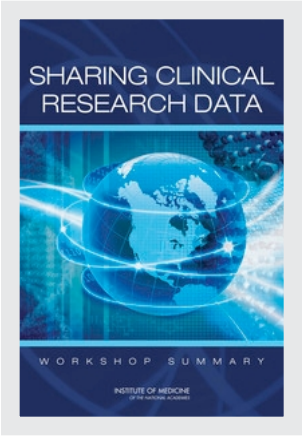

Sharing Clinical Research Data: Workshop Summary

DETAILS

156 pages | 8.5 x 11 | HARDBACK
ISBN 978-0-309-38484-1 | DOI 10.17226/18267

AUTHORS

Steve Olson and Autumn S. Downey, Rapporteurs; Forum on Drug Discovery, Development, and Translation; Forum on Neuroscience and Nervous System Disorders; National Cancer Policy Forum; Roundtable on Translating Genomic-Based Research for Health; Board on Health Sciences Policy; Board on Health Care Services; Institute of Medicine

SHARING CLINICAL RESEARCH DATA

WORKSHOP SUMMARY

Steve Olson and Autumn S. Downey, *Rapporteurs*

Forum on Drug Discovery, Development, and Translation
Forum on Neuroscience and Nervous System Disorders
National Cancer Policy Forum
Roundtable on Translating Genomic-Based Research for Health

Board on Health Sciences Policy
Board on Health Care Services

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

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

THE NATIONAL ACADEMIES PRESS • 500 Fifth Street, NW • 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 study was supported by contracts between the National Academy of Sciences and AbbVie, Alzheimer's Association, American Academy of Nursing, American Association for Cancer Research, American Cancer Society, American College of Medical Genetics and Genomics, American Diabetes Association, American Heart Association, American Medical Association, American Society for Microbiology, American Society of Clinical Oncology, American Society of Human Genetics, Amgen Inc., Association of American Cancer Institutes, Association of American Medical Colleges, Blue Cross and Blue Shield Association, Bristol-Myers Squibb, Burroughs Wellcome Fund, C-Change, Celtic Therapeutics, LLLP, CeNeRx Biopharma, Centers for Disease Control and Prevention (Contract Nos. 200-2005-13434 and 200-2011-38807), CEO Roundtable on Cancer, College of American Pathologists, Critical Path Institute, Department of Health and Human Services (Contract Nos. N01-OD-4-2139 and HHSF22301026T), Department of Health and Human Services' National Institutes of Health (NIH) (Contract No. N01-OD-4-2139) through the National Eye Institute, National Institute of Mental Health, National Institute of Neurological Disorders and Stroke, National Institute on Aging, National Institute on Alcohol Abuse and Alcoholism, National Institute on Drug Abuse, and NIH Blueprint for Neuroscience Research, Department of the Air Force (Contract No. FA7014-10-P-0072), Department of Veterans Affairs (Contract No. V101(93) P-2238), Doris Duke Charitable Foundation, Eli Lilly and Company (Contract No. LRL-0028-07), Fast Forward, LLC, FasterCures, Foundation for the National Institutes of Health, Friends of Cancer Research, GE Healthcare, Inc., Genetic Alliance, GlaxoSmithKline, Health Resources and Services Administration (Contract No. HHSH250201100119P), Janssen Research & Development, LLC, Johnson & Johnson Pharmaceutical Research and Development, LLC, The Kaiser Permanente Program Offices Community Benefit II at the East Bay Community Foundation (Contract No. 20121257), Life Technologies, Lundbeck Research USA, March of Dimes Foundation, Merck & Co., Inc., Merck Research Laboratories, The Michael J. Fox Foundation for Parkinson's Research, National Cancer Institute (Contract Nos. HHSN261200900003C and N01-OD-4-2139, TO#189), National Coalition for Health Professional Education in Genetics, National Heart, Lung, and Blood Institute (Contract No. N01-OD-4-2139, TO#275), National Human Genome Research Institute (Contract Nos. N01-OD-4-2139, TO#264 and HHSN263201200074I, TO#5), National Institute of Mental Health (Contract No. N01-OD-4-2139, TO#275), National Institute on Aging (Contract No. N01-OD-4-2139, TO#275), National Science Foundation (Contract No. OIA-0753701), National Society of Genetic Counselors, Northrop Grumman Health IT, Novartis Oncology, Novartis Pharmaceuticals Corporation, Office of Rare Diseases Research (Contract No. N01-OD-4-2139, TO#275), Oncology Nursing Society, One Mind for Research, Pfizer Inc. (Contract No. 140-N-1818071), Sanofi-Aventis, Society for Neuroscience, and Wellcome Trust. The views presented in this publication do not necessarily reflect the views of the organizations or agencies that provided support for the project.

International Standard Book Number-13: 978-0-309-26874-5
International Standard Book Number-10: 0-309-26874-5

Additional copies of this workshop summary are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; 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

Suggested citation: IOM (Institute of Medicine). 2013. *Sharing clinical research data: 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

PLANNING COMMITTEE ON SHARING CLINICAL RESEARCH DATA: AN INSTITUTE OF MEDICINE WORKSHOP[1]

SHARON TERRY (*Chair*), Genetic Alliance, Washington, DC
JOSEPHINE P. BRIGGS, National Institutes of Health, Bethesda, MD
TIMOTHY COETZEE, National Multiple Sclerosis Society, New York, NY
STEVEN GOODMAN, Stanford University School of Medicine, CA
ROBERT A. HARRINGTON, Stanford University School of Medicine, CA
LYNN HUDSON, Critical Path Institute, Tucson, AZ
CHARLES HUGH-JONES, Sanofi-Aventis, Inc., Bridgewater, NJ
JAN JOHANNESSEN, U.S. Food and Drug Administration, Silver Spring, MD
JEFFREY S. NYE, Janssen Research & Development, LLC, Titusville, NJ
RICHARD PLATT, Harvard Medical School, Boston, MA
WILLIAM Z. POTTER, FNIH Biomarkers Consortium, Philadelphia, PA
FRANK W. ROCKHOLD, GlaxoSmithKline Pharmaceuticals Research and Development, King of Prussia, PA
MICHAEL ROSENBLATT, Merck & Co., Inc., Whitehouse Station, NJ
VICKI SEYFERT-MARGOLIS, U.S. Food and Drug Administration, Silver Spring, MD
DEBORAH A. ZARIN, National Institutes of Health, Bethesda, MD

IOM Staff

AUTUMN DOWNEY, Program Officer
CLARE STROUD, Program Officer
REBECCA A. ENGLISH, Associate Program Officer
CLAIRE GIAMMARIA, Research Associate

[1]Institute of Medicine planning committees are solely responsible for organizing the workshop, identifying topics, and choosing speakers. The responsibility for the published workshop summary rests with the workshop rapporteurs and the institution. Statements, recommendations, and opinions expressed are those of individual presenters and participants, and are not necessarily endorsed or verified by the Institute of Medicine, and they should not be construed as reflecting any group consensus.

FRANK VALLIERE, Research Associate
TONIA E. DICKERSON, Senior Program Assistant
CRYSTI PARK, Senior Program Assistant
BRUCE M. ALTEVOGT, Director, Forum on Neuroscience and Nervous System Disorders
ADAM C. BERGER, Director, Roundtable on Translating Genomic-Based Research for Health
ANNE B. CLAIBORNE, Director, Forum on Drug Discovery, Development, and Translation
ROGER HERDMAN, Director, Board on Health Care Services
SHARYL J. NASS, Director, National Cancer Policy Forum
ANDREW M. POPE, Director, Board on Health Sciences Policy

FORUM ON DRUG DISCOVERY, DEVELOPMENT, AND TRANSLATION[1]

---

[1]Institute of Medicine 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.

JOHN ORLOFF, Novartis Pharmaceuticals Corporation, East Hanover, NJ
AMY PATTERSON, National Institutes of Health, Bethesda, MD
MICHAEL ROSENBLATT, Merck & Co., Inc., Whitehouse Station, NJ
JANET SHOEMAKER, American Society for Microbiology, Washington, DC
ELLEN SIGAL, Friends of Cancer Research, Washington, DC
ELLEN R. STRAHLMAN, GlaxoSmithKline, Research Triangle Park, NC
BRIAN L. STROM, University of Pennsylvania Perelman School of Medicine, Philadelphia
NANCY SUNG, Burroughs Wellcome Fund, Research Triangle Park, NC
JANET TOBIAS, Ikana Media and Mount Sinai School of Medicine, New York, NY
JOANNE WALDSTREICHER, Janssen Research & Development, LLC, Raritan, NJ
JANET WOODCOCK, U.S. Food and Drug Administration, White Oak, MD

IOM Staff

ANNE B. CLAIBORNE, Director, Forum on Drug Discovery, Development, and Translation
REBECCA A. ENGLISH, Associate Program Officer
ELIZABETH F. C. TYSON, Research Associate
ROBIN GUYSE, Senior Program Assistant
ANDREW M. POPE, Director, Board on Health Sciences Policy

FORUM ON NEUROSCIENCE AND NERVOUS SYSTEM DISORDERS[1]

---

[1]Institute of Medicine 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.

RICHARD MOHS, Lilly Research Laboratories, Indianapolis, IN
ALEXANDER OMMAYA, U.S. Department of Veterans Affairs, Washington, DC
ATUL PANDE, GlaxoSmithKline, Inc., Research Triangle, NC
STEVEN PAUL, Weill Cornell Medical College, New York, NY
TODD SHERER, Michael J. Fox Foundation for Parkinson's Research, New York, NY
PAUL SIEVING, National Institutes of Health, Bethesda, MD
JUDITH SIUCIAK, Foundation for the National Institutes of Health, Bethesda, MD
MARC TESSIER-LAVIGNE, The Rockefeller University, New York, NY
WILLIAM THIES, Alzheimer's Association, Chicago, IL
NORA VOLKOW, National Institute on Drug Abuse, Bethesda, MD
KENNETH WARREN, National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD
JOHN WILLIAMS, Wellcome Trust, London, United Kingdom
STEVIN ZORN, Lundbeck USA, Paramus, NJ
CHARLES ZORUMSKI, Washington University School of Medicine

IOM Staff

BRUCE M. ALTEVOGT, Director, Forum on Neuroscience and Nervous System Disorders
DIANA E. PANKEVICH, Program Officer
ELIZABETH K. THOMAS, Senior Program Assistant
ANDREW M. POPE, Director, Board on Health Sciences Policy

NATIONAL CANCER POLICY FORUM[1]

[1]Institute of Medicine 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.

LISA C. RICHARDSON, Centers for Disease Control and Prevention, Atlanta, GA
DEBASISH ROYCHOWDHURY, Sanofi Oncology, Cambridge, MA
YA-CHEN TINA SHIH, University of Chicago, IL
ELLEN SIGAL, Friends of Cancer Research, Washington, DC
STEVEN STEIN, Novartis Oncology, East Hanover, NJ
JOHN A. WAGNER, Merck and Company, Inc., Rahway, NJ
RALPH R. WEICHSELBAUM, University of Chicago, IL
JANET WOODCOCK, U.S. Food and Drug Administration, Rockville, MD

IOM Staff

SHARYL J. NASS, Director, National Cancer Policy Forum
LAURA LEVIT, Program Officer
ERIN BALOGH, Associate Program Officer
PAMELA LIGHTER, Research Assistant
MICHAEL PARK, Senior Program Assistant
PATRICK BURKE, Financial Associate
SHARON B. MURPHY, Scholar in Residence
ROGER HERDMAN, Director, Board on Health Care Services

ROUNDTABLE ON TRANSLATING GENOMIC-BASED RESEARCH FOR HEALTH[1]

---

[1]Institute of Medicine 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.

DEBRA LEONARD, Weill Cornell Medical Center of Cornell University, New York, NY
ELIZABETH MANSFIELD, U.S. Food and Drug Administration, Silver Spring, MD
KELLY MCVEARRY, Northrop Grumman Information Systems, Rockville, MD
GARRY NEIL, Johnson & Johnson, New Brunswick, NJ
ROBERT L. NUSSBAUM, University of California, San Francisco, School of Medicine
MICHELLE A. PENNY, Eli Lilly and Company, Indianapolis, IN
AIDAN POWER, Pfizer Inc., Groton, CT
VICTORIA M. PRATT, Quest Diagnostics Nichols Institute, Chantilly, VA
RONALD PRZYGODZKI, Department of Veterans Affairs, Washington, DC
ALLEN D. ROSES, Cabernet, Shiraz and Zinfandel Pharmaceuticals; Duke University, Durham, NC
KEVIN A. SCHULMAN, Duke University School of Medicine, Durham, NC
JOAN A. SCOTT, National Coalition for Health Professional Education in Genetics, Lutherville, MD
DAVID VEENSTRA, University of Washington, Seattle
MICHAEL S. WATSON, American College of Medical Genetics and Genomics, Bethesda, MD
DANIEL WATTENDORF, DARPA/Defense Sciences Office, Arlington, VA
CATHERINE A. WICKLUND, Northwestern University, Chicago, IL

IOM Staff

ADAM C. BERGER, Director, Roundtable on Translating Genomic-Based Research for Health
SEAN P. DAVID, James C. Puffer, M.D./American Board of Family Medicine Fellow
TONIA E. DICKERSON, Senior Program Assistant
ANDREW M. POPE, Director, Board on Health Sciences Policy

Reviewers

This workshop summary 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 summary as sound as possible and to ensure that the summary 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 process. We wish to thank the following individuals for their review of this workshop summary:

William W. Chin, Harvard Medical School
Carolyn Compton, Critical Path Institute
Barry Gertz, Affairs, Merck Research Laboratories
Jeffrey Nye, Janssen Research & Development, LLC

Although the reviewers listed above have provided many constructive comments and suggestions, they did not see the final draft of the workshop summary before its release. The review of this summary was overseen by **Melvin Worth, Jr.** Appointed by the Institute of Medicine, he 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 summary rests entirely with the authors and the institution.

Contents

Boxes and Figures

BOXES

FIGURES

Acronyms

BRIDG	Biomedical Research Integrated Domain Group
CAMD	Coalition Against Major Diseases
CDA	clinical documentation architecture
CDASH	Clinical Data Acquisition Standards Harmonization
CDE	common data element
CDER	Center for Drug Evaluation and Research
CDISC	Clinical Data Interchange Standards Consortium
CEO	chief executive officer
CFAST	Coalition for Accelerating Standards and Therapies
CLIA	Clinical Laboratory Improvement Amendments
C-Path	Critical Path Institute
CSR	clinical study report
eDISH	Electronic Drug-Induced Serious Hepatoxicity
EHR	electronic health record
EMA	European Medicines Agency
FDA	Food and Drug Administration
FDASIA	FDA Safety and Innovation Act
FNIH	Foundation for the National Institutes of Health
FTC	Federal Trade Commission
HHMI	Howard Hughes Medical Institute
HIPAA	Health Insurance Portability and Accountability Act
HL7	Health Level 7 International

iDASH	Integrated Data for Analysis, Anonymization, and Sharing
IHE	Integrating the Healthcare Enterprise
IOM	Institute of Medicine
IRB	institutional review board
NDA	new drug application
NEWMEDS	Novel Methods for Development of Drugs in Depression and Schizophrenia
NHLBI	National Heart, Lung, and Blood Institute
NIH	National Institutes of Health
NINDS	National Institute of Neurological Disorders and Stroke
NSF	National Science Foundation
PDUFA	Prescription Drug User Fee Act
PHI	protected health information
SDO	standards development organization
SDTM	Study Data Tabulation Model
YODA	Yale University Open Data Access

1

Introduction and Themes of the Workshop[1]

Pharmaceutical companies, academic researchers, and government agencies such as the Food and Drug Administration and the National Institutes of Health all possess large quantities of clinical research data. If these data were shared more widely within and across sectors, the resulting research advances derived from data pooling and analysis could improve public health, enhance patient safety, and spur drug development. Data sharing can also increase public trust in clinical trials and conclusions derived from them by lending transparency to the clinical research process. Much of this information, however, is never shared. Retention of clinical research data by investigators and within organizations may represent lost opportunities in biomedical research.

Despite the potential benefits that could be accrued from pooling and analysis of shared data, barriers to data sharing faced by researchers in industry include concerns about data mining, erroneous secondary analyses of data, and unwarranted litigation, as well as a desire to protect confidential commercial information. Academic partners face significant cultural barriers to sharing data and participating in longer term collaborative efforts that stem from a desire to protect intellectual autonomy and a career advancement system built on priority of publication and citation requirements. Some barriers, like the need to protect patient privacy,

[1]The planning committee's role was limited to planning the workshop, and the summary has been prepared by the workshop rapporteurs as a factual summary of what occurred at the workshop. Statements, recommendations, and opinions expressed are those of individual presenters and participants, and are not necessarily endorsed or verified by the forums, the roundtable, or the Institute of Medicine, and they should not be construed as reflecting any group consensus.

present challenges for both sectors. Looking ahead, there are also a number of technical challenges to be faced in analyzing potentially large and heterogeneous datasets.

Despite these barriers, there is increasing acknowledgment among researchers of the importance and potential benefits to sharing clinical research data at various stages of the research, discovery, and development pipeline. Precompetitive collaboration models promote the sharing of resources and risk among competitors at early stages of the research process, with the goal of providing benefit to all parties. A number of collaborations, including public-private partnerships, have formed to overcome these barriers in order to advance clinical research and accelerate the discovery and development of therapeutics and diagnostic tools.

On October 4-5, 2012, four groups within the Institute of Medicine—the Forum on Drug Discovery, Development, and Translation; the Forum on Neuroscience and Nervous System Disorders; the National Cancer Policy Forum; and the Roundtable on Translating Genomic-Based Research for Health—collectively hosted a workshop to examine and advance this pressing issue. The workshop explored the benefits of sharing clinical research data, the barriers to such sharing, and strategies to address these barriers to facilitate the development of safe, effective therapeutics and diagnostics. Box 1-1 provides the objectives of the workshop. The workshop was designed to provide a neutral venue where stakeholders from government, academia, industry, foundations, public-private partnerships, patient groups, and the public could meet to discuss issues of mutual interest. It is part of a larger effort to build partnerships and enhance collaboration within and among sectors on research, development, and assessment of pharmaceutical products. While acknowledging the importance of other kinds of clinical data, the workshop organizers focused on issues relating to sharing of data from preplanned interventional studies of human subjects. This summary of the workshop presents the observations, viewpoints, and suggestions made during both the presentations and discussion sessions as a way of informing the public, the press, and policy makers about the major issues surrounding the sharing of clinical research data.

ORGANIZATION OF THE REPORT

In the final session of the workshop, the moderators of each workshop session identified the major points that emerged during their sessions

BOX 1-1
Statement of Task for the Workshop

This public workshop focused on strategies to facilitate sharing of clinical research data in order to advance scientific knowledge and public health. While the workshop focused on sharing of data from preplanned interventional studies of human subjects, models and projects involving sharing of other clinical data types were considered to the extent that they provided lessons learned and best practices. The workshop objectives were to

- examine the benefits of sharing of clinical research data from all sectors and among these sectors, including, for example:
  — benefits to the research and development enterprise and
  — benefits to the analysis of safety and efficacy;
- identify barriers and challenges to sharing clinical research data;
- explore strategies to address these barriers and challenges, including identifying priority actions and "low-hanging fruit" opportunities; and
- discuss strategies for using these potentially large datasets to facilitate scientific and public health advances.

along with issues that warrant more focused attention. These points are presented in the next section of this first chapter of the workshop summary as an introduction to the themes of the workshop. Chapter 2 of this report summarizes the benefits of sharing clinical research data and suggests why and when increased data sharing can improve both scientific knowledge and public health. Chapter 3 considers the barriers to data sharing and examines changes that could overcome those barriers. Chapter 4 describes different models of data sharing to demonstrate best practices and lessons learned from each project. Chapter 5 looks at the standardization of clinical data, for both data collected in the past and future data collection efforts. Chapter 6 contains a discussion of mechanisms and incentives to enhance data transparency and sharing across all sectors. Chapter 7 concludes this workshop summary by gathering key take-away points from the workshop that were identified by speakers and other participants as a way to highlight and elaborate on next steps for advancing the sharing of clinical trials data.

THEMES OF THE WORKSHOP

Sharing of Clinical Trials Data: Benefits and Barriers (Chapters 2 and 3)

Many journal articles are at most a synopsis, and in many ways more like an advertisement, for immense quantities of data from which they are derived, said John Ioannidis, C.F. Rehnborg Chair in Disease Prevention at Stanford University. Much of the underlying data are not made available to other researchers or to the public, and are eventually discarded. As Ioannidis put it, the equivalent of several Libraries of Alexandria disappears every day. "How can we regain that information before it is too late?"

The benefits of sharing research data have been amply demonstrated in areas such as cardiovascular disease, where death rates have fallen 40 percent in recent decades, pointing toward the great potential of data sharing to improve human health, said William Potter, co-chair emeritus of the Neuroscience Steering Committee for the Biomarkers Consortium of the Foundation for the National Institutes of Health, in his summary of major messages from the first session of the workshop. Building and sharing datasets within and across the public and private sectors are practices that should be widely emulated. Advances in treatment for the complex illnesses faced by society today are not likely to result from the data from a single study that has been analyzed one time, he said; increased transparency and sharing at the participant level are needed to tackle these challenging diseases.

Several challenges have severely inhibited such sharing, said Jeffrey Nye, vice president for Neuroscience Innovation and Partnership Strategy at Janssen Research & Development, LLC, a Johnson & Johnson pharmaceutical company. For example, data holders have refrained from making data available to others due to privacy concerns. The process of de-identifying and standardizing data in some past cases of data sharing has been expensive and time consuming. Industry is interested in protecting proprietary information that can contribute to commercial products. Academic researchers have incentives to withhold data for their own use so they can advance professionally.

But a cultural shift is occurring, said Potter. Measures such as finding ways to credit academic researchers for sharing data are helping to reduce barriers to sharing. Industry has realized that it needs to collaborate to overcome the major obstacles it faces in developing new drugs, Potter continued. The sharing of data can correct mistaken conclusions and lead to new discoveries that would otherwise go undetected. Transparency in the use and dissemination of data can strengthen public trust in the biomedical research enterprise. Finally, he said, regulatory agencies are also recognizing the importance of facilitating this process and are working with researchers in academia and industry to identify paths forward.

Many stakeholders are involved in the sharing of clinical data, including participants in a trial, researchers, private companies, regulators, and the public, and each has particular interests and expectations. Effective communication and mutual understanding will be essential to identify common values and to take full advantage of current opportunities.

Models of Data Sharing (Chapter 4)

Successful models have demonstrated the value of sharing clinical trials data, said Jeffrey Nye, in his summary of the session on best practices and lessons learned from past experiences with data sharing. These models provide concrete examples of a vision of data sharing that can motivate action and lead to progress.

These models also have demonstrated some of the challenges of data sharing. The partners in data-sharing initiatives can have different cultures, practices, expectations, and rules. These differences need to be resolved, or at least accommodated, for sharing to occur. Sharing also can build trust among the participants in a collaboration, which in turn can provide the foundation for future initiatives.

One important message from the models presented at the workshop, Nye said, is that summary data can be inaccurate. Examination of participant-level data from clinical trials can be essential to draw correct conclusions from shared data.

Standardization to Enhance Data Sharing (Chapter 5)

Disclosure does not equal transparency, said Frank Rockhold, senior vice president for global clinical safety and pharmacovigilance at GlaxoSmithKline Pharmaceuticals Research and Development, who moderated the session on standardization and governance at the workshop. Data need to be understandable and analyzable if they are to be useful.

The need for data standards in clinical trials being conducted today and in the future is clear, Rockhold said. Standards can improve data quality, enable separate studies to be combined, and facilitate regulatory review. The application of data standards retrospectively to trials conducted in the past involves additional considerations. For example, the use of retrospective data to answer specific questions may be preferable to standardizing data and depositing the results in a repository for future use.

To date, most standards development has been done on a volunteer, ad hoc basis. Greater recognition of the value of standardization may lead to more cooperative efforts and more sustainable standards development models, Rockhold concluded.

Changing the Culture of Research (Chapter 6)

Data sharing is a public good, and the actions of the biomedical research enterprise should reflect that good, observed Robert Harrington, Arthur L. Bloomfield Professor of Medicine and chair of the department of medicine at Stanford University. But a variety of disincentives today create a culture that works against sharing. Academic researchers are afraid of losing credit for the work they have done to generate data. Industry is concerned about the loss of proprietary information and potential liability. Patients are worried about privacy. Essentially, every stakeholder associated with clinical trials faces difficulties in moving toward a more open system.

Good will and altruism go only so far in changing a culture, he said. Additional incentives are needed for substantial and enduring change to occur. For example, several factors came together to create powerful motivations for investigators to register trials at ClinicalTrials.gov. New

standards and tools could create new reasons to share data. Funders could consider an investigator's previous experience and future data-sharing plans in making grant funding decisions. Journals could agree on standard practices that authors must follow. A culture of data sharing could be built into the education of the next generation of clinical researchers.

