National Commission on Confidentiality of Health Records. In 2005, with Vivian van Gelder, he wrote *Building Privacy by Design into Emerging Electronic Health Record Systems*. He has made keynote presentations on health privacy issues since the 1960s to more than 120 health conferences, health professional meetings, congressional hearings, and privacy conferences. Dr. Westin has been the designer and academic adviser for 10 national surveys (with Harris Interactive) on health privacy issues, beginning with the 1993 survey on Health Information Privacy and, most recently (2003–2008) in 10 national surveys focused on HIPAA privacy issues, privacy in health research (for the IOM, 2007), and electronic health records-and-privacy developments.

Marcy Wilder, J.D., is a partner at Hogan & Hartson, LLP, where she specializes in health information law, including compliance with HIPAA and federal and state privacy laws. Ms. Wilder assists clients in identifying, evaluating, and managing risks associated with privacy and information security practices and data breaches. She counsels clients on matters related to electronic data sharing, including the use of health information in e-health and Internet ventures, electronic medical and billing systems, research registries, tissue banks, and marketing initiatives. Her clients include information technology vendors, informatics companies, healthcare clearinghouses, e-health organizations, healthcare providers, pharmaceutical and medical device companies, health plans, universities, and research institutions. Prior to joining Hogan & Hartson, Ms. Wilder served as deputy general counsel of HHS, where she was lead attorney in the development of HIPAA privacy regulations. Ms. Wilder has been a featured speaker at HIT and HIPAA conferences and seminars across the country, and she lectures frequently on data privacy and security, data mining, and clinical and records research. After receiving her law degree, Ms. Wilder was a fellow in women's law and public policy at the Georgetown University Law Center.

Janet Woodcock, M.D., is deputy commissioner and chief medical officer at the FDA. She shares responsibility and collaborates with the FDA commissioner in planning, organizing, directing, staffing, coordinating, controlling, and evaluating the agency's scientific and medical regulatory

activities in order to achieve the FDA's mission. Dr. Woodcock has close interactions with diverse constituencies, including the clinical and scientific communities, members of Congress and the administration, national media, patient and consumer advocacy groups, the international drug regulatory community, the regulated industry, and representatives of federal and state agencies. She frequently appears in or is quoted by the national media and has testified repeatedly before Congress. Dr. Woodcock has led many cross-agency initiatives while at the FDA. She introduced the concept of pharmaceutical risk management in 2000 as a new approach to drug safety. She has led the Pharmaceutical Quality for the 21st Century Initiative since 2002. This effort, to modernize pharmaceutical manufacturing and its regulation through the application of modern science and quality management techniques, has been highly successful in meeting its objectives. She has spearheaded an initiative on pharmacogenomics that has led to unprecedented agency-industry interactions on pharmacogenomics use in drug development. Over the past 2 years, she has been leading the FDA's Critical Path Initiative, which is designed to improve the scientific basis for medical product development. Dr. Woodcock was director of the Center for Drug Evaluation and Research from 1994 to 2005. During this period, review times for new and generic drugs were cut in half, while the standards for quality, safety, and effectiveness were improved. Dr. Woodcock also oversaw initiatives to automate submission and review of applications and adverse-event reports. Now nearing completion, these initiatives will allow the center to make much more drug information publicly available.

Appendix C

Workshop Attendee List

Patricia Adams
National Pharmaceutical Council

Charles Alston
Manning Selvage & Lee

Margaret Anderson
FasterCures

Jill Arent
Wyeth Pharmaceuticals

Grant Bagley
American Acadamy of Pharmaceutical Physicians

Richard A. Bankowitz
Premier Inc.

Michael Banyas
NIH—Clinical Center

Dennis J. Barbour
Society for Investigative Dermatology

Galen Barbour
Uniformed Services University

Mercedes Barrs
Amylin Pharmaceuticals, Inc.

Eric Bass

Rachel Behrman
Food and Drug Administration

Rami Ben-Joseph
sanofi-aventis

Mark Benton
Milbank Memorial Fund

Stefano Bertuzzi
National Institutes of Health, Office of the Director

Aman Bhandari
Centers for Medicare & Medicaid Services

Shawn Bishop
U.S. Senate Finance Committee

Cate Blankenburg
American Osteopathic Association

Rachel Block
New York eHealth Collaborative

Meryl Bloomrosen
American Medical Informatics Association

David Blumenthal
Office of the National Coordinator for Health IT

Marilyn Sue Bogner
Institute for the Study of Human Error, LLC

Kathy Bowles
School of Nursing

David Brailer
Health Evolution Partners

Brad Brimhall
TriCore Reference Laboratories

Robert Brown
Tri-Service Infrastructure Management Program Office, Military Health System

Robert Browne
Eli Lilly and Company

Lynda Bryant-Comstock
GlaxoSmithKline

Laird Burnett
Kaiser Permanente

Eric Caplan
United Biosource Company

Tony Cardinelli
TRICARE

Linda Carter
Johnson & Johnson

Margaret Cary
Veterans Health Administration

Alan Cash
Armed Forces Institute of Pathology

Kee Chan
Boston University

John Chelen
Hampshire Research Institute

Kate Cherry
Georgetown University

Peter Cizmadia
KSJ & Associates

John Clarke
Drexel University

Anni Coden
IBM

Andrew Cohen
AGC & Associates

Perry D. Cohen
Parkinson Pipeline Project

Simon Cohn
The Permanente Federation

Elaine Collier
National Center for Research Resources

William Crown
i3 Innovus

Frederick Curro
New York University

Kirstin Dawson
America's Health Insurance Plans

James DeMaioribus
Johnson & Johnson

Anuj Desai
Johnson & Johnson Health Care Systems

Emily DeVoto
National Breast Cancer Coalition Fund

Christopher Dezii
Bristol-Myers Squibb Company

Carol Diamond
Markle Foundation

Dave Domann
Johnson & Johnson

Joyce Dubow
AARP Office of Policy and Strategy

Lammot DuPont
Manatt Health Solutions

Ria Eapen
National Consumers League

Jennifer Eaton
Health Resources and Services Administration

Shahram Ebadollahi
IBM

Brenda Ecken
Booz Allen Hamilton

Jill Eden
Institute of Medicine

Alfred Engelberg
The Engelberg Foundation

Alexander Eremia
MedStar Health, Inc.

Lynn Etheredge
George Washington University

Rebecca Etz
Robert Wood Johnson Medical School

Mary Evans
National Institutes of Health

Jim Fackler
Johns Hopkins University School of Medicine

Nancy Featherstone
AstraZeneca Pharmaceuticals, LLP

Laurie Feinberg
HHS/OS/ASPE/HP

Shelley Fichtner
Pharmaceutical Research and Manufacturers of America

Rosemarie Filart
National Institutes of Health

Steven Findlay
Consumers Union

Jack Fitzgibbons
CarePath, Inc.

Ila Flannigan
Georgetown University

Steve Flowers
IMS Health

Mona Floyd
Research!America

Raymond Formanek, Jr.
Food and Drug Administration

Christopher Forrest
The Children's Hospital of Philadelphia

Renee Fox
Robert Wood Johnson Health Policy

Steven Fox
Agency for Healthcare Research and Quality

Susan Friedman
American Osteopathic Association

Richard Fry
Foundation for Managed Care Pharmacy

Vicki Fung
Kaiser

Allison Garcimonde
Manatt Health Solutions

Steven Garfinkel
American Institutes for Research

Jennifer Garvin
Philadelphia Veterans Affairs Medical Center

Reed Gelzer
Advocates for Documentation Integrity and Compliance

Misrak Gezmu
National Institute of Allergy and Infectious Diseases

Mark Gorman
National Coalition for Cancer Survivorship

Gail Graham
Department of Veterans Affairs, Veterans Health Administration

Tina Grande
Healthcare Leadership Council

Linda Greenberg
Agency for Healthcare Research and Quality

Mary Greene
Booz Allen Hamilton

Ghanshyam Gupta
Division of Biostatistics, CBER, Food and Drug Administration

Kara Haas
Johnson & Johnson

Mary Hager
American Dietetic Association

Michael Halpern
American Cancer Society

Carolyn Hamm
Walter Reed Army Medical Center

Michael Handrigan
Agency for Healthcare Research and Quality

Anthony Hayward
National Center for Research Resources

John Hedstrom
The Society of Thoracic Surgeons

Janet Heinrich
Health Policy R&D

Robin Hertz
Pfizer, Inc.