The advent of organizations outside the traditional biomedical research enterprise offering data analysis and medical advice over the Internet has introduced a new force of cultural change, Harrington observed. If the system does not change from the inside, change may be imposed on it from the outside.

2

The Benefits of Data Sharing

Key Messages Identified by Individual Speakers

- Data sharing can enhance understanding of the results of an individual clinical trial and enable the pooling of data from multiple trials to extend scientific discoveries beyond those derivable from any single study.
- The moral and ethical arguments for data sharing center on fulfilling obligations to research participants, minimizing safety risks, and honoring the nature of medical research as a public good.
- The practical and scientific arguments for data sharing include improving the accuracy of research, informing risk/benefit analysis of treatment options, strengthening collaborations, accelerating biomedical research, and restoring trust in the clinical research enterprise.
- A cultural shift has already begun as leaders in industry, academia, and regulatory agencies recognize the value in increased transparency and data sharing and are focusing on how—instead of why—data should be shared.
- Participant-level data are particularly useful when shared, but care must be taken to avoid drawing inaccurate conclusions from reanalysis of such data.

Clinical data come in a variety of formats (see Box 2-1), from the raw data collected in case report forms during trials to the coded data

BOX 2-1
What Is Participant-Level Data?

Terms such as "participant-level data," "individual patient data," and "raw data" are not well defined, noted Elizabeth Loder of the *BMJ.* A mutual understanding of the way these data are generated and shared can help alleviate ambiguities in nomenclature. In a typical multicenter clinical trial, data originate with case report forms, which can be handwritten or electronic. Study monitors audit the data, either at individual sites or electronically, to ensure accuracy. When a form contains an entry that is difficult to interpret or obviously mistaken, the monitors send a query back to the investigator or study staff to resolve the problem. Each query has to be explained and resolved before the data are entered into the coordinating center database (Kirwan et al., 2008). At several points in this process, a portion of the data is coded or categorized, and additional checks are performed to make sure the data entry is correct. Sometimes in the process of data entry, additional queries about the data are generated that must be addressed by the original investigator and the study staff.

The term "participant-level data" generally refers to the de-identified records of individual patients generated through this process. De-identification is the process by which personal information that can be used to identify an individual is removed. However, even participant-level data may not capture all relevant information recorded in the raw dataset. For example, Loder described several challenges involved in coding adverse events. Misclassification of adverse events in clinical trials can have serious consequences—as when adverse events like suicidal behavior are coded only as emotional liability—so systems have evolved to minimize this possibility. Adverse events usually are categorized using a predefined hierarchy or organizational system. But the symptoms reported by patients do not necessarily fall into this hierarchy or system. As a result, such symptoms can be interpreted in different ways. Because of this ambiguity, some have argued for access to raw data as reported by patients or researchers on the case report forms before any coding has taken place (Gøtzsche, 2011).

stored in computerized databases to the summary data made available through journals and registries like ClinicalTrials.gov. Data sharing can also occur at many levels. Several of the presenters at the workshop de-

scribed these data-sharing continuums and discussed the benefits and risks of data sharing, based on the degree to which participant-level data are made available to researchers and the public.

In some trials, data are not even made available to individual researchers participating in a multicenter trial. Sometimes, data are released to researchers not associated with the study only if they show a genuine research interest in the question and a track record of research capability. In some cases, data are shared with everyone.

THE USES OF SHARED PARTICIPANT-LEVEL DATA

De-identified patient data have two major uses, observed Deborah Zarin, director of ClinicalTrials.gov at the National Library of Medicine. They can improve transparency, helping to understand the results of an individual clinical trial, including what happened to individuals in the trial, and they can be pooled to discover new things not identified in the individual trials.

Data Sharing to Enable Independent Reanalysis

Steven Goodman, associate dean for clinical and translational research and professor of medicine and health policy and research at the Stanford University School of Medicine, discussed the former use case in the context of ensuring that a study was correctly analyzed and interpreted. Independent reanalysis of data is the basis of reproducible research and can be an extremely difficult task. An example he mentioned was a study of childhood asthma that had 72 different study forms, 109 form revisions, and almost 300,000 records in the database. The original manuscript started with 73 tables and 9 figures and underwent 40 revisions. The published manuscript contained three tables and two figures. “How do we begin from this tiny little slice that we see to begin to work backward and figure out is what they did right?” he asked. While the top tier of journals may have methodologists who can begin to check the chain of scientific custody from protocol to conduct to data to analysis to results, other journals have to rely on peer reviewers to detect problems. The authors of published studies can put additional information on the Web in the form of supplementary material and appendixes, but in reality, checking the accuracy of the results for a study like this is extremely difficult.

In talking about the tools that are needed to ensure that published findings are based on sound data and analyses, Goodman referenced a paper titled "Reproducible Epidemiologic Research" that proposes a standard for reproducibility (Peng et al., 2006). The premise behind that paper is that independent replication of research findings is the fundamental mechanism by which scientific evidence accumulates to support a hypothesis. The authors, therefore, argue that datasets and software should be made available to allow other researchers to conduct their own analyses and verify the published results.

Peter Doshi, a postdoctoral fellow at the Johns Hopkins University School of Medicine, also discussed the application of shared data to credible assessment of clinical trial results. Doshi, however, argued for a broader view of what should be considered clinical trial data. He proposed that detailed records of measurements and analyses, as well as narratives—including descriptions of patient dispositions, study protocols, and even correspondence—are needed to evaluate the quality of published trial results.

Data Sharing for Discovery

Participant-level data from multiple trials also can be combined to learn more than can be derived from the results of a single trial. Elizabeth Loder, clinical epidemiology editor at *BMJ*, observed that although meta-analyses historically have been done using summary-level data, the number of meta-analyses of individual participant data has been growing substantially. Furthermore, meta-analyses done with individual patient data are typically more likely to be able to detect treatment effects that differ across subgroups than meta-analyses done with aggregate data (Riley et al., 2010). These subgroup effects are frequently of great interest to clinical investigators. As Loder said, drawing from the title of an essay by Stephen Jay Gould, "the median is not the message."

THE RATIONALE FOR DATA SHARING

The arguments in favor of sharing can be divided into two broad and overlapping categories, Loder explained. The first category consists of moral and ethical arguments. These arguments point to the necessity of fulfilling obligations to research participants, minimizing known risks

and potential harm from unnecessary exposure to previously tested interventions, and honoring the nature of medical research as a public good. Patients participate in clinical trials based at least in part on the understanding that their data may benefit others, and these benefits are more likely to occur if the data are widely available. Also, unpublished information might in some cases prevent the occurrence of adverse events (Chalmers, 2006). Data sharing may take different forms, from simply publishing the results of research to publicly sharing detailed patient-level datasets. Finally, taxpayers provide a large amount of money to support publicly funded research and expect to have access to the benefits of that research.

The second category consists of practical and scientific arguments. These include detecting and deterring selective or inaccurate reporting of research; enabling the replication of results and potential resolution of apparently conflicting results; informing risk/benefit analyses for treatment options; facilitating application of previously generated data to new study questions; accelerating research; enhancing collaboration; and building trust in the clinical research enterprise. Rob Califf, director of the Duke Translational Medicine Institute, professor of medicine, and vice chancellor for clinical and translational research at Duke University Medical Center, who also spoke during the first session, pointed to the need to resolve results that appear conflicting. Clinicians are not able to interpret conflicting clinical trials data based on looking at the data abstractly without any kind of expert synthesis of information. Only through replication can one sort out whether conflicting results are due to chance or true differences.

Califf went on to describe a "cycle of quality" that can generate evidence to inform patient care (see Figure 2-1). Clinical trials generate knowledge, which is then applied in clinical practice. The measurement of patient outcomes then leads both to clinical practice guidelines that define standard of care and to further clinical trials. At the core of the cycle is measurement and education, which in turn depend on access to data. Box 2-2 describes how this paradigm of cumulatively building and sharing datasets has worked to reduce deaths due to heart attacks by 40 percent.

As an example of the kinds of advances that may be possible, Loder cited the case of a high school student who won $75,000 at the Intel International Science and Engineering Fair. The student cited searchable databases and free online science papers as the tools that allowed him to create his prize-winning entry. "How many collaborators are out there,

who we cannot even imagine at this point, who might make use of the data?" said Loder.

Loder also called attention to the need to build trust in the clinical research enterprise. This trust is at "an all-time low," she said, which is causing a crisis in recruitment for clinical trials (Williams et al., 2008).

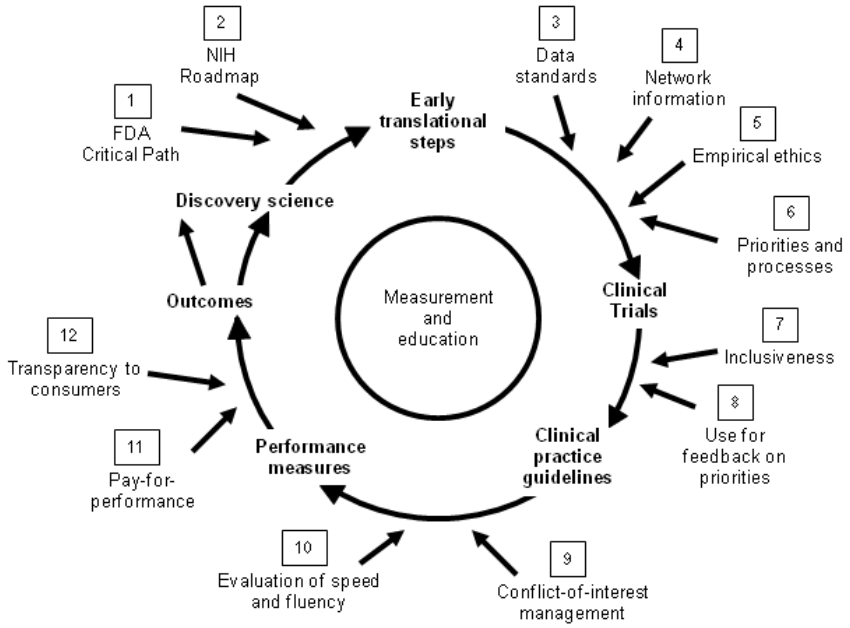

FIGURE 2-1 A "cycle of quality" from discovery science to the measurement of outcomes can generate evidence to inform policy.
SOURCE: Califf et al., 2007.

BOX 2-2
Treatment Advances for Cardiovascular Disease: A Success Story for Data Sharing

The risk of death after a heart attack is now 40 percent lower than it was before the development of medical therapies designed to reduce such deaths (Krumholz et al., 2009) and the development of these therapies relied extensively on clinical trials, said Rob Califf, director of the Duke Translational Medicine Institute, professor of

medicine, and vice chancellor for clinical and translational research at Duke University Medical Center. As an example, he pointed to the Antithrombotic Trialists' Collaboration (2002), which involved 135,000 patients and 287 randomized controlled trials. This study provided compelling evidence that the use of aspirin can reduce deaths from heart attacks. Replication of results from multiple trials has also demonstrated the benefits of fibrinolytics, beta blockers, angiotensin-converting enzyme inhibitors, and other treatments. These studies also showed that particular therapies were more or less useful in different groups of patients and at different times following presentation of symptoms, providing information that then shaped clinical practice guidelines.

Another example involves the effects of statins. By pooling data from multiple trials, it has been possible to show that statins confer benefits regardless of their effects on cholesterol levels (Baigent et al., 2005). In contrast, when data were not released and combined regarding the use of erythropoietin in renal patients who are anemic, the harmful effects of high-dose erythropoietin were overlooked (McCullough and Lepor, 2005). "This could have been detected much earlier if the right trials had been done and the data had been combined," Califf asserted.

The lack of trust extends even to physicians, who tend to discount studies of superior methodological rigor when they perceive that the studies have been funded by industry (Kesselheim et al., 2012). "If doctors do not believe the evidence, what hope is there for evidence-based medicine?" Loder asked.

Sharing data may generate problems that cannot be anticipated today, but it will also generate unanticipated benefits. "We are engaged in one of the great struggles of human knowledge—the struggle to liberate clinical trial information and make sure it is put to its best and highest use now and in the future," Loder concluded. "It is a thrill to be part of this historic meeting."

Commitment to Open Science

Every day, many people face difficult questions about health care, observed Harlan Krumholz, Harold H. Hines, Jr., Professor of Medicine at the Yale University School of Medicine. They need all of the information that is relevant to the options they are considering. If data are

missing, their ability to make informed decisions will be impaired. This is the central argument in favor of open science, Krumholz said.

Krumholz's experience has been that whenever data are shared, whether voluntarily or not, new and important things are learned. In particular, the release of participant-level data has generated vital new information about the risks and benefits of drugs and devices. In some cases, access to this information leads to conclusions that contrast with the prevailing knowledge and changes the use of a drug or device. In other cases, it provides "nuance and understanding." For instance, Krumholz described a study (also described by Loder) which found that unreleased data are about as likely to strengthen evidence for the use of a product as to weaken such evidence (Hart et al., 2012). "What is important is that we support the idea that data are a social good and the best science takes place in the light," he said.

Krumholz shared his vision of a future where data sharing is widely accepted as being in everyone's best interest and will be the cultural norm. "Data sharing [will be] an essential characteristic of being a good scientist and a good citizen," he said. With the full release of data, companies would compete on the basis of science, not marketing. Academic researchers could get credit not only for the papers they publish, but for the knowledge generated from the databases they create.

Industry has the opportunity to demonstrate leadership, restore trust, and reclaim its position of integrity through meaningful actions to share data, Krumholz continued. "You have a meaningful motivation," he said. "The [medical] profession has less trust in your science than in [National Institutes of Health]-sponsored studies and is less likely to act on the results of the trials you sponsored, not just the ones you conduct. The pharmaceutical and device industries no longer have the respect they once held. . . . The result is a situation that does a disservice to the public, the medical profession, and the vast majority of professionals in industry who have extraordinarily high integrity and are in that industry for the right reasons."

Krumholz noted that an important cultural shift is already taking place. Some industry leaders have already taken steps to support data sharing and have contributed to major scientific advances as a result. For example, Medtronic's decision to release the company's data on a product that has nearly a billion dollars in annual sales was a powerful statement that the company was seeking the truth. The individuals who have made these decisions "realize that studies are only possible due to the generosity of people who consented to participate, and that we have an

obligation to ensure that the efforts of those subjects contribute as much as possible to knowledge generation." Such transparency will also be essential to ensure the continuing flow of individuals who are willing to participate in trials, Krumholz added.

In return for the privilege of selling a medical product to the public, industry bears a responsibility to ensure that all the data concerning the risks and benefits are available to everyone, said Krumholz. The current challenge is not to decide whether data should be released, but how to do so while being attentive to the needs and concerns of all stakeholders. In addition, the publication of summary results is not enough, according to Krumholz. Rather, individual patient-level data need to be broadly and freely available for investigators. "We need the protocols and case report forms. We need full sharing of the source data. . . . With the talent in this room, and with those listening on the webinar and those who are interested, I know solutions can be found. If we are committed to the path, we can figure out how to do it."

CAUTIONS ON DATA SHARING

Jesse Berlin, vice president of epidemiology for Janssen Research & Development, LLC, provided a countervailing view by asking whether participant-level data are always needed. Complications can arise when the data are reexamined, he said. Decisions may have been made during a clinical trial that cannot be replicated. Published studies may not always incorporate the appropriate intent-to-treat analysis. Endpoints may be defined differently in different trials. Study designs, patient populations, and treatments can vary from trial to trial. As a result of these and other potential problems, such analyses can go "seriously wrong," Berlin warned. "It is not just a matter of feeling more comfortable having the individual-level data. You can actually get wrong answers."

Although there is a common belief that participant-level data can enable verification and reproduction of trial results, that premise is reliant on the trustworthiness of the shared data, warned Peter Doshi. Even participant-level data can lead investigators astray. For example, a computerized database of participant-level data may not reflect what is actually recorded on a case report form. In some cases, it may be necessary to look beyond what people typically consider data (i.e., numbers) into more narrative forms of documentation depending on the intended use of the shared data.

3

Barriers to Data Sharing

Key Messages Identified by Individual Speakers

- The public expresses a strong demand for biomedical innovation, but privacy issues are also a concern and regulations designed to protect the privacy of personal health information can impact data sharing.
- Incentives in academia to keep data private for purposes of professional advancement can hinder data sharing, but new models for research allow competition to continue in more open environments.
- Means of ensuring the quality of secondary analyses of shared data prior to publication would help ease concerns, particularly for those in industry, regarding misuse of shared data. The development of trust relationships, between patients and researchers but also among researchers, can enable progress that contractual agreements cannot achieve.
- Policies that mandate data sharing are often not observed.
- Technical challenges to data sharing may be more easily addressed by making arrangements to share participant-level clinical data and implementing data standards at the outset of a trial.

Despite the widely acknowledged benefits to be gained by sharing research data, significant barriers to sharing remain to be addressed. Several speakers described these barriers, and others mentioned them in

passing. Some barriers, such as the need to maintain patient privacy, are common to all organizations, but others apply more strongly to some organizations than others. For example, researchers in private industry (and their partners or potential partners in academia) are more concerned about protecting proprietary information, while academic researchers are more interested in keeping data under their control to generate publications and professional acclaim. These differing incentives complicate efforts to disseminate data more widely, as described in the chapters on models for sharing clinical data (Chapter 4), standardization (Chapter 5), and changing in the culture of research (Chapter 6).

LEGAL ISSUES RELATING TO PATIENT PRIVACY

Jennifer Geetter, a partner at McDermott Will & Emery, talked about patient privacy considerations and regulations as barriers to data sharing. Despite a desire to see innovation and progress in biomedical research, the public remains very concerned about a potential loss of privacy. However, the exact nature of these concerns is not well understood, Geetter observed. Are people afraid their personal information will be used in ways they do not intend? Are they afraid their information will be used in ways that have adverse employment or insurance consequences? "It is difficult to know," said Geetter, "but everyone out there perceives a real privacy concern."

Geetter described two opposing approaches to privacy protection. The first holds that access to information should be very restricted with a presumption of nondisclosure. For example, the classic doctor–patient relationship assumes this model, which obviously is a significant impediment to data sharing. The second model balances confidentiality with socially useful sharing and disclosures of data using generally accepted rules for doing so. The public is ambivalent about which of these models to adopt, said Geetter. Despite this ambivalence, Geetter encouraged researchers to involve patients in data-sharing decisions. When patients are asked to share their information, are given a voice in that process, and are thanked for their participation, they are more likely to choose to share health information, she said.

A good deal of legal uncertainty exists regarding privacy regulations and the associated impacts on data sharing, said Geetter. At the time of the workshop, several regulations were being developed and revised that affect data sharing. Release of modified rules promulgated under the

Health Insurance Portability and Accountability Act (HIPAA),[1] which protect the privacy of individually identifiable health information, was imminent and, Geetter noted, will influence the sharing of medical information for research as well as health care. Also, the Common Rule,[2] which provides protection for human research subjects, was undergoing "a massive upgrade for the 21st century," according to Geetter. For example, a proposed change to the rule dealt with whether human biospecimens could ever be considered de-identified, which will have a substantial effect on research. To further complicate matters, there is a lack of harmonization among regulations. For example, Food and Drug Administration (FDA) regulations, the Common Rule, and HIPAA do not agree on aspects of data sharing. Furthermore, state-by-state rules exist that tend to be disease specific, and these can have provisions related to data sharing.

Another uncertainty mentioned by Geetter related to the implications of the FDA's "Part 11" rule,[3] which covers the submission of electronic data to the FDA. Parts of the rule are currently enforced, but other parts are not. This affects data sharing because most data sharing involves electronic data, which is likely to be included in future FDA submissions.

The "preparatory to research exception" is a HIPAA pathway that allows sharing of protected health information (PHI) without an individual's authorization in order to prepare for a clinical trial. However, data sharing using this pathway is limited by a restriction preventing PHI from leaving the premises of a covered entity. "This made sense when you were looking at dusty paper medical records," Geetter noted. "It does not make as much sense, in my view, when you are talking about electronic data that may never be at the covered entity's premises to begin with" because the data may be in an electronic health record held by a vendor who uses cloud computing services.

Geetter concluded by observing that the line between clinical care and research is growing progressively more blurred, especially as electronic health record systems become more common. Opportunities for systematic data collection across institutions will become more plentiful as these systems become increasingly interoperable. Data generated during clinical care, in addition to that resulting from research studies, may become available for mining. However, competing public priorities will likely place limits on the ability of that data to be viewed and shared. For

[1] 45 CFR 160, 164.
[2] 45 CFR 46.
[3] 21 CFR 50, 56.

example, physicians may be encouraged to use that data to troubleshoot the care they provide, but privacy advocates may oppose making it available for discovery purposes.

CULTURAL BARRIERS

Need for Incentives

Andrew Vickers, attending research methodologist in the Department of Epidemiology and Biostatistics at the Memorial Sloan-Kettering Cancer Center, has experienced firsthand the reluctance of investigators to share clinical research data; he described several such instances. In one case, data from the control arm of a trial funded by the National Institutes of Health (NIH) was needed to help design a new study, but the response to his request was "I am not prepared to release the data at this point." Another time, when data were needed for a meta-analysis of results from trials involving chemotherapy, he was told, "I would love to send you these data but my statistician won't allow it." In a third case, Vickers was developing a statistical model for predicting response to treatment for specific patient populations and needed access to a dataset from another NIH-funded trial. Despite providing the investigators with numerous reassurances, including that the data would be used only for a statistical methodological study, that the paper would explicitly state that no clinical conclusions should be drawn from the analysis, that the data would be slightly corrupted, and that he would send a draft of the paper to the investigators and give them veto power, "We never heard back from them," he said.

Among the explanations Vickers received for refusal to share data was the cost and trouble of putting datasets together, typified by the comment, "I would love to help you, but it would take too much time." Vickers labeled this argument as unacceptable. In the case of the 10 papers they surveyed, the authors had just published results based on the requested dataset. He questioned how authors could publish a study without having a clean and well-annotated dataset, which could easily be distributed on request. "You have to do this anyway, right?" he said.

Other arguments against sharing data had more validity, Vickers acknowledged. For example, career advancement in an academic setting depends on the ability of investigators to generate publications from data they might have spent years collecting. This concern was raised repeat-

edly by speakers throughout the workshop as a major barrier to sharing for those working in academia.

In response to his frustrations with the current problems with data access, Vickers wrote an essay in the *New York Times*, drawing attention to this cultural barrier and pointing out the moral obligation that researchers have to share data that has been collected from patients who volunteered to participate in clinical trials, in part, for the benefit of future patients. In addition to the essay, Vickers also helped publish a study on data sharing by authors who publish in *Public Library of Science* (*PLoS*) journals. Despite the journals' data-sharing policies, which require authors to share their raw data, only 1 of the 10 requested datasets was received (Savage and Vickers, 2009), indicating that additional incentives or enforcement mechanisms are necessary to change the culture surrounding data sharing.

John Ioannidis, C.F. Rehnborg Chair in Disease Prevention at Stanford University, also discussed the inadequacy of current incentives and policies promoting data sharing. A study by Ioannidis and his colleagues on data-sharing policies at the 50 scientific journals with the highest impact factors found that most have policies in place for sharing data and making the data available (Alsheikh-Ali et al., 2011). However, when the authors sampled 10 papers from each journal, they found that few authors had actually deposited the data summarized in the paper. "Even though there is a lot of interest and a lot of investment in trying to make data sharing work, in practice we still have some ways to go," Ioannidis said.

Fears Regarding Misuse of Shared Data

Another common concern raised during the workshop, particularly by those in industry, was how clinical data will be reused once they become more accessible. One fear was that data will be misused or misinterpreted if, for example, too little attention is paid to how the data were collected and analyzed or to the nature of the patient population. Such misinterpretations can be published outside the peer-reviewed literature so that standard quality controls do not apply. Ensuring that secondary analyses of data are done responsibly is an important issue, said Michael Rosenblatt, executive vice president and chief medical officer of Merck & Co., Inc., because in many cases the information that comes out of a secondary review will not be subject to the same kind of scrutiny and peer review as was done the first time. But, said Elizabeth Loder, *BMJ*,

in the long view, "we should have confidence in the fact that eventually science is self-correcting." It may take a long time, during which multiple and competing analyses coexist. "The anxiety that we feel about many different people looking at the data and coming up with different interpretations is somewhat misplaced. I do not think we should be afraid of that future."

Other workshop participants worried that researchers will not have the time to rebut all misconceived analyses. Incorrect information can have extremely harmful results, observed Robert Califf, Duke University Medical Center. "There will be consequences of people being killed by poor use of data because, if it hits the news, a lot of people will stop taking their medications . . . based on what is in the news," he said. "You can kill people with bad science very quickly. It is a problem that we are going to have to grapple with."

Building Trust

Kelly Edwards, acting associate dean at the graduate school and associate professor in the Department of Bioethics and Humanities at the University of Washington, described some of the cultural barriers to greater data sharing—along with several ways of overcoming those barriers.

Trust is the foundation for all productive relationships and is at the heart of making data-sharing efforts happen. Researchers and organizations must be able to trust each other and participants must be able to trust those same researchers and organizations. But, she asked, how does one build trust? In research, institutions have relied heavily on contracts to help manage trust relationships. For example, consent forms provide exhaustive detail about expectations and obligations for participants in clinical trials. Data-use agreements, terms of use, and other contracts provide differing ways of ensuring ethical management of research.

These arrangements may be necessary, but they are insufficient, said Edwards. Trust also needs to be relational, with contracts serving as a way to punctuate what has already been agreed to rather than the sum total of how a relationship will work. Different elements enter into relationship trust. In some cases, people share core values and interests or are committed to a common cause. Someone may have another person's best interests in mind, as in the doctor–patient relationship. Relational trust can be built on transparent and consistent (or logical) rules, and trust can

depend on associations. "When someone else I trust trusts you, I can also get in the door."

Reemphasizing a point made by several other participants, Edwards noted that the promotion system in academic settings, which relies on individual credit for grants awarded and papers published, interferes with the establishment of trust relationships and can be antagonistic to data sharing. However, Edwards pointed to several developments that can allow rewards, acknowledgment, and attribution to coexist in a more open research system. The old models of medical research where data are kept close to our chest are beginning to crumble, she said, and new models can be built that still promote competition, but in a more open research space. Other industries are helping to drive more open systems and the democratization of data, such as the information technology industry's move toward mobile and cloud computing. "This culture shift has happened already in other fields," said Edwards. "Let's move it into health research."

One way to encourage openness is to stay grounded in traditional research ethics. The 1978 Belmont Report, subtitled *Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, referred to respect for persons, beneficence, and justice (HEW, 1979). These principles also can be applied to the emergence of more open research systems. For example, respect for persons can be embodied in both partnerships and communication. As a concrete example, Edwards mentioned the simple step of thanking research participants. "I am taking an informal poll of researchers I work with on how often they just say thank you to the participants who are in their studies, and it is embarrassingly low numbers," she said. "Simply saying thank you can go such a long way."

Regulations provide a minimum standard for behavior, Edwards said, and researchers need to do more than just what regulations mandate. Thus, data-sharing policies can provide a scaffolding, but the research community needs to set standards of excellence and strive to meet those standards.

TECHNICAL CHALLENGES

Planning for sharing participant-level data at the outset of a study is important, Ioannidis pointed out. A more difficult, or impossible, task is to unearth data after the paper describing those data has been published. Far

better, he said, is to arrange upfront for the full individual-level data to be available. Issues such as coding, cleaning, and logical queries take some time to resolve, but their difficulty is probably overrated. Post-hoc efforts to standardize data can also prove challenging and costly and may limit the usefulness of such data (see Chapter 5). "If we wait to see what we can do after the fact, it is very difficult." He was involved in one study in which the investigators sought to repeat the analyses of microarray expression studies published in *Nature Genetics* using the datasets deposited with the papers (Ioannidis et al., 2009). Four independent teams of microarray analysts could reproduce only 2 of the 18 tables and figures from the papers. Much of the time, key information was not available, despite a precondition to publication in the journal that data be made available to independent investigators.

Some fields have adopted strong principles of data sharing. One of the best examples is the field of human genomics, which has principles on how to share information among all investigators working in the same area and, in some cases, with other investigators and the public. Without those standards in places, said Ioannidis, the "fantastic growth" in the field of human genomics would have occurred much more slowly. "Clinical trials could learn from such paradigms and try to adopt them."

4

Models of Data Sharing

Key Messages Identified by Individual Speakers

- Registration of clinical trials and summary trial results has been a major step forward, but ambiguous protocols and discrepancies between protocols and results raise concerns about the integrity of clinical research data.
- Greater transparency of study protocols and amendments, statistical analysis plans, informed consent forms, clinical study reports, and adverse event reports would both improve clinical trials and facilitate sharing of trial results.
- The de-identification process can be complicated and expensive when studies are not designed with data sharing in mind.
- Collaborations need to be clear about common goals, realize the unique value each party brings to the effort, and strive for open inclusiveness.
- Companies can be fierce competitors, but still cooperate on precompetitive research to meet common needs.
- If patients provide information for a research project, they should receive information in return that can help them make meaningful health care decisions.
- Treating patients as partners in research would acknowledge their expertise in managing and understanding their conditions.

Clinical trial data are a public good, but many stakeholders in addition to the public have interests in those data, observed Jeffrey Nye, Janssen Research & Development, in his introduction to the session on models of data sharing. Participants in a trial have interests in the information a trial generates, as do the researchers conducting a trial. Pharmaceutical companies are another stakeholder, along with researchers from either the private or public sectors doing reanalyses or meta-analyses of study data. Regulators have the objective of safeguarding public health and guiding and advising companies as they develop new products, while citizen scientists may be studying the data to derive information they can apply in their own lives.

Seven speakers at the workshop described different models designed to increase the sharing of clinical research data. All of these models have strengths and limitations. Although the optimal path forward is not yet clear, all of these models offer lessons that can inform future initiatives.

CLINICALTRIALS.GOV

Three key problems interfere with the practice of evidence-based medicine, said Deborah Zarin, director of ClinicalTrials.gov at the National Library of Medicine, National Institutes of Health (NIH). Not all trials are published. Publications do not always include all of the prespecified outcome measures. Unacknowledged changes made to trial protocols can affect the interpretation of findings.

These problems led to the establishment in 2000 of ClinicalTrials.gov, which serves as a registry of clinical trials at the trials' inception (Zarin et al., 2011). The registry now contains key protocol details of more than 130,000 trials from around the world. In 2008 the registry added a results database, which now contains the summary results of more than 7,000 trials. ClinicalTrials.gov does not accept participant-level data, Zarin emphasized, but it has considerable experience with other kinds of data generated by clinical trials.

Clinical trials data take many forms, from uncoded, participant-level data to analyzed summary data; only the latter are posted at ClinicalTrials.gov. At each step in the process leading from the raw data to the summary data, information is lost (see Figure 4-1). Also, each vertical drop involves subjective judgments that are not transparent, but can

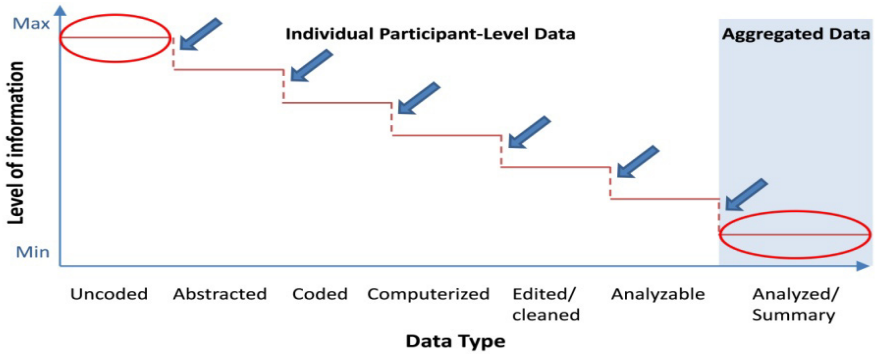

FIGURE 4-1 Information loss as clinical trials data progresses from raw uncoded data to summary data.
SOURCE: Zarin, 2012. Presentation at IOM Workshop on Sharing Clinical Research Data.

influence the reproducibility of results. The users of summary data generally assume that they reflect the underlying participant-level data, with little room for subjectivity. That assumption is not always correct, said Zarin.

The results database at ClinicalTrials.gov was launched in response to the Food and Drug Administration Amendments Act of 2007 and was based on statutory language and other relevant reporting standards. It requires that the sponsors or investigators of trials report the "minimum dataset," which is the dataset specified in the trial protocol in the registry. The data are presented in a tabular format with minimal narrative. They cover participant flows, baseline patient characteristics, outcome measures, and adverse events. The European Medicines Agency is currently developing a similar results database.

Although ClinicalTrials.gov has checks for logic and internal consistencies, it has no way of ensuring the accuracy of the data reported. ClinicalTrials.gov does not dictate how data are analyzed, but does require that the reported data make sense. For example, if the participant flow had 400 people and results are presented for 700, it asks the trial organizers about the discrepancy. Similarly, time to event must be measured in a unit of time, and the mean age of patients cannot be a nonsensical number like 624. "That is the kind of review we do," Zarin said.

ClinicalTrials.gov was established on the assumption that required data are generated routinely after a clinical trial based on the protocol for

the trial, so the burden of reporting to ClinicalTrials.gov would be due mainly to data entry. Instead, the experience at ClinicalTrials.gov has shown that protocols are often vague, are not always followed, or in some cases may not even exist. In addition, summary data are not always readily available even for trials that have already been published. For many trials, no one can explain the structure of the trial or the analysis of the data, said Zarin. "What we learned is there is not an objective, easy-to-describe route from the initial participant-level data to the summary data. Many people and many judgments are involved."

Structural changes to trials are also common. A trial can start as a two-arm study and then become a four-arm study. Participants come and go, so that the number of participants changes over time. Participant flow and baseline characteristic tables describe different populations than the outcomes table. Data providers often cannot explain the "denominators" for their results, the groups from which outcomes or adverse events are collected. Zarin described a study in which a year of close work was required with statisticians to figure out who the people in the study were and where they went as a result of structural changes to the study. "These are brilliant statisticians. They were in charge of the data. [But] this trial was basically too complicated for them to figure out. They were giving outcome measures without actually knowing what the denominators were. That is one kind of problem we have seen."

In other cases, outcome measures were changed: a quality-of-life scale was replaced with a depression scale; 1-month data were replaced with 3-month data; the number of people with an event was replaced with time to an event; and all-cause mortality was replaced with time to relapse. Sometimes discrepancies are obvious. In one study, the mean for hours of sleep per day was listed as 823.32 hours. Another study of 14 people included data on 36 eyeballs. "As a consumer of the medical literature, these are not reassuring things," Zarin observed.

In a study of 100 matched pairs of ClinicalTrials.gov results and publication results, 82 percent had at least one important discrepancy. The inevitable conclusion is that summary data may not always be an accurate reflection of participant-level data. Although the deposition of clinical trial protocols and summary data into registries is a huge step forward in the direction of transparency, the validity and reproducibility of summary data are called into question by such inconsistencies. "This is a big problem," Zarin asserted.

Providing more transparency about the process of converting one type of data into another type would help inspire trust, she said. Docu-

ments that may help explain this journey include the protocol and amendments, the statistical analysis plan, informed consent forms, clinical study reports, and adverse event reports. Greater transparency would also help everyone involved with clinical trials to engage in internal quality improvements.

THE DATASPHERE PROJECT

In contrast to the declining mortality rates for heart disease (see Box 2-2), mortality rates for cancer have dropped only slightly in recent decades, noted Charles Hugh-Jones, vice president and head of Medical Affairs North America for Sanofi Oncology. Changes in risk behaviors, an increase in screening, and new therapeutics have all contributed to this decline in cancer, "but we are not being as effective as we would like to be." At the same time, the price of cancer treatment has skyrocketed, which is not sustainable in an era of fiscal austerity. We need to find better ways of reducing cancer mortality rates, said Hugh-Jones, and "one of the solutions of many that we need to address is data sharing."

Data sharing in the field of oncology could lead to faster and more effective research through improved trial designs and statistical methodology, the development of secondary hypotheses and enhanced understanding of epidemiology, collaborative model development, and smaller trial sizing, said Hugh-Jones. For example, as oncology researchers divide cancers into smaller subgroups with particular molecular drivers, data increasingly need to be pooled to have the statistical power to determine the most effective treatments for each subgroup.

Hugh-Jones described an ideal data-sharing system as simple, systematic, publicly accessible, and respectful of privacy issues. DataSphere, which is an initiative of the CEO Roundtable on Cancer, is designed to achieve these objectives. The CEO Roundtable on Cancer consists of the chief executive officers (CEOs) of companies involved in cancer research and treatment who are seeking to accomplish what no single company can do alone. DataSphere will rely on the convening power of CEOs, together with support from patients and advocacy groups, to secure and provide data. Initially, it will seek to provide comparator arms, genomic data, protocols, case report forms, and data descriptors from industry and academia. DataSphere will include data from both positive and negative studies because a negative study is often as revealing from an epidemiological point of view as a positive study. De-identification

will be standardized, and DataSphere will then work with third-party data aggregators to pool the data in meaningful ways—a significant challenge when hundreds of cancer drugs are being developed at any given time and thousands of studies are registered in ClinicalTrials.gov.

At the outset, said Hugh-Jones, the originators of DataSphere asked three questions. Why would people want to share their data? If I wanted to share my data, how would I do it? Finally, where would I put it once it was ready to post? DataSphere has established incentives for data contributors that call attention to the increased productivity, cost savings, citations, and collaboration that can accompany sharing. It also is looking at micro-attribution software that could extend credit for sharing to the contributors of data. Similarly, incentives for patients emphasize the benefits of making data available and the security precautions that have been taken. It has even been looking into the possibility of competitions among researchers to enhance the sharing of data.

Tools to enable sharing, continued Hugh-Jones, include a standard de-identification system being developed in collaboration with Vanderbilt University that is consistent with Health Insurance Portability and Accountability Act (HIPAA) regulations, a single online data use agreement form, how-to guides for de-identification, and tools for advocacy. Finally, it has been working closely with the database company SAS to produce a simple but secure, powerful, and scalable website where everything needed to share data is automated.

Sanofi is contributing de-identified data from two recent Phase III clinical studies to start the ball rolling. The goal, said Hugh-Jones, is to have at least 30 high-quality datasets in the database by the end of 2013 and then expand beyond that. “With the sort of environment we have demonstrated here, this is something that can be successful.”

THE YALE-MEDTRONIC EXPERIENCE

One paradigm for facilitating dissemination of industry data and ensuring high-quality independent review of the evidence for efficacy is exemplified by the Yale-Medtronic experience, as described by Richard Kuntz, senior vice president and chief scientific, clinical, and regulatory officer of Medtronic, Inc., where proprietary data were released to an external coordinating organization that contracted other organizations to perform systematic reviews of the study results.

In 2002, according to Kuntz, the Food and Drug Administration (FDA) approved a product from Medtronic called INFUSE, which was designed to accelerate bone growth in cases of anterolateral lumbar interbody fusion. Approval was based on one pilot randomized controlled study and two pivotal randomized controlled studies. A series of subsequent peer-reviewed publications supported by Medtronic provided additional data on the use of the product.

In June 2011, Kuntz continued, a major challenge was raised regarding the validity of all the published literature on INFUSE. The principal focus was on the results presented in the peer-reviewed literature and on general study designs and endpoints. The challenge was published in a dedicated issue of a medical journal and consisted of more than 10 articles. The company quickly reviewed its data to ensure that the dossiers it had were accurate. "We are convinced that the data were good, and talked to the FDA immediately to make sure that they felt the same." However, the issue was being discussed extensively in the media. "We had to make some quick decisions," said Kuntz.

Within less than a month, Kuntz said, the company announced its decision to contract with Yale University as an independent review coordinator. In August, Yale announced its plan to establish an independent steering committee and contract with two systematic review organizations to carry out reviews of the research. Medtronic agreed to supply Yale with all de-identified patient-level data, including non-label studies, along with all FDA correspondence and adverse event reports. It also agreed to allow Yale to establish a public transparency policy and process for the entire INFUSE patient-level dataset. The publication of the systematic reviews was scheduled for the fall and winter of 2012, with summary manuscripts prepared and submitted for publication in the *Annals of Internal Medicine* at the time of the workshop.

The project has been undertaken by the Yale University Open Data Access (YODA) project, which, according to Kuntz, serves as a model for the dissemination and independent analysis of clinical trial program data. This project is based on the rationale that a substantial number of clinical trials are conducted but never published, and even among published clinical trials, only a limited portion of the collected data is available. As a result, patients and physicians often make treatment decisions with access to only a fraction of the relevant clinical research data. Clinical trials are conducted with both public and private funding, but several issues are particularly important among industry trials. Industry funds the majority of clinical trial research on drugs, devices, and other products,

both premarket and postmarket. Also, industrial research is proprietary, with no requirement for publication or dissemination, and the public perception is that industry has a financial interest in promoting "supportive" research and not publishing the rest of the data.

The YODA project has been designed to promote wider access to clinical trial program data, increase transparency, protect against industry influence, and accelerate the generation of new knowledge. The public has a compelling interest in having the entirety of the data available for independent analysis, but industry has legitimate concerns about the release of data, Kuntz said. Steps therefore are needed to align the interests of industry and the public, particularly when concerns about safety or effectiveness arise.

Yale and Medtronic spent a year working through issues involved in assembling the data and giving those data in the most unbiased way possible to reviewers so they could do a full systematic review. To maintain transparency and independence, formal documentation of communications between Yale and Medtronic was necessary along with clarity about what kinds of discussions could and could not be held. For example, Kuntz said, Medtronic did not want to send Yale previous reviews or interpretations of the data done by outside groups because the company did not want to taint the information. The query process among the reviewers, Yale, and Medtronic also had to be carefully managed.

The de-identification process was complicated and expensive. De-identifying the necessary HIPAA fields and information took several months and the efforts of about 25 people, which contributed substantially to the overall $2.5 million cost of the project. The HIPAA Privacy Rule was not designed for this kind of activity, Kuntz observed. As a result, the YODA project's approach to de-identification was a "Rube Goldberg contraption" and clearly not scalable. Given that paper case report forms and studies going back to 1997 had to be reviewed, the project was "an outlier example of how complicated it would be to de-identify [data]."

Industry has several reasons for participating in this kind of process, according to Kuntz. It allows fair and objective assessment of product research data, as opposed to speculative analysis based on incomplete data. It supports competition on the basis of science rather than marketing. It promotes transparency and advances patient care. Although committed to transparency, Medtronic was concerned about potential misuses of the data. For example, is everyone seeking access to the data interested in the truth? Litigant firms may be interested in making money, "but

litigant firms also can find the truth," said Kuntz. In the end, Medtronic sought to provide the data and initiate conversations about its use.

However, Kuntz raised a large number of questions that the Yale-Medtronic project has not fully answered:

- Would it be possible for an independent group to determine whether a question requiring the use of data serves the public interest or a special interest?
- Should queries be limited to single questions, and should the methods used to answer the questions be prespecified?
- Should there be an initial time period during which data remain proprietary?
- What portion and level of the dataset are necessary?
- Should there be a time limit or license for data access?
- Who controls the data distribution?
- Are there a priori questions and hypotheses to be tested, or is there an interest in data exploration?
- Is the requester competent to do the proposed analysis?
- Should a trusted third-party analysis center be contracted? May the requester share the data with others?
- Should there be controls on the dissemination of results, such as a requirement for peer review before dissemination?
- What methodological review is required?
- Should industry be involved in the peer review of results derived from its data?

All of these questions need better answers than exist today, said Kuntz. Nevertheless, the bottom line is that industry has a responsibility to do studies with regulatory agencies to produce results in a faithful and trusted way and to disseminate them under the law. It needs to competently and ethically contract or execute the required clinical studies and perform timely filing of the data and results dossier. Industry makes products that "we sell to people," said Kuntz. "We are responsible for the health of those individuals."

The movement from keeping data concealed to sharing data will require foundational changes, Kuntz concluded. One important step will be involving patients as partners rather than "subjects," which will help lower at least some of the barriers to the use of data.

THE BIOMARKERS CONSORTIUM

The Biomarkers Consortium of the Foundation for the National Institutes of Health (FNIH) is a precompetitive collaboration designed to increase the efficiency of biomarkers-related research. Its goals are to facilitate the development and validation of new biomarkers; help qualify these biomarkers for specific applications in diagnosing disease, predicting therapeutic response, or improving clinical practice; generate information useful to inform regulatory decision making; and make Consortium project results broadly available to the entire scientific community.

John Wagner, vice president for clinical pharmacology at Merck & Co., Inc., described the validation of adiponectin as a biomarker as an example of the work of the Consortium. Adiponectin is a protein biomarker discovered in the 1990s that is associated with obesity and insulin sensitivity. Certain drugs can drive up adiponectin levels very quickly in healthy volunteers and in patients, and attention was focused on the use of adiponectin as a predictive biomarker to identify patients who would or would not respond to particular therapies.

Though considerable data about adiponectin existed in the files of companies and academic laboratories, relatively few data about the use of adiponectin as a biomarker were publicly available. The Biomarkers Consortium took on the task of compiling these data as a proof-of-concept project for the collaboration. A number of companies agreed to combine their data into a blind dataset derived from many trials involving more than 2,000 patients. Using these data, the consortium concluded that adiponectin is a robust predictor of glycemic response to peroxisome proliferator–activated receptor agonist drugs used in the treatment of diabetes. The results confirmed previous findings and investigators concluded that "the potential utility of adiponectin across the spectrum of glucose tolerance was well demonstrated" (Wagner et al., 2009).

Wagner drew several important lessons from this experience. The project demonstrated that cross-company collaboration was a robust and feasible method for doing this kind of research. However, the project took a relatively long time to complete, which is a real problem, according to Wagner. The Consortium has since learned how to collaborate more efficiently, but time remains a concern. The pace was set based on the amount of time team members had to dedicate to this project. The Consortium was not the first priority of everyone involved in the project. "It was the evening job for many people, myself included." Good project

management skills have helped to address this problem, as has the development of new collaboration tools.

The Consortium struggled with data-sharing principles and standards, Wagner admitted. Negotiating a data-sharing plan with even a small number of companies was challenging and having a single legal liaison for each of the companies was found to be critical. Standard definitions were not all obvious. In some cases, people would fail to pass on crucial information before leaving for another position. However, in the end the project created a template for the Biomarkers Consortium for data-sharing plans, which should speed the work in subsequent projects. Also, FDA currently has an initiative to require uniform data submissions using standardized data fields, which would result in data that are much more amenable for sharing, Wagner observed. Furthermore, health care reform is also expected to harmonize data practices, in part to reduce costs and improve care.

The existing data had many limitations, Wagner indicated. The original studies were not designed to answer the research question investigated by the Consortium. The adiponectin data also had limitations because different companies used different assays to measure the protein, which required more work to ensure that the data could be combined reliably.

Broader issues also arose. The clarity of the research question is very important for defining the type of collaboration. The existence of a neutral convener—in this case the FNIH—was critical in gaining the trust of all the stakeholders involved in the project. Still, motivations were an issue. Depending on the question being asked, the openness of the contribution and of the output can change. In the case of the Biomarkers Consortium, the output is completely open, which is a good model for generating new knowledge. The nature of the collaboration also depends on whether it is developing standards and tools, aggregating data, creating new knowledge, or developing a product, Wagner said. Collaborations depend on trust and openness. Being clear about common goals, realizing the unique value each party brings to the effort, and striving for open inclusiveness can greatly improve collaborations.

THE NEWMEDS CONSORTIUM

NEWMEDS, which is a project sponsored by the European Union, stands for Novel Methods for Development of Drugs in Depression and Schizophrenia. As discussed by Jonathan Rabinowitz, academic lead of

NEWMEDS at Bar Ilan University, the NEWMEDS consortium was established to facilitate sharing of clinical trials data—in particular, coded participant-level data—from industry and academia to examine research questions in the precompetitive domain. According to Rabinowitz, the schizophrenia database, which includes data from AstraZeneca, Eli Lilly, Janssen, Lundbeck, and Pfizer, encompasses 64 industry-sponsored studies representing more than 25,000 patients, along with studies sponsored by the National Institute of Mental Health and the European Union. The depression database, with data from several of the same companies, includes 26 placebo-controlled, industry-sponsored studies covering more than 8,000 patients.

Rabinowitz went on to describe some of the major findings and lessons learned from the schizophrenia database. When looking at patient response, analysis of the database revealed that results at 4 weeks were nearly the same as at 6 weeks, implying that studies could be shorter. Females show more pronounced differentiation between placebo and active treatment than males. Thus, the inclusion of more females in studies, previously underrepresented, could show heightened differences from placebo. Patients with a later onset of disease showed more pronounced improvements, irrespective of their allocation to active treatment or placebo groups, but differentiation from placebo was not affected by age of onset. For unknown reasons, the active-placebo differentiation varies by geographical region, with considerably more differentiation in Eastern Europe than in North America. All of this information, which is useful in its own right, can be used to design more effective and efficient clinical trials with smaller treatment groups and shorter study durations, Rabinowitz stated, which together could significantly reduce costs of drug discovery trials.