Kathleen Hewitt
The American College of Cardiology

Laura Hoffmeister
CMS

Carmen Hooker Odom
Milbank Memorial Fund

Robert N. Hotchkiss
Hospital for Special Surgery

Han-Yao Huang
Merck

Courtney Hudson
EmergingMed

Susan Ingargiola
Manatt Health Solutions

Andrea Inserra
Booz Allen Hamilton

Belinda Ireland
BJC HealthCare

Sarah Johnson
Georgetown University

Anne Jones
American Association of Colleges of Osteopathic Medicine

Peter Juhn
Johnson & Johnson

Scott Kahan
Johns Hopkins School of Public Health

Jim Karkanias
Microsoft

Elisabeth Kato
Hayes, Inc.

Marcia Kean
Feinstein Kean Healthcare

Brian Kelly
Accenture

Neil Kirschner
American College of Physicians

Ronald Klar
Health Services Analysis, Inc.

Steven Krosnick
National Cancer Institute

Philip Kroth
University of New Mexico

Cara Krulewitch
Food and Drug Administration

Christian Kulak
Harris Interactive

Patricia Kurtz
The Joint Commission

Arnold Kuzmack
Food and Drug Administration

Pierre-André La Chance
Kaiser Permanente Center for Health Research

Mollie Lane
Senate Finance Committee

John Larney
CarePath, Inc.

Jeanne Larsen
Georgetown University Medical Center

Michael Lauer
National Heart, Lung, and Blood Institute

Cato Laurencin
University of Virginia

Martha Lee
Food and Drug Administration

Teresa Lee
Advanced Medical Technology Association (AdvaMed)

Harold Lehmann
Johns Hopkins University

James Leonard
Kennell and Associates

Arthur Levin
Center for Medical Consumers

Jack Lewin
American College of Cardiology

Keith Lind
AARP

Svetlana Lisanti
Bio-Medical Resources Group

Bryan Luce
United BioSource Corporation

Jaime Lucove
Allscripts

Susan Lumsden
Health Resource and Services Administration

Carole J. DeSpain Magoffin
National Minority Quality Forum

Brian Maloney
AstraZeneca Pharmaceuticals, LLP

William Marder
Thomson Healthcare

Danica Marinac-Dabic
Food and Drug Administration

John Martin
Premier, Inc.

Elliot Maxwell
eMaxwell & Associates

Robert Mayes
Agency for Healthcare Research and Quality

Doug McAllaster
DoD TRICARE Management Agency

Mark McClellan
The Brookings Institution

Jessyna McDonald
University of the District of Columbia

Newell McElwee
Pfizer, Inc.

Lori Meyers
HealthCore

Lori Mihalich
Vinson & Elkins, LLP

Amy Miller
Personalized Medicine Coalition

Nancy Miller
Office of the Director, National Institutes of Health

Tom Miller
American Enterprise Institute

JoAnn Minor
Food and Drug Administration, OSHI

Abdollah Moghaddam
Office of the Air Force Surgeon General

Carol Monaco
American Osteopathic Association

Nisha Money
Uniformed Services University of Health Sciences

Don Moran
sanofi-aventis

Hazel Moran
Mental Health America

Anne Mueller
AstraZeneca

Garry Neil
Johnson & Johnson

Chalapathy Neti
IBM

Peter Neupert
Microsoft

Nicole Newburg-Rinn
Health Resources and Services Administration

Christopher Nielson
Veterans Affairs

Georginah Anne Nyambura Munene
Umoja Women Health Care Mobile Clinics & Centers

Eduardo Ortiz
National Heart, Lung, and Blood Institute

Keith Ortiz
VantagePoint Consulting Group

James Ostell
National Library of Medicine

Jeng-Jong Pan
National Institute on Drug Abuse

Kavita Patel
U.S. Senate Committee on Health, Education, Labor, and Pensions

Milesh Patel
American Academy of Otolaryngology-Head and Neck Surgery

Perry Payne
George Washington University

Steve Pelletier
Pelletier Editorial

Eleanor M. Perfetto
Pfizer, Inc.

Gary Persinger
National Pharmaceutical Council

Steve Phurrough
Centers for Medicare & Medicaid Services

Susan Pingleton
Association of American Medical Colleges

Robin Portman
Booz Allen Hamilton

Eva Powell
National Partnership for Women and Families

Peter Quinn
CarePath, Inc.