Rabinowitz described some of the lessons learned from his personal experiences with the Consortium. Just locating the data was a challenge. It might sound mundane, but it can be very complex, he said. For example, companies are bought and sold, and products are exchanged among companies. “To locate who houses data [required] almost the work of a detective.” Also, competing internal resources and priorities mean that data sharing is not necessarily the top priority. Compared with the YODA project’s experience, de-identification was much less expensive and time consuming, said Rabinowitz, requiring about 2 weeks of programming time. In the context of the amounts spent on clinical trials and the potential markets for new products, though, even rather expensive de-identification projects can be justified. The formulation of research ques-

tions and interpretation of data also need to be the result of active collaboration so that understandings are shared as well as data.

Rabinowitz talked about the increasing difficulties of drug discovery as incentive for companies to collaborate through precompetitive challenges. These companies can be fierce competitors elsewhere, but they have common needs. Companies also need to send a clear message of support for collaboration to overcome various kinds of resistance, with ongoing support from the top levels of management. Previous relationships can be very helpful because they help foster the trust that companies need to provide data to a collaborative effort. Peer pressure among companies aided data sharing, in that "if one company [provided] all their data, the others wanted to follow suit. They did not want to feel inferior in terms of their performance."

A paradigm shift is occurring that redefines data sharing as an "ethical imperative," Rabinowitz concluded. Studies should be given extra credit if they are willing to share data. This could be taken into account by institutional review boards (IRBs), for instance, in judging the ethical validity of a study. "Allow yourselves to imagine what you might do in some therapeutic area that is near and dear to you if you had access to almost all of the data out there in your given area," he said. "Just think about that for a second."

PATIENTSLIKEME

PatientsLikeMe is a health information–sharing website for patients where they can form peer-to-peer relationships, establish profiles, provide and share health data, and make de-identified data available for research. Sally Okun, health data integrity manager at PatientsLikeMe, described some of the lessons learned from the website during its 7 years of operation.

A prominent mandate of the site is "give something, get something." If patients provide information for a research project, they should receive information in return that can help them make meaningful decisions, said Okun.

Another motto is "patients first." In a data-sharing environment, the interests of the patients need to come first, Okun said. "They have a lot more skin in this game than any of us in this room do. . . . They have the expertise in managing [their conditions] that as clinicians and as researchers we could never have."

That observation leads to a third mandate: Listen well. Patients want to share their information. When patients were asked in a recent survey whether their health data should be used to help improve the care of future patients who have the same condition, 89 percent agreed (Alston et al., 2012). Yet, when they were asked whether they thought their data were being shared, the majority said they either did not know or did not think so. "We have a huge gap between what patients are telling us they want and what they perceive us to be doing."

The data patients provide involve intimate parts of their daily lives. These patients are not simply human subjects, said Okun; they are actually members of the research team. "I would change our paradigm completely and start thinking of patients as patient researchers or citizen researchers." Okun quoted a recent blog post to the effect that patient engagement is the blockbuster drug of the century. If this is true, she added, and if this "drug" is not currently being used, the research community is essentially engaged in malpractice.

"The system is never going to be perfect," she said. But the biomedical research system has evolved to the point that all stakeholders can be involved in decisions. "Without patients, we would have no research. Let's start thinking about how we can best honor them, respect them, and allow them to develop the trust that they need to participate with us."

DISTRIBUTED SYSTEMS FOR CLINICAL RESEARCH INFORMATION SHARING

An alternative to widespread data sharing was described by Richard Platt, professor and chair in the Department of Population Medicine, Harvard Medical School, and executive director of Harvard Pilgrim Health Care Institute. Platt proposed that sharing information derived from the data while minimizing the sharing of data themselves nullifies some of the barriers discussed previously (Chapter 3). He went on to describe the Query Health Initiative, a system for sharing clinical information that has been promulgated by the Office of the National Coordinator for Health Information Technology. It uses the approach of sending the question to the data rather than bringing the data to the question. The question, in this case, is an executable program sent from the originator to the holder of data. The program then operates on a remote dataset and returns the answer to the sender.

An alternative approach based on the same idea, Platt indicated, is to let a user log onto a remote system and do the analyses. The user needs to be able to access the system through a firewall, which many organizations are hesitant to permit. Other protections can be built into the system as well, such as a mechanism for determining whether the research has oversight by an IRB. A steering committee or IRB could be involved in reviewing and approving queries. Network management could provide for auditing, authentication, authorization, scheduling, permissions, and other functions. Local controls at the source of the data could monitor what kind of question is being asked, who is asking the question, and whether the question is worth answering.

A logical extension of such a system would be a multisite system in which research data from several different organizations are behind several different firewalls (see Figure 4-2). According to Platt, a single question could be distributed to multiple sites and the responses compiled to produce an answer. Source data, such as information from electronic health records, could flow into research systems through firewalls. The result would be a system in which remote investigators can gain the information they need to answer a question while data are protected.

Platt described a system developed by his group that implements this concept. The system, called Mini-Sentinel, is being used by FDA to do postmarket medical product safety surveillance. It has a distributed database with data on more than 125 million people, 3 billion instances of drug

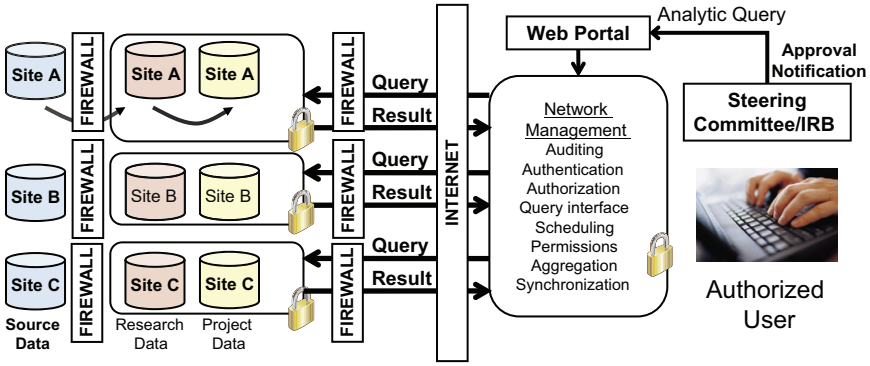

FIGURE 4-2 Distributed networks can facilitate working remotely with research datasets derived from routinely collected electronic health information, often eliminating the need to transfer sensitive data.
SOURCE: Platt, 2012. Presentation at IOM Workshop on Sharing Clinical Research Data.

dispensing, and 2.4 billion unique patient encounters, including 40 million acute inpatient stays. Each of the 17 data partners involved in the project uses a common data format so that remote programs can operate on the data. Data checks ensure that the data are correct. Data partners have the option of stopping and reviewing the queries that arrive before the code is executed. They also can stop and inspect every result before it is returned to the coordinating center. The amount of patient-level data that is transferred is minimized, with most of the analysis of patient-level data done behind the firewall of the organization that has the data. "Our goal is not to never share data. Our goal is to share as little data as possible." The analysis dataset is usually a small fraction of all the data that exist, and the data can usually be de-identified.

As an example of the kinds of projects that can be done using this system, Platt described a study looking at comparative risks of angioedema related to treatment with drugs targeting the renin-angiotensin-aldosterone system. The results of the study had not yet been released at the time of the workshop, but Platt concluded from the experience that data from millions of people could be accessed to do the study without sharing any patient-level data. Yet, from the perspective of the investigators, "essentially everything that was interesting in those datasets that could answer this question was accessible and was used to address the questions of interest."

Using such a system, it would be possible to address a large fraction of the questions thought to require data sharing by instead sharing programs among organizations that are prepared to collaborate on distributed analyses, Platt insisted. Organizations also could participate in multiple networks, further expanding the uses of the data they hold. At the same time, every network could control its own access and governance.

Today, only FDA can submit questions to Mini-Sentinel, but FDA believes it should be a national resource and is working on ways to make it accessible to others. Toward that end, the week before the workshop, the NIH announced the creation of the Health Care Systems Research Collaborative, which will develop a distributed research network with the capability of communicating with the Mini-Sentinel distributed dataset. Such systems, by sharing information rather than data, could make progress faster than waiting for all the issues surrounding data sharing to be resolved, said Platt.

5

Standardization to Enhance Data Sharing

Key Messages Identified by Individual Speakers

- Standardization can improve clinical research through increased data quality, better data integration and reusability, facilitation of data exchange with partners, increased use of software tools, improvements in team communication, and facilitation of regulatory reviews and audits.
- Collection of clinical research data using predetermined standards is preferable to post-hoc conversion of data to meet a standard.
- The development of standards requires collaborative expert input, analysis, and consensus.
- Clinical and scientific expertise is also needed to determine how to fit data retroactively to standards and harmonize terminology.
- Standards need to be used to the greatest extent possible, but they do not ensure data quality.

Speakers at the workshop addressed issues of standardization in two ways. They described the general principles that should underlie such efforts and they drew lessons from specific projects that could be applied more broadly.

Standards can be applied to clinical research data in multiple ways: standardized core datasets can provide a minimum set of variables that should be measured or recorded during a trial. Standards can also specify

how demographic (e.g., gender) and clinical information is recorded or defined.

The value of shared clinical data is undermined when those data cannot be used to answer new questions through secondary analysis. Standards can help facilitate pooling of data from disparate sources, either to increase sample sizes or for comparison purposes. By harmonizing vocabularies standards can also help to ensure that researchers are "speaking the same language."

HOW STANDARDS BENEFIT SHARING

Meredith Nahm, associate director for clinical research informatics at the Duke Translational Medicine Institute, emphasized that a major function of data sharing is reuse of data for purposes other than those intended by the people who collected the data. If the data are not defined well enough that others can use them, then the original researchers have not done their jobs well, she said. Data reuse requires both standards and a level of rigor and semantic specificity sufficient not just for human, but also for computational analysis. For example, she briefly described an effort by the Clinical Trials Network at the National Institute on Drug Abuse to de-identify data, align the data to Clinical Data Interchange Standards Consortium (CDISC) standards, and make the data available on the Web. Because data elements and tools were defined and implemented uniformly across the network, the mapping of the data onto the CDISC Study Data Tabulation Model (SDTM) standard was relatively straightforward, facilitating pooled analysis and cross-product comparisons.

As an example of the difficulties in synthesizing a common set of data element standards retrospectively from case reports, rather than having them defined upfront, she mentioned the different ways in which sponsors operationalized critical variables in clinical trials on treatment of schizophrenia. As a result, each trial examined yielded fewer and fewer instances of new semantic content. Authoritative clinical definitions are essential, she said, to reduce the burden on clinical investigational sites and to support the compilation and reuse of data for health care, research, and regulatory decision making. "It all depends on the data element as the atomic level of information exchange."

To demonstrate the need for standards to ensure that shared data can be pooled and compared, Rebecca Kush, CDISC president and chief executive officer, described the many different systems for reporting the

gender of study participants. Some use 1 and 0 for male and female, others 1 and 2, others M and F, and others an arbitrary designation. Health Level 7 (HL7) has about 15 options for the gender field, she said, depending on how people define themselves. With so many systems and no standards for data collection and reporting, data often have to be examined by hand just to determine something as simple as how many males or females are in a study. Using data standards, such as those being developed by CDISC and other standards development organizations (SDOs), can save significant time and cost, especially when implemented in the early stages of the study, said Kush. She reemphasized the value of developing standards a priori around a core dataset that is required across all trials. Information is lost when data are gathered in different ways and later mapped to common standards. Standardization also provides opportunities for additional impact on clinical research through increased data quality, better data integration and reusability, facilitation of data exchange and communication with partners, interoperability of software tools, and facilitation of regulatory reviews and audits.

Laura Lyman Rodriguez, director of the Office of Policy, Communications, and Education at the National Human Genome Research Institute, observed that the Institute has been thinking about issues of standardization as it has constructed large data repositories that combine genomic information with phenotype information. For example, the PhenX project, through a consensus process, has been working to create standard measures of phenotypes and environmental exposures for use in population-based genomic studies to facilitate cross-study comparisons and analysis. Standardized taxonomies to describe phenotypes ensure that different studies share a common vocabulary. Agreeing with several earlier speakers, Rodriguez emphasized that standards do not ensure quality and that the value of standardization is best realized when it is done upfront. However, aligning interests in the development of data standards and the sharing of data is not easy, she said. The search for common interests requires identifying common values and integrating them into the research enterprise. Communication and transparency can help identify and spread these common values while also building public trust.

Sharing and accessing clinical information is a global issue, said Neil de Crescenzo, senior vice president and general manager at Oracle Health Sciences. For a number of projects in which Oracle has been involved, there has been heavy emphasis on data standardization. Innovation and progress in clinical research and care will depend on immense

quantities of complex data being passed among organizations, and standardization can help overcome some of challenges posed by the "3 Vs of big data"—variety, volume, and the velocity at which data are needed. Outside the United States, de Crescenzo has seen great progress in implementing requirements for the use of standards in national-level research projects and electronic health record (EHR) systems. These efforts over the past decade have yielded many lessons to be learned.

Cautions on Standardization

While acknowledging the value of data standards, Vicki Seyfert-Margolis, senior advisor for science innovation and policy at the Food and Drug Administration's (FDA's) Office of the Chief Scientist, brought up some points researchers should remember when thinking about standardizing data. "Standardization does not ensure quality," she said. If not done well, conversion to a standard format has the potential to adversely affect data quality and analysis. For example, standardized formats for indicating patient race can still lead to inaccurate information if the categories used in questionnaires do not adequately capture the complexity of a person's racial identity. It can also result in loss of traceability from the source. Standardization does not imply that data are fit for purpose either, she warned. Standardized data may or may not answer the questions of interest and may or may not be useful for future analysis. It may not be possible to predefine all standards, and not all data must be standardized. FDA is working to identify minimum sets of data points that must be standardized for analysis. The effort devoted to standards needs to be weighed against these other considerations, she said, to determine how much time and money to invest in standardization, especially given that the data gathered will never be perfect.

Standards solve some problems, Seyfert-Margolis said, but they do not solve problems with data quality, disease definition, basic understanding, or data analysis. She emphasized the importance of defining diseases and having a clear understanding of clinical phenotypes as part of the standardization process. Especially as genomics begins to play a larger role in medicine, a taxonomy of disease will be needed to define patient subpopulations, "because we know not every type 2 diabetes patient is the same, yet we call them all that." In that respect, case report forms should not treat all patients identically because patient characteristics need to be probed carefully to clarify patient populations.

DEVELOPING STANDARDS TO ENABLE DATA SHARING

The Role of Standards Development Organizations

Kush described the desired criteria for data standards to facilitate clinical research. They should be fit for purpose; global; based on good clinical practices, guidelines, and regulations; harmonized and semantically consistent; developed through a recognized standards development process; consensus based; and platform independent. They should also encourage innovation and support links with health care. "There is no right or wrong in standards," said Kush, "[just] how are we going to agree to go forward."

CDISC, a nonprofit SDO, works to create standards that meet these criteria in order to support the acquisition, sharing, submission, and archiving of clinical research data. Such standards enable information system interoperability, thereby improving the efficiency and quality of medical research. The development of CDISC consensus standards requires the expert input from thousands of volunteers around the world, said Kush. In some cases, CDISC also works with other SDOs, like HL7 International, which generates standards for the exchange, integration, sharing, and retrieval of electronic health information to support clinical practice and health services areas.

For example, CDISC and HL7, along with partners at the FDA and National Cancer Institute, developed the Biomedical Research Integrated Domain Group (BRIDG) model to ensure that standards spanning the entire clinical research process are harmonized. "The idea is to have all these standards working together and go from end to end in the clinical research process," said Kush. The BRIDG model, which serves to bridge standards and research organizations, as well as the gap between clinical research and health care, provides a shared view of the semantics for the field of protocol-driven research and associated regulatory activities (e.g., postmarketing adverse event reporting). It was predicated on the need to pass information seamlessly between patient care and clinical research arenas in order to shorten the time lag between basic research and the implementation of new knowledge in patient care processes, said Charles Jaffe, chief executive officer of HL7.

An important current challenge in the arena of clinical research is how health care providers at different sites can use different electronic health records yet share a core high-quality set of clinical research data. One solution discussed by Kush that has been developed by CDISC and

its partners at Integrating the Healthcare Enterprise (IHE) is a tool that allows the EHR user to remotely collect a key set of data out of electronic health records, making it available for secondary uses like research and adverse event reporting. While complying with 21 CFR 11—regulations that require clinical researchers to implement controls such as audit trails and system validations to ensure that electronic health records are trustworthy and reliable—it can produce a standard core clinical research dataset, such as that defined by the CDISC Clinical Data Acquisition Standards Harmonization (CDASH) standard. This integration profile is being implemented in Europe and in Japan and has been used in some postmarketing studies in the United States. More than just a standard, Kush explained, it is a workflow tool. When implemented among ambulatory care physicians at Harvard, the time to report an adverse event dropped from 35 minutes to less than a minute, and the number of reports increased dramatically. "This is a real workflow improvement effort," said Kush.

FDA Efforts

New drug application data submitted to FDA have extremely variable and unpredictable formats and content, which presents a major obstacle to timely, consistent, and efficient review with currently mandated time frames. "This has been the problem for years," said Ron Fitzmartin, senior advisor in the Office of Planning and Informatics at FDA's Center for Drug Evaluation and Research (CDER). "It is unbelievable that in 2012 we are still saying that."

This lack of standardization has serious implications for FDA reviewers. It limits their ability to address in-depth questions and late-emerging issues in a timely manner. It also impedes timely safety analysis to inform risk evaluation and mitigation strategy decisions, and limits the ability to transition to more standardized and quantitative approaches to benefit–risk assessment. Given the "tremendous workload" facing reviewers at FDA, there is a great need for standards and tools that can expedite the review process, Fitzmartin said.

Toward this end, the Prescription Drug User Fee Act (PDUFA), which Fitzmartin noted was reauthorized in 2012 by the FDA Safety and Innovation Act (FDASIA), mandates that FDA develop "standardized clinical data terminology through open standards development organizations with the goal of completing clinical data terminology and detailed

implementation guides by FY 2017." It also calls for FDA to "periodically publish final guidance specifying the completed data standards, format, and terminologies that sponsors must use to submit data in applications." Fitzmartin presented a list of 58 therapeutic and disease areas where standards will be developed by the end of 2017, which will require extensive collaboration between FDA and other organizations.

Already, a number of organizations are converging on this challenge, said Fitzmartin. Recently, for example, CDISC and the Critical Path Institute, an independent, nonprofit organization committed to accelerating the pace of drug and diagnostics development through data, method, and measurement standards, have collaboratively formed an entity called the Coalition for Accelerating Standards and Therapies (CFAST). CFAST is an initiative established with the objective of defining, developing, and maintaining an initial set of data standards for priority therapeutic areas identified by FDA. HL7 is also working with CDISC and CFAST on clinical data standards, as is a recently formed industry group called TransCelerate BioPharma. FDA is participating in these efforts by providing scientific and technical direction to prioritize therapeutic areas, advising on work streams, and publishing draft and final guidance on completed standards. Because the standards are coming out under PDUFA, Fitzmartin said, they will be enforceable.

Fitzmartin concluded his presentation with three examples showing how standardization can help expedite the review process. In the first, Fitzmartin described a clinical data integration tool developed by SAS that could be used to map clinical trial submission data in the CDISC-developed standard SDTM format to the format required by a liver toxicity assessment product called Electronic Drug-Induced Serious Hepatoxicity, or eDISH. Using this tool, reviewers do not have to spend time piecing these data together and can quickly drill down to the patient-level data to look at outliers and elevated values.

In his second example, Fitzmartin described another adverse event diagnostic tool, this one developed by reviewers at FDA's CDER, which uses input data in the SDTM format to perform more than 200 automated, complex safety signal detection assessments. Within 1 month of being available to reviewers, medical officers using this tool discovered multiple cases of adverse events that previously had gone undetected, including anaphylaxis and pancreatitis.

Finally, Fitzmartin discussed how more than 50 common CDER review analyses, including demographics, exposure, adverse events, disposition, and liver toxicity, can now be automated through the use of

standard review analysis panels. A typical clinical review has not previously been able to produce this degree of output, which can corroborate sponsors' analyses and improve reviewer efficiency, consistency, and quality. However, standard panels require standardized data to run successfully, Fitzmartin said.

RETROSPECTIVE VERSUS PROSPECTIVE APPROACHES TO DATA STANDARDIZATION

Seyfert-Margolis described an experiment at FDA that was designed to evaluate the return on investment from normalization of raw data on an as-needed basis as compared to conversion of legacy data into a standard format. In this experiment, two different approaches were taken to the standardization of data. In the first, legacy data from about 100 new drug applications (NDAs) were converted to a standard data format with no predetermined scientific questions, then the standardized dataset was used for research. In the second approach, the converted data and the unconverted data were both used to answer a specific scientific question using a program called Amalga, Microsoft software designed to integrate patient data from disparate sources and in different formats.

The first approach yielded several key insights regarding the standardization of legacy data to allow for integration and comparison of studies and products, according to Seyfert-Margolis. First, scientific questions drove the details of the conversion. Clinical and scientific subject-matter expertise was essential to determine how to reorganize the data into the standard format required to address a particular question and to harmonize the terminology. Statisticians were needed to translate scientific questions into analyzable components. In addition, quality control of the converted data was essential, but time consuming, and the conversion activity in general was resource intensive and expensive, costing about $7 million in this case.

Seyfert-Margolis went on to describe lessons learned from the second approach to data standardization, where FDA used a tool that could take data from a variety of formats and transform it on the fly depending on the questions that were being asked of the data. Such tools allow integration of multiple types of data without having to go through all the time to standardize it first. Using this approach, they were able to integrate disparate regulatory datasets, including postmarket and premarket data, and in the process answer many interesting questions. Together,

standardization and advanced tools that integrate data on the fly could help move advanced analytics forward, she said, whether for the evaluation of multiple clinical trials or single product applications.

Collection of data using standards—versus conversion to a standard—is optimal, Seyfert-Margolis concluded. These standards should be implemented in the same way across studies, which would facilitate analysis across studies at FDA.

DATA-SHARING APPROACHES THAT HAVE BENEFITED FROM THE USE OF STANDARDS

The Alzheimer's Clinical Trials Database

Without standards, said Carolyn Compton, president and chief executive officer of the Critical Path Institute (C-Path), integrating datasets and pooling data is difficult. The Critical Path Institute acts as a trusted third party that works with partners in FDA, industry, and academia to develop consensus measurement, method, and data standards. The standards are then submitted to FDA for qualification. After qualification is achieved, the standards enter the public domain and can be used by everyone. C-Path "convenes consortiums to bring together the best science and in this fashion create shared risk and shared cost for the creation of these standards."

Compton went on to describe one of six such global consortiums organized by C-Path, the Coalition Against Major Diseases (CAMD), which focuses on diseases of the brain and peripheral nervous system. The coalition is seeking to advance drug development tools and standards as a means of addressing the challenge of the unsustainable time and cost required to get a new drug to market. In particular, the focus of its efforts is process improvement to advance effective treatments for Alzheimer's and Parkinson's diseases. CAMD is working to qualify biomarkers as drug development tools and has also been developing standards to create integrated databases drawn from clinical trials. These databases have been used to model clinical trials to optimize trial design.

Nine member companies agreed to share placebo control data from 22 clinical trials on Alzheimer's disease, but the data were not in a common format and needed to be combined in a consistent manner, Compton explained. All data were remapped to the CDISC standards and pooled. The resulting database was used to develop a new computerized clinical

trial simulation and modeling tool. To get there, however, the contributing companies had to go through a corporate approval process to share and de-identify the data, after which C-Path did further de-identification to ensure compliance with Health Insurance Portability and Accountability Act requirements.

The modeling tool allowed for accurate quantitative predictions of defined patient populations, Compton said. By merging data from diverse sources, 65-year-old males who looked alike in the databases could be divided into three classes with different trajectories of disease. "Seeing this kind of distinction emerge from the modeling tool would allow you to design a trial much more wisely," said Compton. "It would inform patient selection, study size, study duration, study feasibility, and even study costs."

Compton cited several key insights gained from the project. First, as others noted previously, legacy data conversion is resource dependent, but worthwhile for specific projects. In this case, de-identifying data and converting it to a standard format took 9 months, but generated a database with 6,100 Alzheimer's disease patients. To get the value back from the conversion process, it is important to assess upfront that the database will be useful for achieving specific objectives, like qualifying a new tool. If it will be, selectivity is beneficial, she recommended. "Convert the data you need, [but] maybe not everything." Once data are converted to a common standard and aggregated, the addition of standardized data from other sources, whether prospective or retrospective, becomes simplified and expands the power and utility of a standardized data resource. "Your database continues to grow over time and in power," Compton said.

Based on the success with Alzheimer's, the approach is now being applied to other research projects, including the development of new tools for Parkinson's disease, polycystic kidney disease, and tuberculosis. According to Compton, this approach could cut drug development times "by 4 to 5 years." Such tools also have applications to postapproval safety monitoring and data gathering.

Translational Medicine Mart

Eric Perakslis, chief information officer and chief scientist for informatics at FDA, provided an overview on an initiative he spearheaded while working at Johnson & Johnson in 2008-2009. The company asked

him to bring together data and informatics across their immunology, oncology, and biotechnology franchises, which originally had been different companies with many different clinical trials and standards. Rather than reinventing the wheel, he and his colleagues built their system off a data warehousing tool called i2b2 that had been developed by researchers at Harvard for data from electronic health records. They made it open source and ran it through Amazon's cloud computing service, which Perakslis termed "heretical" for the time.

The system, known as Translational Medicine Mart (tranSMART), was designed for research and development, specifically to generate hypotheses for biomarker research. The requirements for this kind of system are different from those required for automating premarket review of FDA submission data, said Perakslis. Johnson & Johnson wanted to be able to ask secondary research questions using the substantial amount of clinical trials data they had already collected. For example, many clinical trials have been done on potential asthma medications. What else can be learned from those trials?

At the time, the Innovative Medicines Initiative in Europe had gotten under way, and its first project was to look at severe asthma in 5,000 patients. Perakslis worked with the consortium to integrate the system he had helped build with the European effort. Within 3 months, the group had set up a pilot study and was able to combine data from several pharmaceutical companies and begin analyzing it. "Nobody could believe it had happened so early," said Perakslis, "but what happened more than anything else was the incentives aligned. We all had one goal."

Several lessons emerged from the experience, according to Perakslis. First, use the standards that are available because "patients are waiting." At some point, human curators are going to be necessary to align the data and insert it into a database, but to get the project moving forward, start with what already works. Second, an important goal for a project such as this one is to rule out options quickly. Clinical trials should not waste patients' time on drugs that are not going to work. "Get me 60 or 70 hypotheses that I can rule out, and then I can be really interested in the one that I cannot."

Perakslis concluded that he prefers light and agile data "marts," or databases generated to answer specific questions or test hypotheses, over large data warehouses. "That sounds like IT speak, but what I am saying is aggregate the source around the question quickly and effectively." That way, as technologies, standards, and definitions change, tools are flexible and can change accordingly.

ePlacebo

Michael Cantor, senior director of clinical informatics and innovation at Pfizer Inc., described an ongoing data-sharing project being undertaken by Pfizer as part of its "Data Without Borders" initiative. The project, called ePlacebo, pools data from placebo and control arms across multiple clinical trials in a variety of therapeutic areas. The result is a large comparison group that can be used to evaluate events that might not be seen in a single trial, study placebo effects, and possibly reduce the size of placebo arms needed in future clinical trials. So far, data from about 20,000 patients have been compiled from hundreds of trials, and Pfizer is hoping to expand the utility of this data source by soliciting participation from other organizations.

The goal for ePlacebo is to provide a resource that is inclusive, rests on standards, and spans disease areas. The intent is to set it up as a self-service dataset that could be used for any legitimate research purpose. However, consistent data standards have only been implemented at Pfizer within the past decade and as a result, only relatively recent studies were used for ePlacebo because of the difficulties combining data from trials that did not use standards or implemented them in different ways.

GOVERNANCE ISSUES

Compton discussed several important governance issues that arose during the CAMD initiative and in other C-Path efforts. First, rules for developing the data standards require collaborative expert input and consensus. Disease definitions need to come from the bottom up, said Compton, from the clinicians who are dealing with patients and diseases. A system cannot be imposed on them from the outside. However, the National Institutes of Health can use its purse strings to enforce clinician-driven, evidence-based guidelines, and perhaps some degree of evidence-based standardization could be regulated. Also, best practices for merging the data call for the use of high-quality data and FDA-accepted standards that work together along the process, from beginning to end. With regard to rules for accessing the data, the broadest possible data use agreements are needed, and access controls need to be appropriate to the use objectives. Finally, qualified drug development tools should be placed in the public domain to maximize their use.

With data as complex as those produced by clinical trials, standardization is needed upfront, added Cantor. But which standards should be used and how should they be implemented? Political will is needed to enforce standards—for example, by using the funding process to encourage standardization. Standards make it much easier to overcome the technical hurdles to broad-based cooperative projects, but people and institutions need the right incentives to contribute their data. As detailed in the next chapter, the social and cultural aspects of sharing clinical data are much more challenging than the technical issues.

6

Changing the Culture of Research

Key Messages Identified by Individual Speakers

- Given the substantial time and cost involved in conducting clinical trials, as well as the current system of career advancement in academia, academic researchers need incentives for releasing datasets, such as receiving credit for secondary analyses of their data.
- Clinical trial funders can influence the data-sharing actions of researchers by making grants contingent on compliance with data-sharing policies.
- Where journals can agree on principles and the means of enforcing those principles, they, too, can shape data-sharing policies.
- New policies at the European Medicines Agency on the release of clinical trials data could have implications for data sharing worldwide.
- Engaging patients in research and being open and honest with them can lead to patient-driven mechanisms for data sharing.
- As organizations increasingly offer data analysis services and medical advice over the Internet, the traditional health care and biomedical research enterprise may need to adapt to keep up with the changing culture.

The norms and expectations of the various groups involved in data sharing came up repeatedly during the workshop. In the context of the previous chapters, for example, rigorous standards and working mod-

els for data sharing can go only so far if not supported by the prevailing culture.

Many participants acknowledged the difficulty of changing well-established cultures. Educational preparation, vigorous enforcement, and consistent leadership are steps in the process of cultural change. Cultures also can be transformed by profound technological change, as is happening today with the application of Internet-based technologies and practices to health care.

THE ROLE OF REGULATORS

Hans-Georg Eichler, senior medical officer at the European Medicines Agency (EMA), which regulates drugs and biologicals in Europe, discussed recent major policy changes at EMA regarding sharing of data from clinical trials. EMA is a public agency, said Eichler, and as a public body it is obliged to be fully transparent. The only exceptions, he said, are personal protected data and commercial confidential information. Given the overriding public health interest, EMA recently has taken the position that clinical trial data will no longer be considered commercial confidential information (Eichler et al., 2012). This has "huge implications," according to Eichler.

Currently, EMA is providing trial reports retroactively with personal information redacted. In the future, however, it will publish trial reports proactively. The next step is making all data held by EMA publicly available, including data from prelicensing clinical trials, pharmacovigilance, and observational data. If someone asks for de-identified patient-level data, "we will make it available," said Eichler. EMA is approaching this objective "gingerly," he continued. The release of data puts many people, particularly from industry, outside their comfort zones. However, science is moving toward a new model of openness in which data are made available for others to reanalyze and combine with other data.

The open question is whether making clinical trials data available will be a boon or a bane for drug development and public health. One barrier to making data available is that clinical trials data include personal information that needs to be protected. However, Eichler said, "that's probably an addressable problem." What will likely happen is that EMA will tell industry that as of a certain date, all clinical trials data submitted to the agency will be available to anyone else, so it should not

include protected patient health information that could be used to identify individuals.

Another risk of data sharing is that reanalysis of data may produce phantom risk and health scares. Neither regulators nor industry like to be blindsided by reports that a drug or vaccine has an unreported side effect, but Eichler predicted that many licensed drugs could come under attack based on such reanalyses. As an example, he cited a meta-analysis of a drug called tiotropium bromide that found a slightly increased risk (relative risk of 1.6) of adverse cardiovascular events in chronic obstructive pulmonary disease patients using the drug (Singh et al., 2008). In this case, however, the company responsible for the drug had a study under way looking at long-term clinical endpoints that ultimately found no increased risk (Michele et al., 2010). This is a risk of meta-analyses and the use of observational data. How many beneficial drugs will be lost through mistaken analyses, and how will people be persuaded to participate in postmarketing trials of drugs if they perceive a possibility of drug-related harm, Eichler asked.

Despite the risks of data sharing, there are clearly considerable benefits to patients and the research community. Open science could support the development of predictive models for patient selection to appropriate treatments or doses based on patient characteristics. A second advantage is that different therapies could be compared to determine relative efficacies without the expense of direct comparison trials. "This will be a boon for comparative effectiveness research," said Eichler.

EMA is still in the process of determining how best to make data available, and it is engaging many stakeholder groups in this discussion. However, Eichler said that data will not be released until the agency has made a regulatory decision on the product based on its assessment of the data. Also, EMA intends to ask for the preregistration of protocols for data reanalysis to avoid data dredging that is unlikely to produce meaningful results. EMA wants to know in advance whether studies on requested data will be exploratory or confirmatory in nature.

Shortly after the workshop summarized in this volume, EMA conducted its own workshop to bring together stakeholders to provide input to the development of its policies. Issues addressed included standards for storing and sharing of data, the level of data to be released, standards for protection of personal data, quality standards for meta-analyses, and rules of engagement among stakeholders (EMA, 2012).

The implication of the EMA policy change for the Food and Drug Administration's (FDA's) policy of nondisclosure was raised during the

discussion period. Unlike EMA, said a participant, FDA is prohibited from releasing patient-level data by statute and regulation, presenting a major legal barrier to data sharing. Robert Califf, Duke University Medical Center, contended that the reports companies send to FDA should be made public, along with the internal FDA analysis. Today, if a drug does not get to the market, federal law prohibits the release of these documents, but companies still could make these reports public, he said, even if FDA currently cannot.

THE ROLE OF JOURNALS

Steven Goodman, who, in addition to his academic appointment at Stanford University School of Medicine is also associate editor at *Annals of Internal Medicine* and editor at *Clinical Trials*, discussed the role of journals in promoting data sharing and the challenges they face. In a paper published in *Annals of Internal Medicine* in 2007, Goodman and several colleagues announced a new policy the journal was adopting to require that manuscripts include a reproducible research statement (Laine et al., 2007a). Such a statement would say whether the study protocol, code, and dataset are available and how to get each. Goodman labeled this a "weak" solution, but he also said that if *Annals of Internal Medicine* makes demands that are difficult to fulfill, authors will simply publish their articles elsewhere. "Journals are competitive with each other. They also want to publish the best stuff. And they can't put up barriers that nobody else is putting up," Goodman said. The requirement has at least shined a light on the problem, but some authors have simply said that data are not available or have referred readers to the large databanks from which the data in the study were derived. Polling by journal staff has indicated that the number of requests authors are receiving for data, statistical code, and protocols is still fairly low.

Journals cannot be effective acting alone, said Goodman. To really shift the culture surrounding data sharing, journals will need to agree on a common set of principles and sanctions, such as requiring that the authors of articles share data on request. Although a few other journals have adopted the reproducible research statement policy, in general, journals are taking their own approaches to dealing with data sharing and the issue of reproducibility, and some have no such policies at all. One success story mentioned by Goodman was clinical trial registration. Though the system still needs to be improved, he said, it has worked well

because it is done through one central repository, and having a legislative mandate and collective support by the medical journals has helped with enforcement (Laine et al., 2007b). Journals cannot be the custodians of all research data and protocols, and they cannot be the sole guarantors of scientific quality because they have neither the staff nor often the technical capability.

THE ROLE OF FUNDERS

Goodman also briefly touched on the role that funders have in promoting data sharing. For example, the National Institutes of Health (NIH) requires a data-sharing plan for research projects funded at levels above half a million dollars, and the National Science Foundation recently started requiring all grantees to have plans for sharing data in a timely fashion and at nominal cost. Goodman pointed out that Howard Hughes Medical Institute (HHMI) has a very detailed requirement that funded researchers make any materials, data, databases, and software deemed integral to the publication freely and expeditiously available for use by other scientists, with no restrictions on use. Interestingly, HHMI actually specifies that researchers may not insist on collaboration, coauthorship, or prior review of manuscripts generated using their shared data and materials. Other funders have their own policies, but the extent to which these policies are being followed is difficult to determine, said Goodman. However, a recent joint statement by a group of funding organizations that was published in the *Lancet* (Walport and Brest, 2011) indicates that funders are aware of the role they can play in changing the culture of data sharing. The statement indicated an intention to work together to increase access of the scientific community to research data that is funded by their organizations.

As an alternative to making funding contingent on adherence to specified data sharing policies, a suggestion was raised during a discussion period that funders consider track records in data sharing as a significant factor in the scoring of funding proposals.

NIH Perspectives

The issues associated with data sharing are a major concern of the NIH leadership, said Josephine Briggs, director of the National Center

for Complementary and Alternative Medicine and acting director of the Division of Clinical Innovation in the National Center for Advancing Translational Sciences. Almost a decade ago, the NIH implemented a data-sharing policy for cooperative agreements (through which many NIH-run clinical trials are funded) and for grants exceeding a half-million dollars, but that is just a "baby step" compared with the many things that need to be done to promote data sharing, said Briggs. She briefly described three ways in which the NIH is investing in data sharing. First, it is investing in data standards, which, as described in Chapter 5, can facilitate pooling of shared data from different sources and comparison of results from independent studies. For example, the National Institute of Neurological Disorders and Stroke has developed sets of common data elements (CDEs) for specific disease areas and now requires researchers who receive funding from the Institute to ensure that their data collection is compatible with those standards. One concern she raised, however, was that data elements defined for different disease areas will use different demographic variables. The trans-NIH BioMedical Informatics Coordinating Committee, which is being led by the National Library of Medicine, is collating a list of available CDEs, which may draw attention to needs for harmonization. Second, the NIH is supporting data-sharing resources in order to make datasets easier to find, accessible, and available. These resources are generally disease specific and are led by a single Institute or Center. Briggs pointed to the National Heart, Lung, and Blood Institute (NHLBI) as a great model, given its clear and unambiguous expectations for data sharing in large trials and to a certain extent even in smaller trials. Data can be shared through the data repository managed by the NHLBI's Biological Specimen and Data Repository Information Coordinating Center (BioLINCC) or directly among investigators as part of their continued collaboration on NHLBI-funded work.

The final way that the NIH promotes data sharing is by funding secondary analyses on existing datasets. Clinical trials can be extremely complex to organize and run, often requiring large collaborations, but secondary analyses of trials are "an incredibly important way for individual investigators to participate in the generation of new knowledge," said Briggs.

For studies with budgets of less than $500,000, NIH policies are not clear regarding expectations for data sharing, Briggs acknowledged. But the NIH controls the purse strings, and by creating expectations for smaller grants that datasets should be shared, it could exert a powerful influence.

INCENTIVIZING CHANGE BY ENSURING CREDIT

Ensuring that researchers who generate data get credit for it was raised by several workshop participants as an important incentive to promote data sharing, particularly in the academic community, where career advancement depends on publications and citations. Throughout the workshop, different mechanisms for giving trial organizers credit were discussed, including offering coauthorship, listing trialists as collaborators, and assigning datasets unique identifiers that researchers can track to show downstream use of their data.

Code of Conduct for Conducting Secondary Analyses

While discounting many other arguments commonly raised against data sharing, Andrew Vickers, Memorial Sloan-Kettering Cancer Center, acknowledged the validity of one of the major cultural arguments against data sharing—that researchers have a right to exploit data that they have spent years collecting. Researchers do need incentives to collect data. But blocking access to data forever is far from the only available alternative. The investigators who have collected the data will have the opportunity to publish the first paper on those results and an embargo period during which they alone can use the data would be simple to arrange, said Vickers. Systems conferring credit for the reuse of data are being discussed and are needed to incentivize data sharing in academia. We know already that papers for which the data are made available are cited more than papers for which the data are not available (Piwowar et al., 2007).

Vickers (2006) suggested that a code of conduct governing the use of shared raw data could help to ensure that the original data collectors get fair credit for their work. He suggested that a code of conduct could include the following: an independent investigator planning to publish a new analysis of previously published data should contact the trialists, those who ran and published on the original clinical trial, before undertaking those analyses; if a reanalysis of the data is to be published, the trialists should be offered coauthorship or an opportunity to write a commentary to be published alongside the new analysis; journals should refuse to publish the new analysis unless this step has been taken; and finally, the original publication should be cited in any new analysis of the data.

Researchers and companies that continue to resist the release of data are swimming against the tide of history, Vickers said. When open access to scientific papers was first proposed, it was widely resisted, as was clinical trial registration, yet today both are widely accepted. "A whole bunch of things seemed very radical at the time. I think data sharing is one of those," Vickers observed.

Trial Organizers as Collaborators on Secondary Analyses

Myles Axton, editor of *Nature Genetics*, has been involved in several experiments to allow greater access to research data, including databases of genotypes and phenotypes, micro-attribution as a way to incentivize community annotation of the human genome, and peer review on an open data platform. However, at the workshop, he focused on a different means for ensuring that investigators get credit for data they generate. He argued against the separation of people who have invested their time in a clinical trial from the data generated by the trial. The trial organizers should, of course, be able to continue to use their data. But in a second track, the trial organizers should be cited as collaborators and not authors. This would allow the original trial organizers to distance themselves from the conclusions of others who reuse their data while remaining associated with those data. Data need to be analyzed independently, but the people who spent years organizing the trial also should receive credit for the generation of those data—even if subsequent conclusions end up being critical of the trial, Axton said. An additional step forward would be to universally identify exactly what each person did in the production of new knowledge. "There should never be a discussion again about authorship order," he asserted.

Unique Dataset Identifiers

Steven Goodman, Stanford University School of Medicine, proposed yet another mechanism by which due credit could be ensured. Currently, academic researchers have only two ways to gain credit for their work. They are an author on a paper, or their paper is cited. What is needed, said Goodman, is a way to measure use of someone's data for the generation of novel findings and publications. This would require that each dataset has a unique identifier, like the PubMed ID for a paper. "Every

single time that dataset is used, [that identifier] needs to be in the paper that used it." These citations could then go on the CVs of academic researchers and factor into hiring and promotion decisions. Some organizations are already doing this. For example, iDASH (Integrated Data for Analysis, Anonymization, and Sharing), a data repository established at the University of California, San Diego, specifically for health research, assigns unique identifiers to all datasets it provides via its Web-based distribution system. According to Goodman, applying the approach more broadly is key to solving the incentive problem. "We have to create a culture and a reality where people benefit as much from everyone sharing their data for all purposes as they currently do from protecting it."

PROTECTING AGAINST MISUSE OF SHARED DATA

One of the major barriers to data sharing identified by those in industry is fear over the misuse of data. Several workshop participants raised the possibility of controlled access as a means of protecting against the potential harms from poor-quality secondary analyses of shared data. Goodman described different models of data sharing that are intermediate between full access, where the data can be used for any purpose with no restrictions, and no access. For example, he said, data can be shared only for the purpose of reproducing the results that were published or for commenting on the results via a letter to the editor, with no original findings based on the data published without explicit permission from the original investigators. Alternatively, the data can be used to generate new findings, but any modifications to the data also need to be made available and/or the authors of the original data need to be cited. "There are ways to mediate this relationship that are not 'I give you the data' or 'I don't give you the data,'" Goodman said.

Similarly, Axton proposed that one way to obtain access to research results could be to have anyone wanting to reuse the results document his or her status as a bona fide researcher and provide a research plan detailing the objectives of the research to be performed. Such a request could specify the dataset that is necessary and sufficient to conduct the proposed research. It also could provide a detailed documentation of processes designed to ensure that the data will not be distributed to third parties and will be protected to safeguard the privacy of the research subjects. Under these conditions, the default should be that access is granted rapidly by the trial organizers and owners of the data. If this default is not achieved,

a data access moderation committee could be a source of recourse. According to Axton, such a committee should include members of the trial group, independent researchers, and participant representatives. It would be responsible for advising those who have been denied access on how to comply with conditions for access. In this way, it could protect research subjects while making data more available for useful research questions. It would be quicker than existing procedures and should work better because the trial group would remain involved in data reuse analyses and in publications.

PATIENT-DRIVEN SHARING OF CLINICAL RESEARCH DATA

Institutions that participate in the clinical research enterprise must comply with regulations such as the Health Insurance Portability and Accountability Act Privacy Rule and the Common Rule, which place clear boundaries on use of patient data. But when patients take data they have generated themselves to the Internet, these regulations do not apply, said Deven McGraw, director of the Health Privacy Project at the Center for Democracy and Technology, making such "patient-facing pathways" enormously attractive. People dealing with a serious illness often have different conceptions of privacy than someone who is not and, therefore, may be more willing to share health information. "We need to acknowledge that there is a great range in the extent to which people care about their privacy and give more flexibility in that realm, she said. As a privacy advocate, McGraw stressed the importance of protecting patients' personal health information and postulated that both institutional and patient-facing pathways to data sharing rely too much on the consent process for this purpose. Most patients will sign almost anything put in front of them if they trust the person asking them to sign, but consent does not necessarily protect privacy, McGraw observed. Consent forms therefore create an obstacle for researchers without providing patients with much protection.

McGraw contended that the general type of consent form often used in online research is not specific enough with regard to how the patient's information will be used. When someone gives consent to do research with their data using such forms, others define what is and is not research, not the person giving consent. The same observation applies to other uses of the data, including commercial uses. "We need another

framework for thinking about how we make sure in this environment, both on the regulated side and on the unregulated side, that there is public trust and understanding of what we're doing."

McGraw suggested a different approach based on what are known as Fair Information Practices. These are models of data stewardship that build both privacy protections and public trust into the process. She presented a set of such practices drawn from the Markle Common Framework, which was issued by the Markle Foundation in 2006 as a framework for the exchange of information among health professionals:

- openness and transparency about how data will be used;
- purpose specification and minimization;
- collection limitation to only those data actually needed;
- use limitation;
- individual participation and control (e.g., patient consent);
- data integrity and quality;
- security safeguards and controls;
- accountability and oversight; and
- remedies.

McGraw expounded on some of these principles as follows: the users of data need to be open and transparent about the purposes for which they are using the data; investigators should take only the data they need to address a research question and not take data that are not needed; if data are to be used for purposes significantly outside the context for which they were collected (e.g., sale to third parties), permission needs to be obtained. The purpose of this kind of framework is to create a system "that works without necessarily relying on the patient to evaluate and say yes to each and every research question that we want to bring to the data," said McGraw.

The concept of data ownership is not very helpful in considering the sharing of health data, McGraw observed. A better and more workable concept is that holders of data have rights and responsibilities that accompany them. "The patient has a right to transparency about data, to be able to get copies of data, to take data and to use it in ways that they want to, including to donate it for research projects if they want to do that," she said. Research organizations that have data in their possession have a responsibility to think about sharing that data in ways that protect the rights of patients. "If we're struggling with notions of who owns [data]

and when can it be given away, we're starting in the wrong place. The holders of the data have responsibilities."

Beginning in 2014, clinicians participating in the Meaningful Use of Electronic Health Records incentive program will be required to provide patients with the capability to view, download, and transmit clinical data that are part of an electronic medical record. This will create a "very interesting dynamic," said McGraw, as patients gain more control over their health data.

Williams Syndrome as an Example

Beth Kozel, instructor of pediatrics in the Division of Genetics and Genomic Medicine at St. Louis Children's Hospital and the Washington University School of Medicine, works with individuals who have Williams syndrome, a rare genetic condition affecting approximately 1 in 10,000 individuals. Kozel described health effects associated with Williams syndrome, including significant cardiovascular anomalies, hypertension, neurocognitive effects, predisposition for obesity and diabetes, and endocrine abnormalities. However, each characteristic varies in severity among people with Williams syndrome, which is likely caused by differences in genetic background and environment exposures. This constellation of features leads to a complicated health picture for these individuals, but it also leads to the confluence of research groups interested in these many different phenotypes.

As a clinical geneticist, said Kozel, she would like to have genomic or environmental information that she could present to families to let them know what might happen to a child, rather than giving families a long list of things that might go wrong. The problem is that the sample sizes needed to study the effects of genetic backgrounds or environmental exposures are large; several hundred patients may be needed in a study to detect an association. To do such studies, people who work with Williams syndrome need to pool their data because most investigators work with relatively small numbers of people. But several major barriers have limited such sharing to date. Kozel works with the Williams Syndrome Association, which is an organization that brings together people with the syndrome and their families. It provides information for families, teachers, and others who work with people with Williams syndrome. It also includes a registry that allows families interested in research to interact directly with researchers. As part of its efforts to promote collab-

orations among researchers, it sent a survey to 30 individuals and groups known to be active in research on the syndrome. Only 15 surveys were returned, and of those 15 respondents, 9 said they had no samples. Six said they had samples, but went on to cite various challenges to sharing. "There were absolutely zero investigators who said, 'Yes, I have samples and I would love to share them with you,'" Kozel said.

One important barrier identified by Kozel involves the issues that arise in other genetic studies. Genetic signatures may be identifiable in public databases, particularly with a small community where people know each other. Contributors of data may expect to receive results back. Some of the people in studies were consented before the molecular diagnosis was even known, and reconsenting them for new studies would be a challenge. Other samples were collected when someone was a child and is now an adult. Regulations or restrictions imposed by institutional review boards (IRBs) may place limits on doing research on genetic material collected in the past.

Other barriers involve the culture of academia. Investigators may be worried about getting credit for contributing samples. Being included in the middle of a long list of authors is not going to help a junior investigator receive tenure. Scientific "clout" may be associated with an investigator's access to rare samples. Some investigators "have accumulated hundreds of samples and have reputations with the families—and that is who they are," said Kozel. "If they let that go, their clout in the community becomes different."