Greg Raab
Raab & Associates, Inc.

Barbra Rabson
Massachusetts Health Quality Partners

John Rayburn
Healthcare Leadership Council

Alison Rein
AcademyHealth

Karin Remington
National Institute of General Medical Sciences

Mary Lacey Reuther
BGR Holding, LLC

Craig Richardson
Johnson & Johnson Health Care Systems

John Ring
American Heart Association

Wayne Rosenkrans
AstraZeneca Pharmaceuticals, LLP

Susan Ross
S.D. Ross Consulting

Jaewon Ryu
Department of Veterans Affairs

Jody Sachs
National Institutes of Health

Rajni Samavedam
Booz Allen Hamilton

Karen Sanders
American Psychiatric Association

Michael Sayre
National Institutes of Health

Leslie Schafer
TRICARE

Kristin Schneeman
FasterCures

Victor Schneider
Office of the Chief Health and Medical Officer

Stephen Schoenbaum
The Commonwealth Fund

Sven Seyffert
Medtronic

Chad Shearer
Center for Health Care Strategies

Gail Shearer
Consumers Union

George Silberman
Cancer Policy Group, LLC

Rebecca Singer Cohen
United BioSource Corporation

Eve Slater
Pfizer, Inc.

Isabella Sledge
United BioSource Corporation

Karen Smith
AstraZeneca Pharmaceuticals, LLP

Peter Smith
Duke University Medical Center

Scott R. Smith
Agency for Healthcare Research and Quality

James Sorace
U.S. Department of Health and Human Services

Michelle Spetman
Brigham & Women's Hospital

Melissa Stegun
George Washington University

Donald Steinwachs
Johns Hopkins University

Charles Stemple
Humana

Eugene Steuerle
Urban Institute

Melissa Stevens
FasterCures

Ansalan Stewart
Office of the Secretary of Health and Human Services

Harriette Stewart
Army Long Term Family Case Management

Mark Stewart
American College of Cardiology

Robin Stombler
Auburn Health Strategies, LLC

Michael Stoto
Georgetown University

David Sugano
Schering-Plough

Betty Tai
Center for the Clinical Trials Network

Robert Temple
Food and Drug Administration

Nicolas Terry
Center for Health Law Studies

Susan Thaul
Congressional Research Service

Jordana Thomas
U.S. Army—Telemedicine and Advanced Technology Research Center

Beverly Thompson
Uniformed Services University

Richard Toselli
Johnson & Johnson

Deborah Trautman
Office of Speaker Pelosi

Jonathan Treadwell
ECRI Institute

Brian Turcsanyi
Pittsburgh Regional Health Initiative

Nathan Vafaie
National Health Council

Allison Viola
American Health Information Management Association

Mary Walchak
Becton, Dickinson and Company

Steven Waldren
American Academy of Family Physicians

Christian Weaver
Tri-Service Infrastructure Management Program Office, Military Health System

Jeff Weinfeld
HRSA

Rachel Weinstein
Johnson & Johnson, Pharmaceutical Research & Development

Kathleen Weis
Pfizer, Inc.

Alan Westin
Columbia University

Christine Weston
Johns Hopkins School of Public Health

Kimberly Westrich
National Pharmaceutical Council

Edward Westrick
UMass Memorial Health Care

James Whitlock
Air Force Medical Service

Marcy Wilder
Hogan & Hartson, LLP

Claudia Williams
Markle Foundation

Karen Williams
National Pharmaceutical Council

Deborah Willis-Fillinger
Health Resources and Services Administration

Catherine Witkop
Agency for Healthcare Research and Quality/Johns Hopkins University

Kim Wittenberg
Agency for Healthcare Research and Quality

Hui-Hsing Wong
OASPE

Janet Woodcock
Food and Drug Administration

Chantal Worzala
Health Policy R&D

Laura Zick
Eli Lilly and Company

Appendix D

The IOM Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule

RECOMMENDATIONS SUMMARY

The committee's foremost recommendation is the following:

I. Congress should authorize HHS and other relevant federal agencies to develop a new approach to protecting privacy that would apply uniformly to all health research. When this new approach is implemented, HHS should exempt health research from the HIPAA Privacy Rule.