Kozel suggested that patients and patient groups have a role to play in overcoming some of these barriers. IRBs could allow patients and families to become active partners in making decisions about issues such as genetic confidentiality. For example, the registry of the Williams Syndrome Association has an online forum where families can discuss changes in protocols and then make decisions about whether to continue with research. Social media and new technologies also could increase the engagement of patients and families, which could lead to better acquisition of data. As an example, longitudinal data could be acquired on changes in phenotypes over time. Family groups can educate their members about the pros and cons of data sharing. They could ask members to look for and ask about sharing statements in consent forms, and when data sharing is not allowed, ask why. If "researchers are aware that the individuals giving their time to the study want the data shared, [it] may put more impetus on the researchers to make it happen."

Some barriers are beyond the reach of patients and family groups, and other stakeholders will need to step up, said Kozel. For example, the expense of well-run biobanks is too large for small family groups to support. Funding organizations could consider establishing central biobanks for rare diseases. When samples are limited, the provision of downstream data, such as sequence or expression data, may be preferable to storing and distributing samples. Journals can continue to require such genomic data to be deposited in protected but accessible sites online. They also could consider mechanisms to connect authors of underpowered research instead of allowing publication of lower-powered studies that can later be reexamined by meta-analysis. "It doesn't serve our rare disease community or science itself for all of this data to be sitting in people's drawers," Kozel concluded. But the acquisition of large numbers of rare samples will require coordinated efforts among multiple groups, and changes in practice will likely be needed from all stakeholders.

Public-Driven Sharing of Clinical Research Data

Clinical research data, said John Wilbanks, director of Sage Bionetworks, is more than the information historically contained in folders at a physician's office. Those folders, which have now been reproduced in electronic medical records, contain only the information generated during episodic trips to the doctor. New technologies, biomedical as well as ubiquitous sensors such as cell phones and computers, now enable people to collect longitudinal data on their health and other aspects of their lives, regardless of whether they are in a traditional clinical research study.

A week before the workshop, Wilbanks got his genotype from the company 23andme and posted it on openSNP, which is a wiki based in Europe created by a postdoctoral fellow to enable genomics research. Within 2 days he got an e-mail from another wiki called SNPedia with an annotation of his genotype, which indicated that he had a genetic variant conferring an increased risk of hypertension, along with another variant that seems to prevent baldness. This is happening "outside of any sort of regulated direct-to-consumer system," said Wilbanks. Although he would prefer that he got this kind of information from health care providers who have the training and resources to substantiate the information they provide, "I'm not getting this service from the health system as an individual and my capacity as an individual to generate data about myself is exploding,"

he said. Services in the marketplace now enable an individual to obtain their genotype and distribute it to people who will interpret it and return the results via e-mail. "People who are frustrated are increasingly going to find these services and start using them" despite a lack of standards, site protections, and privacy.

Wilbanks also uploaded the genome file he received from 23andme into the Sage Bionetworks Synapse system, which is a self-contributed data repository for genomics research. The system, which includes an online informed-consent process, allows data scientists to conduct collaborative research on individual-level data that are provided in a standard format and have been cleared with respect to privacy protections.

With the computational and consent infrastructures in place, the last piece in the democratization of clinical research is something that begins to change the role of the individual, "so it's not just 'I'm a patient and I see my doctor *x* number of times a year.' You can be a participant," said Wilbanks. Bridge, which is the newest piece of the Sage system, demonstrates the power of this kind of model. It provides a means for people who have data about themselves to come together and commission researchers to build the computational disease models. For example, he said, "50 people with early-onset Parkinson's could come in and say, 'we've got genomics data, we've got all sorts of other omics data, we've got metabolic and molecular data, it's in a standard format—$50,000 prize to the first person who builds a successful computational model.'"

Wilbanks proposed a simple set of standards to guide this kind of public-driven data sharing. First, he said, be honest with people. If people send their genomes to a shared system where data are at least moderately public, their privacy is unlikely to be permanently protected. Contributors of data need to know about the risks they face, but society should also have some tolerance for people who think the value of sharing their data is greater than the risks, such as those with a rare disease. Second, data should be reusable, which to Wilbanks meant computationally useful. Scans of paper records that patients have typically received from their doctors when requesting their medical records, for example, are not reusable. Finally, data should be portable so they can be shared among institutions, doctors, laboratories, and studies. When the control group from one study can also serve as a cohort control for another, "it begins to accelerate the system exponentially," said Wilbanks.

7

Final Reflections on Sharing Clinical Research Data

Key Messages Identified by Individual Speakers

- Data sharing does not have to be all or nothing; sharing of clinical study reports, including protocols, may be an intermediate step representing "low-hanging fruit" that would be relatively easy to implement now. However, even this approach brings its own challenges.
- Although notable progress has been made in sharing of placebo and comparator arm data, companies need to think about the boundaries of precompetitive space and what is gained and risked by sharing active arm data.
- Given their access to valuable data and their urgency to advance treatment alternatives, patients need to be engaged as partners in the clinical research process. When this is done successfully, much can be learned about the natural history of diseases and how best to match patients with promising treatment options.
- Clinical trials participants deserve to receive information from trials that will help them make health decisions, though decisions about what information to return to patients are the province of institutional review boards.

Throughout the workshop and during the closing discussion period, low-hanging fruit and priority actions were identified by speakers and other participants. Those key take-away points from the workshop are

gathered in this final chapter of the summary as a way to highlight and elaborate on next steps for advancing the sharing of clinical trials data.

LOW-HANGING FRUIT

As a possible intermediate step in the release of participant-level data, several participants raised the possibility of sharing clinical study reports, which are much more detailed reports than typically appear in a publication or even in an online publication. These reports, which generally contain the methodology, the subgroup analyses, the sensitivity analyses, and other detailed analyses of the data, are "low-hanging fruit" that could add to the information available in publications, without engendering the kinds of concerns raised when sharing patient-level data, said Jesse Berlin, Janssen Research & Development. Peter Doshi, Johns Hopkins School of Medicine, agreed that clinical study reports may be good initial targets for data sharing because they already exist for nearly all trials. Making them available would not be expensive, but would promote research integrity, medical knowledge, and public health (Rodwin and Abramson, 2012).

In fact, some trial organizers have already instituted this practice. Sachin Jain, chief medical information and innovation officer at Merck & Co., Inc., said that starting in July 2011, Merck has included the protocol and statistical analysis plan as part of its submission package to journals. Upon a journal's acceptance of a manuscript for publication, Merck also provides the journal with the opportunity to post on its website the key sections of the protocol, including the objectives and hypotheses, patient inclusion and exclusion criteria, study design and procedures, efficacy and safety measures, the statistical analysis plan, and any amendments relating to those sections. "The response was terrific," said Jain. "This is an idea whose time has come." If external investigators need additional data at the patient level, Merck also has an initiative through which they can approach the company and ask for access. The data request goes to the product teams, who have the opportunity to review the request and ask questions.

Study protocols, which show how the investigators intended to run the trial, collect data, and perform analyses, provide far more detail than can be condensed into the methods section of a journal article, and often include important information from protocol amendments, such as changes to outcome measures. To emphasize the potential impact of

sharing study protocols on evaluation of trial results, Doshi cited a trial on the use of celecoxib (Celebrex) in the treatment of osteoarthritis and rheumatoid arthritis (Silverstein et al., 2000). Availability of the original study protocol allowed investigators to determine that the trial had been misreported, including the key claim that Celebrex was safer than alternative treatments (Hrachovec and Mora, 2001).

However, although the partial release of data may seem preferable to not having any data released, challenges with this model need to be considered. For example, partial release of data where trial organizers are allowed to decide which information to publicize creates the potential for selection bias.

THE BOUNDARIES OF PRECOMPETITIVE COLLABORATION

Many of the data-sharing initiatives discussed during the workshop, such as the Datasphere Project, the Coalition Against Major Diseases and ePlacebo, involve the pooling of data from placebo or comparator arms of trials. The value of such initiatives was clearly acknowledged; for example, Janet Woodcock, director of the Food and Drug Administration's (FDA's) Center for Drug Evaluation and Research (CDER), indicated that by sharing the placebo group, one can decrease the numbers of people one needs to enroll in a clinical trial. However, added value of and barriers to sharing active treatment arm data was a notable topic of discussion. The successes that emerged for cardiovascular disease (see Box 2-2) were raised as motivation for making the sharing of these kinds of data a high-priority action.

Carolyn Compton, Critical Path Institute, talked about this issue in the context of defining the boundaries for precompetitive collaboration. Companies need to think about what they can collectively gain by pooling not only data, but expertise and other resources, "so that they all get something that none of them could acquire on their own," she said. People will need to get past their anxieties over intellectual property (IP) in order to get more benefit out of money they have already spent and data they have already collected, both from successful and failed trials. Michael Cantor, Pfizer Inc., added that it is not just about the IP risk. There is also a lot of concern about the potential of uncovering previously undetected safety signals when one starts pooling data. Questions remain about what role the FDA could play in helping to sort out the

validity of safety concerns resulting from pooled analyses but, said Compton, there is the potential for enormous benefit to both the public and private sectors if people would be willing to contribute these kinds of data to a common source that was accessible to everyone.

PARTNERING WITH PATIENTS

In her closing statement, Josephine Briggs, National Institutes of Health, emphasized the need for clinical research organizations across all sectors to build systems that incorporate "patient-facing" pathways to data sharing, leveraging the urgency patients feel around advancing treatments, their access to other sources of valuable information, and the fact that they often have different tolerances for privacy risks as compared to the well public. During the closing discussion period and throughout the workshop, many participants talked about forming stronger partnerships with patients as a priority action for advancing clinical research. There was also discussion on how best to give back to patients who donate their time and personal information, and willingly take on the risks associated with novel treatments and protocols.

Learning from Patients

Jay "Marty" Tenenbaum, founder and chair of Cancer Commons, echoed Briggs' sentiments. Health care is witnessing the dawn of personalized genomic medicine, but not enough information exists to make informed clinical choices on the basis of genomic data. For example, asserted Tenenbaum, cancer is not just a handful of diseases, but hundreds or thousands of diseases depending on the particular molecular drivers of a person's tumor. Clinical trials using small cohorts of volunteers are not enough; 95 percent of patients will not be treated in clinical trials. Tenenbaum suggested implementing the Institute of Medicine's vision of a rapid learning community in cancer (NRC, 2010) as the driving application behind personalized genomic medicine. Genomic and clinical information should be captured "from every cancer patient, from every conceivable source, from clinical trials, from electronic health records, and from the patients themselves, because they are the ones who have the natural instincts and incentives to share." Thousands of experiments occur every day as oncologists try to extend the lives of their patients and

researchers need to be capturing these longitudinal data. This information could be used to create molecular models of cancer subtypes. As additional data are developed following treatments for those subtypes, models can be updated and subtypes can be split or modified, if necessary.

To realize this vision, patients will need to donate their data, which will increasingly include genomic information as such data are placed in patient medical records, Tenenbaum added. Information about potential improvements in treatments will create incentives to participate for both patients and physicians, as will transparency on what happens with the data and assurances that patients participating in research will be the first to benefit from resulting treatment advances. Patients' data could be available online, which will "unleash an ecosystem of third-party applications," according to Tenenbaum, offering value "that we can't even anticipate but that will have a dramatic impact on health."

Janet Woodcock, CDER, agreed with Tenenbaum that the Internet and social media could be used to learn much more about patient populations and the natural history of diseases, particularly those that are rare and therefore difficult to study. "It is a tragedy that [information about] most people's course of illness is not used to further treatment of that illness," she said. However, in contrast to Tenenbaum, she advocated for conducting research in a systematic manner, as can be done with clinical trials. She mentioned the I-SPY trial, which is being used to study biomarkers for predicting response to cancer therapies. Moving that kind of initiative out into the community, for example, by inviting people who have cancer to get their tumors genotyped, could enable the design of large-scale trials in which people could be matched to the drugs that are most likely to have effects on that tumor, and then the trial could be adapted over time. "Everyone agrees that sharing clinical data is a public good . . . and that we really need to figure out how to leverage these data as much as possible because [they're] so precious," said Woodcock.

Giving Back

In the context of incentivizing patients to donate their clinical data, Cindy Geoghegan, principal at Patient and Partners, asked about the extent to which participant-level data should be made available to the participants in a clinical trial, whether in a de-identified form or more personally, so that, for example, incidental findings are conveyed to participants.

Harlan Krumholz, Yale University School of Medicine, had previously reminded the workshop participants that to restore trust in clinical trials, investigators need to listen to people so as not to seem "self-serving even as we are trying to promote society's best interest." With that in mind, principles could be adopted that would govern the return of information to research participants. Many trialists do not make enough effort to let participants know the results of the trial and what the implications are for society, let alone for them as individual patients, he said. It is easier to end a trial and "not have an individual interaction with the participants to make sure that they understood what had accrued as a result." People who are persistent often can get information about their assignment in a trial, but most trials do not provide the information as a service and in an easy-to-understand way.

Deborah Zarin, National Library of Medicine, observed that such decisions are the province of the institutional review board. Each trial would have its own guidelines, which would be laid out during the informed consent process. Rob Califf, Duke University Medical Center, agreed that there could be a standard and that it could be articulated fairly easily. "In general, there should be an obligation to inform participants in clinical trials about the result of the trial," he said. In most cases, participants can be informed about whether they got a treatment or placebo after the trial concludes. But making sure that a participant can put his or her individual assignment in the context of the results of a trial will take time and effort.

The Need for Leadership

Sharon Terry, president and chief executive officer of Genetic Alliance, concluded the workshop by observing that far too few people are participating in clinical research. When fewer than 10 percent of the population is engaged, either through participation in trials or through citizen science, the system is broken, she said. We need to find ways to involve many more people in biomedical research. Disease-specific and cross-disease advocacy or interest groups already exist that could foster this involvement. Industry has demonstrated its willingness to participate and increase its sharing of data. Researchers in specific areas have moved toward models of cooperation and data sharing.

Terry reminded the group that it is a mistake to refer to patients as "them" and to researchers as "us." Everyone is either a patient or a po-

tential future patient. "When we say we need to communicate with them, that's us." Public engagement in clinical trials has been lacking but as involvement becomes more commonplace, the resulting changes could be abrupt and dramatic.

What is needed, she concluded, is leadership from each of the stakeholders involved in clinical research. Each group needs to ask how it is impeding the flow of information and the conduct of research and take steps to remove those barriers. Each needs to move forward, said Terry, "for the sake of all the people who need us to do this quickly, efficiently, and carefully."

References

Alsheikh-Ali, A. A., W. Qureshi, M. H. Al-Mallah, and J. P. Ioannidis. 2011. Public availability of published research data in high-impact journals. *PLoS ONE* 6(9):e24357.

Alston, C., L. Paget, G. Halvorson, B. Novelli, J. Guest, P. McCabe, K. Hoffman, C. Koepke, M. Simon, S. Sutton, S. Okun, P. Wicks, T. Undem, V. Rohrbach, and I. Von Kohorn. 2012. *Communicating with patients on health care evidence.* Discussion Paper, Institute of Medicine, Washington, DC. http://www.iom.edu/evidence (accessed March 7, 2013).

Antithrombotic Trialists' Collaboration. 2002. Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients. *BMJ* 324(7329):71-86. Erratum in: *BMJ*, 2002, 324(7330):141.

Baigent, C., A. Keech, P. M. Kearney, L. Blackwell, G. Buck, C. Pollicino, A. Kirby, T. Sourjina, R. Peto, R. Collins, R. Simes, and Cholesterol Treatment Trialists' (CTT) Collaborators. 2005. Efficacy and safety of cholesterol-lowering treatment: Prospective meta-analysis of data from 90,056 participants in 14 randomised trials of statins. *Lancet* 366(9493):1267-1278. Errata in: *Lancet*, 2008, 371(9630):2084; *Lancet*, 2005, 366(9494):1358.

Califf, R. M., R. A. Harrington, L. K. Madre, E. D. Peterson, D. Roth, and K. A. Schulman. 2007. Curbing the cardiovascular disease epidemic: Aligning industry, government, payers, and academics. *Health Affairs (Millwood)* 26(1):62-74.

Chalmers, I. 2006. TGN1412 and the *Lancet's* solicitation of reports of Phase I trials. *Lancet* 368(9554):2206-2207.

Eichler, H.-G., E. Abadie, A. Breckenridge, H. Leufkens, and G. Rasi. 2012. Open clinical trial data for all? A view from regulators. *PLoS Medicine* 9(4):e1001202.

EMA (European Medicines Agency). 2012. *Access to clinical-trial data and transparency: Workshop report*. http://www.ema.europa.edu/docs/en_GB/

document_library/Report/2012/12/WC500135841.pdf (accessed March 7, 2013).

Gøtzsche, P. C. 2011. Why we need easy access to all data from all clinical trials and how to accomplish it. *Trials* 12:249.

Hart, B., A. Lundh, and L. Bero. 2012. Effect of reporting bias on meta-analyses of drug trials: Reanalysis of meta-analyses. *BMJ* 344:d7202.

HEW (Department of Health, Education and Welfare). 1979. *The Belmont Report: Ethical principles and guidelines for the protection of human subjects of research.* http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html (accessed March 7, 2013).

Hrachovec, J. B., and M. Mora. 2001. Reporting of 6-month vs. 12-month data in a clinical trial of celecoxib. *JAMA* 286(19):2398; author reply 2399-2400.

Ioannidis, J. P., D. B. Allison, C. A. Ball, I. Coulibaly, X. Cui, A. C. Culhane, M. Falchi, C. Furlanello, L. Game, G. Jurman, J. Mangion, T. Mehta, M. Nitzberg, G. P. Page, E. Petretto, and V. van Noort. 2009. Repeatability of published microarray gene expression analyses. *Nature Genetics* 41(2):149-155.

Kesselheim, A. S., C. T. Robertson, J. A. Myers, S. L. Rose, V. Gillet, K. M. Ross, R. J. Glynn, S. Joffe, and J. Avorn. 2012. A randomized study of how physicians interpret research funding disclosures. *New England Journal of Medicine* 367(12):1119-1127.

Kirwan, B. A., J. Lubsen, S. de Brouwer, F. J. van Dalen, S. J. Pocock, T. Clayton, N. Danchin, P. A. Poole-Wilson, and ACTION (A Coronary disease Trial Investigating Outcome with Nifedipine GITS) Investigators. 2008. Quality management of a large randomized double-blind multi-centre trial: The ACTION experience. *Contemporary Clinical Trials* 29(2):259-269.

Krumholz, H. M., Y. Wang, J. Chen, E. E. Drye, J. A. Spertus, J. S. Ross, J. P. Curtis, B. K. Nallamothu, J. H. Lichtman, E. P. Havranek, F. A. Masoudi, M. J. Radford, L. F. Han, M. T. Rapp, B. M. Straube, and S. L. Normand. 2009. Reduction in acute myocardial infarction mortality in the United States: Risk-standardized mortality rates from 1995-2006. *JAMA* 302(7):767-773.

Laine, C., S. N. Goodman, M. E. Griswold, and H. C. Sox. 2007a. Reproducible research: Moving toward research the public can really trust. *Annals of Internal Medicine* 146(6):450-453.

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

McCullough, P. A., and N. E. Lepor. 2005. The deadly triangle of anemia, renal insufficiency, and cardiovascular disease: Implications for prognosis and treatment. *Reviews in Cardiovascular Medicine* 6(1):1-10.

Michele, T. M., S. Pinheiro, and S. Iyasu. 2010. The safety of tiotropium—the FDA's conclusions. *New England Journal of Medicine* 363(12):1097-1099.

NRC (National Research Council). 2010. *A foundation for evidence-driven practice: A rapid learning system for cancer care: Workshop summary*. Washington, DC: The National Academies Press.

Peng, R. D., F. Dominici, and S. L. Zeger. 2006. Reproducible epidemiologic research. *American Journal of Epidemiology* 163(9):783-789.

Piwowar, H. A., R. S. Day, D. and B. Fridsma. 2007. Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3):e308. doi:10.1371/journal.pone.0000308

Platt, R. 2012. *Distributed networks can facilitate working remotely with research datasets derived from routinely collected electronic health information, often eliminating the need to transfer sensitive data*. Presentation at IOM Workshop on Sharing Clinical Research Data, October 4-5, Washington, DC.

Riley, R. D., P. C. Lambert, and G. Abo-Zaid. 2010. Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ* 340:c221.

Rodwin, M. A., and J. D. Abramson. 2012. Clinical trial data as a public good. *JAMA* 308(9):871-872.

Savage C. J., and A. J. Vickers. 2009. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. *PLoS ONE* 4(9):e7078. doi:10.1371/journal.pone.0007078.

Silverstein, F. E., G. Faich, J. L. Goldstein, L. S. Simon, T. Pincus, A. Whelton, R. Makuch, G. Eisen, N. M. Agrawal, W. F. Stenson, A. M. Burr, W. W. Zhao, J. D. Kent, J. B. Lefkowith, K. M. Verburg, and G. Geis. 2000. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: The CLASS study: A randomized controlled trial. *JAMA* 284(10):1247-1255.

Singh, S., Y. K. Loke, and C. D. Furberg. 2008. Inhaled anticholinergics and risk of major adverse cardiovascular events in patients with chronic obstructive pulmonary disease: A systematic review and meta-analysis. *JAMA* 300(12):1439-1450. Erratum in: *JAMA*, 2009, 301(12):1227-1230.

Vickers, A. J. 2006. Whose data set is it anyway? Sharing raw data from randomized trials. *Trials* 7:15.

Wagner, J. A., E. C. Wright, M. M. Ennis, M. Prince, J. Kochan, D. J. Nunez, B. Schneider, M. D. Wang, Y. Chen, S. Ghosh, B. J. Musser, and M. T. Vassileva. 2009. Utility of adiponectin as a biomarker predictive of glycemic efficacy is demonstrated by collaborative pooling of data from clinical trials conducted by multiple sponsors. *Clinical Pharmacology & Therapeutics* 86(6):619-625.

Walport, M., and P. Brest. 2011. Sharing research data to improve public health. *Lancet* 377(9765):537-539.

Williams, B., V. Entwistle, G. Haddow, and M. Wells. 2008. Promoting research participation: Why not advertise altruism? *Social Science & Medicine* 66(7):1451-1456.

Zarin, D. A. 2012. *Information loss as clinical trials data progresses from raw uncoded data to summary data*. Presentation at IOM workshop on Sharing Clinical Research Data, October 4-5, Washington, DC.

Zarin, D. A., T. Tse, R. J. Williams, R. M. Califf, and N. C. Ide. 2011. The ClinicalTrials.gov results database—update and key issues. *New England Journal of Medicine* 364(9):852-860.

A

Workshop Agenda

Sharing Clinical Research Data: An Institute of Medicine Workshop

October 4–5, 2012

National Academy of Sciences Building, Room 125
2101 Constitution Avenue, NW
Washington, DC 20418

Background:

Pharmaceutical companies, academic institutions, advocacy organizations, and government agencies such as the Food and Drug Administration and the National Institutes of Health have large quantities of clinical research data. Increased data sharing could facilitate scientific and public health advances, among other potential benefits to patients and society. Much of this information, however, is not transparent or shared beyond the data owner. More specifically, study results are not always published and where results are published, they typically only include summary-level data; participant-level data are privately held and rarely shared or revealed publicly.

This workshop will explore the benefits of and barriers to the sharing of clinical research data and will help identify strategies for enhancing the sharing both within and across sectors. To facilitate identification of key issues and potential solutions, the workshop will focus on data resulting from preplanned interventional studies of human subjects. While recognizing the importance of other data sources such as observational

studies and electronic health records, this focus was selected to encourage concrete problem-solving discussions over the course of a day-and-a-half–long meeting. Models and projects that involve sharing of other types of data will be considered during the workshop to the extent that these models provide lessons and best practices applicable to sharing preplanned interventional clinical research data.

The workshop is being jointly organized by the Institute of Medicine's Forum on Drug Discovery, Development, and Translation; Forum on Neuroscience and Nervous System Disorders; National Cancer Policy Forum; and Roundtable on Translating Genomic-Based Research for Health.

Meeting Objectives:

- Examine the benefits of sharing of clinical research data, and specifically clinical trial data, from all sectors and among these sectors, including, for example:
  - Benefits to the research and development enterprise
  - Benefits to the analysis of safety and efficacy
- Identify barriers and challenges to sharing clinical research data.
- Explore strategies to address these barriers and challenges, including the identification of priority actions and "low-hanging fruit" opportunities.
- Discuss strategies for using these potentially large datasets to facilitate scientific and public health advances.

October 4, 2012
Day One

8:30 a.m. Opening Remarks

SHARON TERRY, *Workshop Chair*
President and Chief Executive Officer
Genetic Alliance

SESSION I: BENEFITS OF SHARING CLINICAL RESEARCH DATA

Session Objectives:

- Provide an overview of the benefits of sharing clinical research data, specifically clinical trial data, and discuss advantages and disadvantages of sharing participant- versus summary-level data from individual trials as well as pooling data across multiple studies.
- Consider examples of scientific success stories that illustrate what can be accomplished when clinical trial data are shared.

8:40 a.m. Background and Session Objectives

WILLIAM POTTER, *Session Co-Chair*
Co-Chair Emeritus
Neuroscience Steering Committee
Foundation for the National Institutes of Health (FNIH)
Biomarkers Consortium

DEBORAH ZARIN, *Session Co-Chair*
Director, ClinicalTrials.gov
National Library of Medicine
National Institutes of Health

8:50 a.m. ***Fundamentals and Benefits of Sharing Participant-Level Clinical Trial Data***

ELIZABETH LODER
Clinical Epidemiology Editor, *BMJ*

9:10 a.m. ***Pooling Data from Multiple Clinical Trials to Answer Big Questions***

ROBERT CALIFF
Director, Duke Translational Medicine Institute
Professor of Medicine
Vice Chancellor for Clinical and Translational Research
Duke University Medical Center

9:30 a.m. ***Panel Discussion: Perspectives on the Benefits of Sharing Clinical Trial Data***

- Data sharing—what does it mean from your perspective?
- Considering the benefits and risks of sharing clinical research data, how extensively should it be shared to maximize new knowledge and ultimately patient benefit?

Panelists

HARLAN KRUMHOLZ
Harold H. Hines, Jr., Professor of Medicine and Epidemiology and Public Health
Yale University School of Medicine

MYLES AXTON
Editor
Nature Genetics

JESSE BERLIN
Vice President of Epidemiology
Janssen Research & Development, LLC

Panel Moderators

WILLIAM POTTER, *Session Co-Chair*
Co-Chair Emeritus
Neuroscience Steering Committee
FNIH Biomarkers Consortium

DEBORAH ZARIN, *Session Co-Chair*
Director, ClinicalTrials.gov
National Library of Medicine
National Institutes of Health

10:30 a.m. BREAK

SESSION II: DATA-SHARING MODELS: DESIGN, BEST PRACTICES, AND LESSONS LEARNED

Session Objectives:

- Present examples, best practices, and lessons learned from projects across the continuum of data-sharing opportunities (e.g., rapid publication of participant-level data, increased access to participant-level data for qualified researchers, or maximizing the use of clinical research data that are currently held in centralized locations by requiring sharing or access to subsets of data).
- Distill best practices and lessons learned that can be applied broadly to new projects to maximize the use of data from individual trials and/or data-pooling initiatives.

10:45 a.m. Background and Session Objectives

JEFFREY NYE, *Session Chair*
Vice President
Neuroscience Innovation and Partnership Strategy
Janssen Research & Development, LLC

10:55 a.m. ***The Limits of Summary Data Reporting: Lessons from ClinicalTrials.gov***

DEBORAH ZARIN
Director, ClinicalTrials.gov
National Library of Medicine
National Institutes of Health

11:10 a.m. ***Models That Increase Access and Use of Data from Individual Clinical Trials***

The DataSphere Project

CHARLES HUGH-JONES
Vice President, Medical Affairs North America
Sanofi Oncology, on behalf of the Life Sciences Consortium
CEO Roundtable on Cancer

Yale/Medtronic Experience

RICHARD KUNTZ
Senior Vice President
Chief Scientific, Clinical and Regulatory Officer
Medtronic, Inc.

11:40 a.m. ***Models That Foster Pooling and Analysis of Data***

FNIH Biomarkers Consortium Adiponectin Project

JOHN WAGNER
Vice President, Clinical Pharmacology
Merck & Co., Inc.

Novel Methods Leading to New Medications in Depression and Schizophrenia (NEWMEDS) Consortium

JONATHAN RABINOWITZ
Academic Lead, NEWMEDS Work Package on Advanced Data Analysis Techniques
Bar Ilan University

12:10 p.m. ***Series of Brief Presentations on Overcoming Challenges Facing Clinical Trial Data Sharing***

Challenge #1: Permissions

JENNIFER GEETTER
Partner
McDermott Will & Emery

Challenge #2: Techniques and Methodologies

JOHN IOANNIDIS (*via video conference*)
C.F. Rehnborg Chair in Disease Prevention
Stanford University

Challenge #3: Culture

KELLY EDWARDS
Acting Associate Dean, The Graduate School
Associate Professor, Bioethics and Humanities
University of Washington

12:40 p.m. Discussion among speakers, panelists, and audience
Discussant:

- Sally Okun, Health Data Integrity & Patient Safety, PatientsLikeMe

Discussion Moderator

JEFFREY NYE, *Session Chair*
Vice President
Neuroscience Innovation and Partnership Strategy
Janssen Research & Development, LLC

1:00 p.m. LUNCH

KEYNOTE CASE STUDY: DISTRIBUTED SYSTEMS FOR CLINICAL RESEARCH INFORMATION SHARING

1:30 p.m. RICHARD PLATT
Professor and Chair
Department of Population Medicine
Harvard Pilgrim Health Care Institute and Harvard Medical School

1:50 p.m. Discussion with Speaker and Audience

Discussion Moderator

SHARON TERRY, *Workshop Chair*
President and Chief Executive Officer
Genetic Alliance

SESSION III: STANDARDIZATION AND GOVERNANCE

Session Objectives:

- Receive an update on recent legislative and regulatory language regarding standardization of clinical research data and discuss how stakeholders are designing and implementing data standardization plans in response.
- Discuss the relative cost–benefit of data conversion of existing trial data versus building an infrastructure to improve data collection and sharing moving forward.
- Present case studies from data-sharing projects using different data standardization and governance models and consider lessons learned or best practices for the future.

2:00 p.m. Background and Session Objectives

FRANK ROCKHOLD, *Session Co-Chair*
Senior Vice President, Global Clinical Safety and Pharmacovigilance
GlaxoSmithKline Pharmaceuticals Research and Development

LYNN HUDSON, *Session Co-Chair*
Chief Science Officer and Executive Director
Coalition Against Major Diseases
Critical Path Institute

2:10 p.m. ***PDUFA Update on Data Standards***

RON FITZMARTIN
Senior Advisor, Office of Planning and Analysis
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

2:25 p.m. ***Standardization to Facilitate Data Sharing: Opportunities and Limitations***

CDISC Efforts to Support Clinical Research Data

REBECCA KUSH
President and Chief Executive Officer
Clinical Data Interchange Standards Consortium

HL7 Efforts to Support Clinical Care Data

CHARLES JAFFE
Chief Executive Officer
Health Level 7 International

Health Information Technology Perspective on Clinical Research Data Standards

SACHIN JAIN
Chief Medical Information and Innovation Officer
Merck & Co., Inc.

3:10 p.m. Discussion with speakers and audience

3:30 p.m. BREAK

3:45 p.m. ***Cost–Benefit Analysis of Retrospective vs. Prospective Data Standardization***

VICKI SEYFERT-MARGOLIS
Senior Advisor, Science Innovation and Policy
Office of the Chief Scientist
U.S. Food and Drug Administration

4:00 p.m. ***Case Studies: Standardization and Governance Models in Data Sharing***

Critical Path Institute and Coalition Against Major Diseases Alzheimer's Clinical Trial Database

CAROLYN COMPTON
President and Chief Executive Officer
Critical Path Institute

Translational Medicine Mart (tranSMART)

ERIC PERAKSLIS
Chief Information Officer and Chief Scientist, Informatics
U.S. Food and Drug Administration

4:30 p.m. ***Panel Discussion***

- Catalog new data-sharing challenges not yet discussed and provide suggestions for overcoming these challenges.
- Given the data standardization and governance models discussed, suggest a framework to guide the development of new data-sharing projects based on their purpose (e.g., regulatory approval with FDA, detecting safety signals, testing secondary hypotheses, etc.).

Panelists

LAURA LYMAN RODRIGUEZ
Director
Office of Policy, Communications and Education
National Human Genome Research Institute

MEREDITH NAHM
Associate Director for Clinical Research Informatics
Duke Translational Medicine Institute

NEIL DE CRESCENZO
Senior Vice President and General Manager
Oracle Health Sciences

MICHAEL CANTOR
Senior Director
Clinical Informatics and Innovation
Pfizer Inc.

Panel Moderators

FRANK ROCKHOLD, *Session Co-Chair*
Senior Vice President, Global Clinical Safety and Pharmacovigilance
GlaxoSmithKline Pharmaceuticals Research and Development

LYNN HUDSON, *Session Co-Chair*
Chief Science Officer and Executive Director
Coalition Against Major Diseases
Critical Path Institute

5:30 p.m. Adjourn Day One

October 5, 2012
Day Two

8:00 a.m. Opening Remarks

SHARON TERRY, *Workshop Chair*
President and Chief Executive Officer
Genetic Alliance

SESSION IV: INCENTIVIZING POLICY AND CULTURAL SHIFTS TO ENHANCE DATA SHARING

Session Objectives:

- Receive an update on clinical trial data transparency decisions in Europe.

- Explore current incentives for and against (i.e., benefits and risks of) data sharing within and across sectors and suggest mechanisms to encourage stakeholders to engage in a culture of data sharing.
- Identify existing and potential strategies, including technology-based approaches, for protecting patient privacy and confidentiality while facilitating data sharing.

8:10 a.m. Background and Session Objectives

ROBERT HARRINGTON, *Session Chair*
Arthur L. Bloomfield Professor of Medicine
Chair, Department of Medicine
Stanford University

8:20 a.m. ***Clinical Trial Data Transparency: European Medicines Agency Perspective***

HANS-GEORG EICHLER
Senior Medical Officer
European Medicines Agency

8:40 a.m. ***Clinical Research Data Sharing Practices and Attitudes***

ANDREW VICKERS
Attending Research Methodologist
Department of Epidemiology and Biostatistics
Memorial Sloan-Kettering Cancer Center

8:55 a.m. ***Overview of Data-Sharing Policies: Research Funders and Publishers***

STEVEN GOODMAN
Associate Dean for Clinical and Translational Research
Professor of Medicine & Health Research and Policy
Stanford University School of Medicine

9:10 a.m. ***Series of Presentations: Incentives for Data Sharing Within and Across Sectors***

Academic Perspectives

PETER DOSHI
Postdoctoral Fellow
Johns Hopkins University School of Medicine

BETH KOZEL
Instructor of Pediatrics
Division of Genetics and Genomic Medicine
St. Louis Children's Hospital and Washington University School of Medicine

Federal Research Funder Perspective

JOSEPHINE BRIGGS
Director, National Center for Complementary and Alternative Medicine
Director, National Center for Advancing Translation Sciences, Division of Clinical Innovation
National Institutes of Health

9:55 a.m. Discussion with speakers and audience

10:30 a.m. BREAK

10:45 a.m. ***Facilitating Patient Ownership of Clinical Trial Data: Technical Challenges and Opportunities***

JOHN WILBANKS
Director
Sage Bionetworks

DEVEN MCGRAW
Director, Health Privacy Project
Center for Democracy and Technology

11:15 a.m. Discussion with speakers and audience

SESSION V: NEXT STEPS AND FUTURE DIRECTIONS

Session Objectives:

- Discuss key themes from the workshop.
- Based on workshop presentations and discussions, identify potential next steps and priority actions for data-sharing stakeholders to take action.
- Highlight potential opportunities and challenges that are currently on the horizon, but may become more salient as technology evolves and/or data sharing becomes more pervasive.

11:30 a.m. Background and Session Objectives

SHARON TERRY, *Workshop Chair*
President and Chief Executive Officer
Genetic Alliance

11:40 a.m. ***Session Chair Reports*** (*5 minutes per session*)

WILLIAM POTTER, *Session I Co-Chair*
Co-Chair Emeritus
Neuroscience Steering Committee
FNIH Biomarkers Consortium

DEBORAH ZARIN, *Session I Co-Chair*
Director, ClinicalTrials.gov
National Library of Medicine
National Institutes of Health

JEFFREY NYE, *Session II Chair*
Vice President
Neuroscience Innovation and Partnership Strategy
Janssen Research & Development, LLC

FRANK ROCKHOLD, *Session III Co-Chair*
Senior Vice President
Global Clinical Safety and Pharmacovigilance
GlaxoSmithKline Pharmaceuticals Research and Development

LYNN HUDSON, *Session III Co-Chair*
Chief Science Officer and Executive Director
Coalition Against Major Diseases, Critical Path Institute

ROBERT HARRINGTON, *Session IV Chair*
Arthur L. Bloomfield Professor of Medicine
Chair, Department of Medicine
Stanford University

12:00 p.m. ***Closing Discussion with Session Chairs, Panelists, and Audience Led by Workshop Chair***

JOSEPHINE BRIGGS
Director, National Center for Complementary and Alternative Medicine
Director, National Center for Advancing Translation Sciences, Division of Clinical Innovation
National Institutes of Health

MICHAEL ROSENBLATT
Executive Vice President and Chief Medical Officer
Merck & Co., Inc.

JAY "MARTY" TENENBAUM
Founder and Chair
Cancer Commons

JANET WOODCOCK
Director, Center for Drug Evaluation and Research
U.S. Food and Drug Administration

12:45 p.m. ADJOURN

B

List of Data-Sharing Initiatives[1]

Alzheimer's Disease Neuroimaging Initiative (ADNI)
http://adni-info.org

Analgesic Clinical Trials Innovation, Opportunities and Networks (ACTION) Initiative
http://www.fda.gov/AboutFDA/PartnershipsCollaborations/PublicPrivatePartnershipProgram/ucm231130.htm

Arch2POCM (Archipelago to Proof of Clinical Mechanism [Phase IIa])
http://sagebase.org/WP/arch

Biogrid Australia
http://www.biogrid.org.au/wps/portal

Biomarkers Consortium Project on Adiponectin
http://www.biomarkersconsortium.org/press_release_adiponectin_predictive_biomarker.php

[1]This list of data-sharing projects is not an exhaustive list, and inclusion does not denote endorsement. The list includes projects that share different types of data and health information (e.g., genetic information, observational data, etc.) as potential sources of best practices/lessons learned that may be applicable to initiatives focused on sharing data from preplanned interventional studies of human subjects.

Biosense (CDC)
http://www.cdc.gov/biosense

caBIG (Cancer Biomedical Informatics Grid)
http://cabig.cancer.gov

CAMD (Coalition Against Major Diseases) and C-Path Alzheimer's Database
http://www.c-path.org/News/CDISCTAStds%20PR-24June2012.pdf

Cancer Commons
http://cancercommons.org/wp-content/themes/cancer_commons/docs/cancer_commons_whitepaper.pdf
http://www.cancercommons.org/about

CDC's Chronic Fatigue Syndrome Wichita Clinical Study
http://www.cdc.gov/rdc/B1DataType/Dt132.htm
http://www.cfids.org/advocacy/testimony-vernon-oct2008.pdf

Clinical Trial Comparator Arm Partnership (CTCAP)
http://sagebase.org/partners/CTCAP.php

ClinicalTrials.gov
http://www.nlm.nih.gov/pubs/factsheets/clintrial.html

The database of Genotypes and Phenotypes (dbGaP)
http://www.ncbi.nlm.nih.gov/gap

DataSphere Project (CEO Roundtable on Cancer)
http://ceo-lsc.org/projectdatasphere

The "ePlacebo" Database
http://www.genome.gov/19518664

Informatics for Integrating Biology and the Bedside (i2b2)
https://www.i2b2.org

Innovative Medicines Initiative (IMI) European Medical Information Framework (EMIF)
http://www.imi.europa.eu/content/home

International Severe Adverse Events Consortium (iSAEC)
http://www.saeconsortium.org

Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and UCSF Biobank
http://www.dor.kaiser.org/external/DORExternal/rpgeh/index.aspx

Mini-Sentinel
http://mini-sentinel.org/about_us

NEWMEDS
http://www.newmeds-europe.com

One Mind Initiative
http://1mind4research.org/programs

Parkinson's Progression Markers Initiative (PPMI)
http://ppmi-info.org

PatientsLikeMe
http://www.patientslikeme.com/about

Sage Bionetworks
http://sagebase.org/info/index.php

Structural Genomics Consortium (SGC)
http://www.thesgc.org

tranSMART
http://www.transmartproject.org/index.html

C

Participant Biographies

Sharon F. Terry, M.A. (*Workshop Chair*), is president and CEO of the Genetic Alliance, a network of more than 10,000 organizations, 1,200 of which are disease advocacy organizations. Genetic Alliance improves health through the authentic engagement of communities and individuals. It develops innovative solutions through novel partnerships, connecting consumers to smart services. She is the founding CEO of PXE International, a research advocacy organization for the genetic condition pseudoxanthoma elasticum (PXE). As codiscoverer of the gene associated with PXE, she holds the patent for ABCC6 and has assigned her rights to the foundation. She developed a diagnostic test and is conducting clinical trials. Ms. Terry is also a cofounder of the Genetic Alliance Registry and Biobank. She is the author of more than 90 peer-reviewed articles. In her focus at the forefront of consumer participation in genetics research, services, and policy, she serves in a leadership role on many of the major international and national organizations, including the Institute of Medicine (IOM) Health Sciences Policy Board, the National Coalition for Health Professional Education in Genetics board, and the International Rare Disease Research Consortium Interim Executive Committee. She is a member of the IOM Roundtable on Translating Genomic-Based Research for Health. She is on the editorial boards of several journals. She was instrumental in the passage of the Genetic Information Nondiscrimination Act. In 2005, she received an honorary doctorate from Iona College for her work in community engagement; the first Patient Service Award from the University of North Carolina Institute for Pharmacogenomics and Individualized Therapy in 2007; the Research!America Distinguished Organization Advocacy Award in

2009; and the Clinical Research Forum and Foundation's Annual Award for Leadership in Public Advocacy in 2011. She is also an Ashoka Fellow.

Myles Axton, Ph.D., is editor of *Nature Genetics*. He was a university lecturer in molecular and cellular biology at the University of Oxford and a fellow of Balliol College from 1995 to 2003. He obtained his degree in genetics at Cambridge in 1985, and his doctorate at Imperial College in 1990. Between 1990 and 1995, he did postdoctoral research at Dundee and at the Massachusetts Institute of Technology's (MIT's) Whitehead Institute. His research made use of the advanced genetics of Drosophila to study genome stability by examining the roles of cell cycle regulators in life cycle transitions. His interests broadened into human genetics, genomics, and systems biology through lecturing and from tutoring biochemists, zoologists, and medical students from primary research papers. Helping to establish Oxford's innovative research M.Sc. in integrative biosciences led Dr. Axton to realize the importance of the integrative overview of biomedical research. As a full-time professional editor, he is now in a position to use this perspective to help coordinate research in genetics.

Jesse A. Berlin, Sc.D., spent 15 years as a faculty member at the University of Pennsylvania, in the Center for Clinical Epidemiology and Biostatistics, under the direction of Dr. Brian Strom. He left the University of Pennsylvania to join Janssen Research & Development, where he is currently vice president of epidemiology. He has authored or coauthored more than 230 publications in a wide variety of clinical and methodological areas, including papers on the study of meta-analytic methods as applied to both randomized trials and epidemiology. He served on an Institute of Medicine committee that developed recently released recommendations for the use of systematic reviews in clinical effectiveness research. He currently serves on the Scientific Advisory Committee to the Observational Medical Outcomes Partnership, a public-private partnership aimed at understanding methodology for assessing drug safety in large administrative databases. He is also a fellow of the American Statistical Association.

Josephine P. Briggs, M.D., is director of the National Center for Complementary and Alternative Medicine (NCCAM), and acting director, Division of Clinical Innovation, National Center for Advancing Transla-

tional Sciences, National Institutes of Health (NIH). An accomplished researcher and physician, Dr. Briggs received her A.B. in biology from Harvard–Radcliffe College and her M.D. from Harvard Medical School. She completed her residency training in internal medicine and nephrology at the Mount Sinai School of Medicine, followed by a fellowship at Yale University. She then worked as a research scientist at the Physiology Institute at the University of Munich. In 1985, Dr. Briggs moved to the University of Michigan, where she held several academic positions, including associate chair for research in the Department of Internal Medicine and professorships in the Division of Nephrology, Department of Internal Medicine, and the Department of Physiology. She joined the NIH in 1997 as director of the Division of Kidney, Urologic, and Hematologic Diseases at the National Institute of Diabetes and Digestive and Kidney Diseases. In 2006, Dr. Briggs accepted a position as senior scientific officer at the Howard Hughes Medical Institute. In 2008, she returned to the NIH as director of the NCCAM. Dr. Briggs has published more than 175 research articles, book chapters, and scholarly publications. She has served on the editorial boards of several journals, and was deputy editor for the *Journal of Clinical Investigation*. Dr. Briggs is an elected member of the American Association of Physicians and the American Society of Clinical Investigation and a fellow of the American Association for the Advancement of Science. She is a recipient of many awards and prizes, including the Volhard Prize of the German Nephrological Society, the Alexander von Humboldt Scientific Exchange Award, and NIH Director's Awards for her role in the development of the Trans-NIH Type I Diabetes Strategic Plan and her leadership of the Trans-NIH Zebrafish committee. Dr. Briggs is also a member of the NIH Steering Committee, the highest governing board at the NIH.

Robert M. Califf, M.D., is the vice chancellor for Clinical and Translational Research, director of the Duke Translational Medicine Institute (DTMI), and professor of medicine in the Division of Cardiology at Duke University Medical Center in Durham, North Carolina. He leads a multifaceted organization that seeks to transform how scientific discoveries are translated into improved health outcomes. Prior to leading the DTMI, he was the founding director of the Duke Clinical Research Institute (DCRI). He is editor in chief of the *American Heart Journal*, the oldest cardiovascular specialty journal, and a practicing cardiologist at Duke University Medical Center. Dr. Califf attended Duke University, graduating summa cum laude and Phi Beta Kappa. He remained at Duke

for medical school, where he was selected for the Alpha Omega Alpha medical honor society. After graduating from Duke University School of Medicine, he completed a residency in internal medicine at the University of California, San Francisco, then returned to Duke for a cardiology fellowship. Dr. Califf is board certified in internal medicine and cardiology, and was named a Master of the American College of Cardiology in 2006. An international leader in the fields of cardiovascular medicine, health care outcomes, quality of care, and medical economics, he has authored or coauthored more than 1,000 peer-reviewed articles and is among the most frequently cited authors in medicine. He is also a contributing editor for TheHeart.org, an online information resource for health care professionals working in the field of cardiovascular medicine. As founder and a decade-long director of the DCRI, Dr. Califf led many landmark clinical trials and health services research projects, and remains actively involved in designing, leading, and conducting multinational clinical trials. Under his guidance, DCRI grew into an organization with more than 1,000 employees and an annual budget of more than $100 million; its umbrella organization, the DTMI, now has an annual budget of more than $300 million. Supported in part by a Clinical and Translational Science Award (CTSA) from the National Institutes of Health, the DTMI works with government agencies, academic partners, research foundations, and the medical products industry to conduct innovative research spanning multiple therapeutic arenas and scientific disciplines. Dr. Califf serves as co-chair of the first Principal Investigators Steering Committee of the CTSA. He has served on the Food and Drug Administration's (FDA's) Cardiorenal Advisory Panel, and on the Institute of Medicine's Committee on Identifying and Preventing Medication Errors, and Committee on Nutritional Biomarkers. In 2008, he was part of the subcommittee of the FDA's Science Board that recommended sweeping reform of the agency's science base. He was also a member of the IOM committees that recommended Medicare coverage of clinical trials and the removal of ephedra from the market. Dr. Califf is currently a member of the Institute of Medicine (IOM) Forum on Drug Discovery, Development, and Translation and a member of the National Advisory Council on Aging. Reflecting his interests in health care quality, Dr. Califf was the founding director of the coordinating center of the Centers for Education & Research on Therapeutics, a public-private partnership that seeks to improve the use of medical products through research and education. He is currently co-chair of the Clinical Trials Transformation Initiative, a public-private partnership focused on improving the clinical trials sys-

tem. He is also chair of the Clinical Research Forum, an organization of academic health and science system leaders devoted to improving the clinical research enterprise.