- Apply privacy, security, transparency, and accountability obligations to all health records used in research.

If national policy makers choose to continue to rely on the HIPAA Privacy Rule rather than adopt a new federal approach (Recommendation I), the committee recommends the following:

II. HHS should revise the HIPAA Privacy Rule and associated guidance.

A. HHS should reduce variability in interpretations of the HIPAA Privacy Rule in health research by covered entities, IRBs, and Privacy Boards through revised and expanded guidance and harmonization.

1. HHS should develop a dynamic, ongoing process to increase empirical knowledge about current "best practices" for privacy protection in responsible research using protected health information (PHI), and promote the use of those best practices.

2. HHS should encourage greater use of partially deidentified data called "limited datasets" and develop clear guidance on how to set up and comply with the associated data use agreements more efficiently and effectively, in order to enhance privacy in research by expanding use and usability of data with direct identifiers removed.
3. HHS should clarify the distinctions between "research" and "practice" to ensure appropriate IRB and Privacy Board oversight of PHI disclosures for these activities.
4. HHS guidance documents should simplify the HIPAA Privacy Rule's provisions regarding the use of PHI in activities preparatory to research and harmonize those provisions with the Common Rule, in order to facilitate appropriate IRB and Privacy Board oversight of identification and recruitment of potential research participants.

B. HHS should develop guidance materials to facilitate more effective use of existing data and materials for health research and public health purposes.

1. HHS should develop guidance that clearly states that individuals can authorize use of PHI stored in databases or associated with biospecimen banks for specified future research under the HIPAA Privacy Rule with IRB/Privacy Board oversight, as is allowed under the Common Rule, in order to facilitate use of repositories for health research.
2. HHS should develop clear guidance for use of a single form that permits individuals to authorize use and disclosure of health information in a clinical trial and to authorize the storage of their biospecimens collected in conjunction with the clinical trial, in order to simplify authorization for interrelated research activities.
3. HHS should clarify the circumstances under which DNA samples or sequences are considered PHI, in order to facilitate appropriate use of DNA in health research.
4. HHS should develop a mechanism for linking data from multiple sources so that more useful datasets can be made available for research in a manner that protects privacy, confidentiality, and security.

C. HHS should revise provisions of the HIPAA Privacy Rule that entail heavy burdens for covered entities and impede research without providing substantive improvements in patient privacy.

1. HHS should reform the requirements for the accounting of disclosures of PHI for research.
2. HHS should simplify the criteria that IRBs and Privacy Boards use in making determinations for when they can waive the requirements to obtain authorization from each patient whose PHI will be used for a research study, in order to facilitate appropriate authorization requirements for responsible research.

Regardless of whether Recommendation I or II is implemented, the following recommendation, which are independent of the Privacy Rule, should be adopted:

III. Implement changes necessary for both policy options above (Recommendations I and II).

A. All institutions (both covered entities and non-covered entities) in the health research community should take strong measures to safeguard the security of health data.

- HHS should also support the development and use of new security technologies and self-evaluation standards.

B. To encourage service on Institutional Review Boards, HHS—or, as necessary, Congress—should provide reasonable protection against civil suits for members of Institutional Review Boards and Privacy Boards who serve in good faith.

- But no protection for willful or wanton misconduct.

C. HHS and researchers should take steps to provide the public with more information about health research by:

1. Disseminating research results to study participants and the public.
2. Educating the public about how research is done and what value it provides.

OTHER PUBLICATIONS IN THE LEARNING HEALTHCARE SYSTEM SERIES

The Learning Healthcare System

Evidence-Based Medicine and the Changing Nature of Healthcare

Leadership Commitments to Improve Value in Healthcare: Finding Common Ground

Value in Health Care: Accounting for Cost, Quality, Safety, Outcomes, and Innovation

Redesigning the Clinical Effectiveness Research Paradigm: Innovation and Practice-Based Approaches

Clinical Data as the Basic Staple of Health Learning: Creating and Protecting a Public Good

Learning What Works: Infrastructure Required for Comparative Effectiveness Research (forthcoming)

The Healthcare Imperative: Lowering Costs and Improving Outcomes (forthcoming)

Engineering a Learning Healthcare System: A Look at the Future (forthcoming)