Michael N. Cantor, M.D., M.A., FACP, is senior director, Information Strategy and Analytics, in Pfizer's Clinical Informatics and Innovation group. His work focuses on leveraging data reuse and integration to support future horizons of scientific decision support for precision medicine. He is currently coleading several initiatives around the secondary use of clinical data, including Pfizer's ePlacebo/eControls database, as well as its comprehensive Clinical Lab Data Catalog. He created and coleads the MEDIC (Multisite Electronic Data Infectious Disease Consortium) project, which aims to partner with academic medical centers to perform observational studies using data from electronic medical record (EMR) systems. He has served as an advisor to programs across each of Pfizer's Business Units, as well as the Worldwide Research and Development organization, on the role of health care IT in advancing their strategic priorities. Dr. Cantor previously led Pfizer Business Technology's "Data Without Borders" strategy, with the aim of advancing data sharing and reuse, both internally and externally, to advance Precision Medicine. He has been a member of American Medical Informatics Association's public policy committee for 6 years, and led the committee's initiative to update its positions on data stewardship and reuse. Prior to joining Pfizer, Dr. Cantor was the chief medical information officer for the South Manhattan Healthcare Network of the New York City Health and Hospitals Corporation, based at Bellevue Hospital in Manhattan. His work there focused on developing the network's EMR system to improve patient safety and on using the network's clinical data warehouse for research. He continues to see patients at Bellevue, and is a clinical assistant professor of medicine at the New York University School of Medicine. Dr. Cantor completed his residency in Internal Medicine and Informatics Training at Columbia. He has an M.D. from Emory University and an A.B. from Princeton.

Carolyn Compton, M.D., Ph.D., is the president and chief executive officer of Critical Path Institute. She was most recently the director of the Office of Biorepositories and Biospecimen Research and the executive director of the Cancer Human Biobank project at the National Cancer Institute (NCI). In these capacities, she had leadership responsibility for strategic initiatives that included the Innovative Molecular Analysis

Technologies for Cancer program; the Biospecimen Research Network program; and the NCI Community Cancer Centers project. She is an adjunct professor of pathology at the Johns Hopkins School of Medicine. She received her M.D. and Ph.D. from Harvard Medical School and the Harvard Graduate School of Arts and Sciences. She trained in pathology at Harvard's Brigham and Women's Hospital and is boarded in both Anatomic Pathology and Clinical Pathology. She came to the NCI from McGill University, where she had been the Strathcona Professor and Chair of Pathology and the pathologist in chief of McGill University Health Center from 2000 to 2005. Prior to this, she had been a professor of pathology at Harvard Medical School, the director of gastrointestinal pathology at the Massachusetts General Hospital, and the pathologist in chief of the Shriners Hospital for Crippled Children, Boston Burns Unit for 15 years. During this time she served as the chair of the Pathology Committee of the Cancer and Leukemia Group B for 12 years. Her research interests are in colon and pancreatic cancer as well as epithelial biology and wound healing. Dr. Compton has held many national and international leadership positions in pathology and cancer-related professional organizations. She is a fellow of the College of American Pathologists (CAP) and a fellow of the Royal Society of Medicine. Currently, she is chair of the American Joint Committee on Cancer, serves on the Executive Committee of the Commission on Cancer (COC) of the American College of Surgeons, and serves as the pathology section editor for *Cancer*. She is a past chair of the Cancer Committee of the College of American Pathologists and was editor of the first edition of the CAP Cancer Protocols (Reporting on Cancer Specimens) used as standards for COC accreditation. Among her awards are the International Society for Biological and Environmental Repositories Award for Outstanding Achievement in Biobanking, the National Institutes of Health (NIH) Director's Award, the NIH Award of Merit, and the CAP Frank W. Hartman Award. She has published more than 500 original scientific papers, reports, review articles, and books.

Neil de Crescenzo, M.B.A., is senior vice president and general manager for Health Sciences at Oracle. He is responsible for managing Oracle's solution groups, strategic planning, product development, and sales, service, and support for the industry solutions sold into the health care and life sciences markets worldwide. Mr. de Crescenzo brings more than 20 years of operational and information technology (IT) leadership across health care and life sciences to his work with customers and partners

worldwide. Prior to joining Oracle, Mr. de Crescenzo held a number of leadership positions during his decade at IBM Corporation, working with health care and life sciences clients throughout the world. Prior to entering the IT industry, he held leadership positions in health care operations at multiple medical centers and a major health insurer. Mr. de Crescenzo began his career in investment banking, working with U.S. and European clients in the areas of corporate finance and mergers and acquisitions. Mr. de Crescenzo has been a keynote speaker at numerous industry conferences worldwide and is quoted frequently on industry issues. In 2005, he was named one of the "Top 25 Most Influential Consultants" by *Consulting* magazine. Mr. de Crescenzo has a B.A. in political science from Yale University and an M.B.A. in high technology from Northeastern University.

Peter Doshi, Ph.D., is a postdoctoral fellow in comparative effectiveness research at the Johns Hopkins University School of Medicine. His overarching research interests are in improving the basis for credible evidence synthesis to support and improve the quality of evidence-based medical and health policy–related decision making. In 2009, he joined a Cochrane systematic review team evaluating neuraminidase inhibitors for the treatment and prevention of influenza. Rather than focusing on publications, the review evaluates regulatory information, including clinical study reports. He received his A.B. in anthropology from Brown University, A.M. in East Asian studies from Harvard University, and Ph.D. in history, anthropology, and science, technology and society from the Massachusetts Institute of Technology.

Kelly Edwards, Ph.D., is an associate professor in the University of Washington (UW) School of Medicine's Department of Bioethics and Humanities. Dr. Edwards also is a core faculty member for the UW Institute for Public Health Genetics. She received both her M.A. in medical ethics and her Ph.D. in philosophy of education from the UW. Dr. Edwards' work incorporates communication and public engagement as an ethical obligation for clinicians and researchers. She is the director of the Ethics and Outreach Core for the UW Center for Ecogenetics and Environmental Health, which is funded by the National Institute of Environmental Health Sciences. She also is a codirector of the Regulatory Support and Bioethics Core for the Institute for Translational Health Sciences (ITHS), a partnership of the UW, Fred Hutchinson Cancer Research Center, Seattle Children's, and other regional institutions and

community and tribal groups. Funded by the National Institutes of Health (NIH), the ITHS assists researchers with translating their scientific discoveries into practice. In addition, Dr. Edwards is a lead investigator with the UW Center for Genomics and Healthcare Equality, funded by the NIH's National Human Genome Research Institute. Since 2004, she has been the faculty advisor for the Forum on Science, Ethics and Policy, groups of graduate and professional students and postdoctoral fellows at the UW and University of Colorado who promote dialogue on issues concerning science and society. To further engage people in conversations about ethical dimensions of science and medicine, Dr. Edwards has facilitated Community Conversations and the Public Health Café, a series of events hosted in Seattle by the Northwest Association for Biomedical Research. Dr. Edwards contributes to issues of ethical research practices with the Genetic Alliance, a health advocacy organization; Sage Bionetworks, a local nonprofit; and the Institute of Medicine. Her courses include "Inquiry-Based Science Communication," "Applied Research Ethics," "Community-Based Participatory Research: A Model for Genetics Research with Native American Communities?" and "Public Commentary on Ethical Issues in Public Health Genetics." She is associate editor of *BMC Medical Research Methodology* and a reviewer for several journals. Dr. Edwards serves on the UW School of Medicine's Continuous Professional Improvement Committee and is a former member of Medicine's Standing Committee on Issues of Women Faculty, the Student Progress Committee, and the Committee on Research and Graduate Education. She is a current member of the UW Graduate School Committee on Interdisciplinary Education.

Hans-Georg Eichler, M.D., M.Sc., is the senior medical officer at the European Medicines Agency (EMA) in London, where he is responsible for coordinating activities among the EMA's scientific committees and advising on scientific and public health issues. In 2011, Dr. Eichler was the Robert E. Wilhelm Fellow at the Massachusetts Institute of Technology (MIT) Center for International Studies, participating in a joint research project under the MIT's New Drug Development Paradigms initiative. He divided his time between the MIT and the EMA in London. Prior to joining the EMA, Dr. Eichler was at the Medical University of Vienna in Austria for 15 years. He was vice rector for Research and International Relations since 2003, and professor and chair of the Department of Clinical Pharmacology since 1992. His other previous positions include president of the Vienna School of Clinical Research and co-chair

of the Committee on Reimbursement of Drugs of the Austrian Social Security Association. His industry experience includes time spent at Ciba-Geigy Research Labs in the United Kingdom, and Outcomes Research at Merck & Co. in New Jersey. Dr. Eichler graduated with an M.D. from Vienna University Medical School and an M.Sc. in toxicology from the University of Surrey in Guildford, United Kingdom. He trained in internal medicine and clinical pharmacology at the Vienna University Hospital as well as at Stanford University.

Ron Fitzmartin, Ph.D., M.B.A., is senior advisor, Office of Planning and Informatics, Center for Drug Evaluation and Research, Food and Drug Administration (FDA). Prior to joining FDA, Dr. Fitzmartin held scientific and technical leadership positions at Decision Analytics, Daiichi Sankyo, Inc., Daiichi Medical Research, Inc., and Purdue Pharma L.P. In addition, he served as statistician at both the U.S. Census Bureau and the U.S. Department of the Navy. Dr. Fitzmartin was elected a member of the Board of Directors and President of the Drug Information Association from 2007 to 2009. Dr. Fitzmartin has been a frequent presenter at many industry meetings and has authored numerous articles in areas such as informatics, pharmacovigilance, clinical data management, regulatory compliance, and R&D strategy. Dr. Fitzmartin received a Ph.D. in statistics from the University of Maryland, an M.B.A. from the University of New Haven, and an M.S. and B.S. from Southern Connecticut State University.

Jennifer S. Geetter, J.D., is a partner in the law firm of McDermott Will & Emery LLP and is based in the firm's Washington, DC, office. She focuses her practice on emerging biotechnology and safety issues, advising hospital, industry, insurance, and provider clients on matters relating to research, drug and device development, off-label use, personalized medicine, formulary compliance, privacy and security, electronic health records and data strategy initiatives, patient safety, conflicts of interest, scientific review and research misconduct, internal hospital disciplinary proceedings, and emerging issues in secondary research concerning biological samples and data warehousing. She also assists health care clients in implementing research strategies, structuring research operational and compliance infrastructure, and developing guidelines for the appropriate relationships between providers and industry. She is a frequent speaker on these topics. Ms. Geetter is a member of the firm's Life Sciences Affinity Group and Personalized Medicine Team. She is also a co-chair of

the Pro Bono and Community Service Committee for the Washington, DC, office. She sits on McDermott's National Pro Bono Committee and the American Health Lawyers Association Life Sciences Section Steering Committee. Ms. Geetter is listed as a leading individual in health care in Washington, DC, in *Chambers USA 2008: America's Leading Lawyers for Business*. She was recognized as a STAR Mentor in the McDermott University Mentoring Program for 2006-2007. She is also a member of McDermott's Gender Diversity Committee. In 2003, Ms. Geetter was awarded the ACE Founders Award for her pro bono efforts as part of a team of McDermott lawyers on behalf of a group of low-income residents of a Boston neighborhood. She received the firm's Outstanding Achievement Award for Commitment to Pro Bono and Service to the Community in 2004 and serves on the firm's national coordinating committee for pro bono activities. Ms. Geetter is admitted to practice in Massachusetts, New York, and Washington, DC.

Steven N. Goodman, M.D., M.H.S., Ph.D., is professor of medicine and health policy, and research and associate dean for clinical and translational research at Stanford University. Before joining Stanford in 2011, Dr. Goodman spent two decades on the Johns Hopkins medical faculty as professor of oncology in the division of biostatistics, with appointments in the departments of Pediatrics, Biostatistics, and Epidemiology in the Johns Hopkins Schools of Medicine and Public Health. He has been editor of *Clinical Trials: Journal of the Society for Clinical Trials* since 2004 and senior statistical editor for the *Annals of Internal Medicine* since 1987. He has served on a wide range of Institute of Medicine committees, including Agent Orange and Veterans, Immunization Safety, the Committee on Alternatives to the Daubert Standards, Treatment of PTSD (Posttraumatic Stress Disorder) in Veterans, and most recently co-chaired the Committee on Ethical and Scientific Issues in Studying the Safety of Approved Drugs, whose report was released in 2012. Dr. Goodman was appointed by the Government Accountability Office to serve on the Methodology Committee of the Patient Centered Outcomes Research Institute and is a scientific advisor to the Medical Advisory Panel of the National Blue Cross/Blue Shield Technology Evaluation Center. He served on the Surgeon General's committee to write the 2004 report on the health consequences of smoking. Dr. Goodman received a B.A. from Harvard; an M.D. from New York University; trained at Washington University in pediatrics, in which he was board certified; and received an M.H.S. in biostatistics and a Ph.D. in epidemiology from

Johns Hopkins University. He writes and teaches on evidence evaluation and inferential, methodologic, and ethical issues in epidemiology and clinical research.

Robert A. Harrington, M.D., is the Arthur L. Bloomfield Professor of Medicine and chair of the Department of Medicine at Stanford University. He received his undergraduate degree in English from the College of the Holy Cross in Worcester, MA. He attended Dartmouth Medical School and received his M.D. from Tufts University School of Medicine in 1986. He was an intern, resident, and the chief medical resident in internal medicine at the University of Massachusetts Medical Center. He was a fellow in cardiology at Duke University Medical Center, where he received training in interventional cardiology and research training in the Duke Databank for Cardiovascular Diseases. Dr. Harrington was previously the director of the Duke Clinical Research Institute. His research interests include evaluating antithrombotic therapies to treat acute ischemic heart disease and to minimize the acute complications of percutaneous coronary procedures; studying the mechanism of disease of the acute coronary syndromes; understanding the issue of risk stratification in the care of patients with acute ischemic coronary syndromes; and trying to better understand and improve on the methodology of clinical trials. He is the recipient of a National Institutes of Health Roadmap contract to investigate "best practices" among clinical trial networks. He has authored more than 400 peer-reviewed manuscripts, reviews, book chapters, and editorials. He is an associate editor of the *American Heart Journal* and an editorial board member for the *Journal of the American College of Cardiology*. He is a senior editor of the 13th edition of Hurst's *The Heart*. He is a fellow of the American College of Cardiology, the American Heart Association, the European Society of Cardiology, the Society of Cardiovascular Angiography and Intervention, and the American College of Chest Physicians. He recently served as a member and the chair of the Food and Drug Administration Cardiovascular and Renal Drugs Advisory Committee.

Lynn D. Hudson, Ph.D., serves as the chief science officer for the Critical Path Institute and executive director of the Multiple Sclerosis Consortium. She received a B.S. in biochemistry at the University of Wisconsin, a Ph.D. in genetics and cell biology at the University of Minnesota, and postdoctoral training at Harvard Medical School and Brown University. For most of her career, Dr. Hudson has been a bench neuroscientist at the

National Institutes of Health (NIH), where she directed the Office of Science Policy Analysis from 2006 to 2011. As a major source for policy analysis within the NIH Office of the Director, her office covered a wide spectrum of sensitive and emerging issues and oversaw a number of programs, including the American Association for the Advancement of Science/NIH Science Policy Fellowship program, the NIH's contract with the National Academy of Sciences, and the Public–Private Partnership Program. Her policy team's awards cite contributions to the NIH's congressional justification, biennial report, implementation of the NIH Reform Act, stem cell guidelines, and comparative effectiveness research. Dr. Hudson has conducted research in the National Institute of Neurological Disorders and Stroke (NINDS) intramural program, where she was chief of the Section of Developmental Genetics for 16 years. She received the NIH Merit Award for her discovery of the causative mutations in the neurologic disorder Pelizaeus-Merzbacher Disease (PMD), and an NINDS Award for educational outreach efforts. Her research focused on defining the network of genes involved in the development of glial cells, with the goal of designing strategies to overcome glial dysfunction in inherited or acquired neurological diseases. She served as an officer for the American Society for Neurochemistry, as an officer on the scientific advisory board of the PMD Foundation, and as an advisor for a number of granting agencies and disease foundations, including the National Multiple Sclerosis Society.

Charles Hugh-Jones, M.D., is vice president and head, Medical Affairs North America, Oncology, Hematology, and Solid Organ Transplant at Sanofi. He has previously served as vice president and head, Oncology Medical Affairs, North America, for Sanofi-Aventis; vice president of brand management and vice president of medical affairs at Enzon Pharmaceuticals; executive director, Medical Affairs, at Schering AG/Berlex/Bayer-Schering; director of medical affairs for Schering AG, Global Business Unit; senior medical advisor for Schering AG, United Kingdom; and specialist registrar at Hammersmith Hospital, United Kingdom. Dr. Hugh-Jones received his M.D. from the University of London. He is board certified in internal medicine and has additional Higher Specialist Training (United Kingdom) in diagnostic and interventional radiology.

John P. A. Ioannidis, D.Sc., M.D., holds the C.F. Rehnborg Chair in Disease Prevention at Stanford University and is professor of medicine, professor of health research and policy, and director of the Stanford Pre-

vention Research Center at Stanford University School of Medicine, professor of statistics (by courtesy) at Stanford University School of Humanities and Sciences, member of the Stanford Cancer Center and of the Stanford Cardiovascular Institute, and affiliated faculty of the Woods Institute for the Environment. From 1999 until 2010, he chaired the Department of Hygiene and Epidemiology at the University of Ioannina School of Medicine in Greece, where he had been a tenured professor since 2003. Dr. Ioannidis graduated in the top rank of his class from the School of Medicine, University of Athens, in 1990 and earned a doctorate in biopathology. He trained at Harvard and Tufts, specializing in internal medicine and infectious diseases, and then held positions at the National Institutes of Health, Johns Hopkins University School of Medicine, and Tufts University School of Medicine before returning to Greece in 1999. He has been adjunct faculty for the Tufts School of Medicine since 1996, with the rank of professor since 2002. Since 2008 he led the Genetics/Genomics component of the Tufts Clinical and Translational Science Institute and the Center for Genetic Epidemiology and Modeling of the Tufts Institute for Clinical Research and Health Policy Studies at Tufts Medical Center. He is also an adjunct professor of epidemiology at the Harvard School of Public Health and a visiting professor of epidemiology and biostatistics at Imperial College London. Dr. Ioannidis is a member of the executive board of the Human Genome Epidemiology Network. He has served as president of the Society for Research Synthesis Methodology, as a member of the editorial board of 27 leading international journals (including *PLoS Medicine*, *Lancet*, *Annals of Internal Medicine*, *Journal of the National Cancer Institute*, *Science Translational Medicine*, *Molecular and Cellular Proteomics*, *AIDS*, *International Journal of Epidemiology*, *Journal of Clinical Epidemiology*, *Clinical Trials*, *Cancer Treatment Reviews*, *Open Medicine*, and *PLoS ONE*), and as editor in chief of the *European Journal of Clinical Investigation* from 2010 to 2014. He has given more than 250 invited and honorary lectures. He has received several awards, including the European Award for Excellence in Clinical Science for 2007, and was inducted into the Association of American Physicians in 2009 and into the European Academy of Cancer Sciences in 2010. His *PLoS Medicine* paper "Why Most Published Research Findings Are False" has been the most accessed article in the history of the Public Library of Science. The *Atlantic* selected Ioannidis as the Brave Thinker scientist for 2010, claiming that he "may be one of the most influential scientists alive."

Charles Jaffe, M.D., is the CEO of Health Level 7, where he serves as the organization's global ambassador, fostering relationships with key industry stakeholders. A 37-year veteran of the health care IT industry, Dr. Jaffe was previously the senior global strategist for the Digital Health Group at Intel Corporation, vice president of life sciences at Science Applications International Corporation, and director of medical informatics at AstraZeneca Pharmaceuticals. He completed his medical training at Johns Hopkins and Duke Universities, and was a postdoctoral fellow at the National Institutes of Health and at Georgetown University. Formerly, he was president of InforMed, a consultancy for research informatics. Over the course of his career, he has been the principal investigator for more than 200 clinical trials, and has served in various leadership roles in American Medical Informatics Association. He has been a board member of leading organizations for information technology standards, and served as chair of a national institutional review board. Most recently, he held an appointment in the Department of Engineering at Penn State University. Dr. Jaffe has been the contributing editor for several journals and has published on a range of subjects, including clinical management, informatics deployment, and health care policy.

Sachin H. Jain, M.D., M.B.A., is chief medical information and innovation officer (CMIO) at Merck and a lecturer in health care policy at Harvard Medical School. He also serves as an attending hospitalist physician at the Boston VA-Boston Medical Center. In his capacity as Merck's CMIO, Dr. Jain is working to establish and manage global alliances to use health data to improve medication safety, efficacy, use, and adherence. In addition, Dr. Jain is leading several efforts for new ventures and business development. Previously, he was senior advisor to the administrator of the Centers for Medicare & Medicaid Services (CMS). At CMS, he helped lead the launch of the Center for Medicare & Medicaid Innovation, briefly serving as its acting deputy director for policy and programs. At CMS, Dr. Jain advocated for speedier translation of health care delivery research into practice and an expanded use of clinical registries. Dr. Jain also served as special assistant to the National Coordinator for Health Information Technology at the Office of the National Coordinator for Health Information Technology (ONC). At ONC, Dr. Jain worked with David Blumenthal to implement the Health Information Technology for Economic and Clinical Health provisions of the American Recovery and Reinvestment Act that provide incentives for physicians and hospitals to become meaningful users of health information technology. He

received his undergraduate degree magna cum laude in government from Harvard College; his M.D. from Harvard Medical School; and his M.B.A. from Harvard Business School. He was a recipient of the Paul and Daisy Soros Fellowship and the Dean's Award at Harvard Business School. He completed his residency in internal medicine at Brigham and Women's Hospital, where he was honored with the Resident Mentor Award. Dr. Jain is a founder of several nonprofit health care ventures, including the Homeless Health Clinic at the Harvard Square Homeless Shelter; the Harvard Bone Marrow Initiative; and ImproveHealthCare.org. He worked with DaVita-Bridge of Life to bring charity dialysis care to rural Rajasthan, India, and Medical Missions for Children to bring cleft lip and palate surgery to that region. He maintained a faculty appointment at Harvard Business School's Institute for Strategy and Competitiveness and worked with Strategy Professor Michael Porter on a new case literature on health care delivery innovation. He presently serves as an honorary senior institute associate at Harvard Business School. Dr. Jain has worked at WellPoint, McKinsey & Co., and the Institute for Healthcare Improvement. He has also served as an expert consultant to the World Health Organization. He has authored more than 50 publications on health care delivery innovation and health care reform in journals such as the *New England Journal of Medicine*, *JAMA*, and *Health Affairs*. His work has been cited in the *New York Times*, CNN, the *Wall Street Journal*, and other media outlets. He serves on the editorial boards of the *American Journal of Managed Care*, *Harvard Medicine*, and *Wing of Zock*. The book he coedited with Susan Pories and Gordon Harper, *The Soul of a Doctor*, has been translated into Chinese.

Beth Kozel, M.D., Ph.D., is an instructor of pediatrics in the Division of Genetics and Genomic Medicine at St. Louis Children's Hospital and Washington University School of Medicine. Clinically, she cares for children with genetic conditions that include chromosomal changes, inherited disorders, and multiple medical problems of unknown etiology. In her research, she has studied the biology of elastic fiber formation and works with patients with elastic fiber diseases such as Williams-Beuren syndrome, supravalvular aortic stenosis, and cutis laxa. Her current research is aimed at identifying modifiers of vascular disease severity among affected individuals. Her work is supported by a K08 award from the National Institutes of Health and she is a scholar of the Children's Discovery Institute at Washington University. She is a member of the American Society of Matrix Biology and the American Society of Hu-

man Genetics. She is a registered researcher with the Williams Syndrome Patient and Clinical Research Registry.

Harlan Krumholz, M.D., is the Harold H. Hines, Jr., Professor of Medicine and director of the Robert Wood Johnson Clinical Scholars Program at Yale University School of Medicine, and director of the Yale-New Haven Hospital Center for Outcomes Research and Evaluation. His research focuses on determining optimal clinical strategies and identifying opportunities for improving the prevention, treatment, and outcome of cardiovascular disease. Using applied clinical research methods, his work seeks to provide critical information to improve the quality of health care, monitor changes over time, guide decisions about the allocation of scarce resources, and inform decisions made by patients and their clinicians. Studies from his research group have directly led to improvements in the use of guideline-based medications, the timeliness of care for acute myocardial infarction, the information available to patients who are making decisions about clinical options, the public reporting of outcomes measures, and the current national focus on reducing the risk of readmission. He has also advanced the application of quality improvement to the elimination of disparities. His work influenced several provisions in the health reform bill. Dr. Krumholz serves on the board of trustees of the American College of Cardiology, the board of directors of the American Board of Internal Medicine, and the board of governors of the Patient-Centered Outcomes Research Institute. He is an elected member of the Association of American Physicians, the American Society for Clinical Investigation, and the Institute of Medicine. He is a Distinguished Scientist of the American Heart Association and received its Distinguished Service Award. He received a B.S. from Yale, an M.D. from Harvard Medical School, and a master's in health policy and management from the Harvard University School of Public Health.

Richard E. Kuntz, M.D., M.Sc., is senior vice president and chief scientific, clinical, and regulatory officer of Medtronic, Inc. In this role, which he assumed in 2009, Dr. Kuntz oversees the company's global regulatory affairs, health policy and reimbursement, clinical research, ventures and new therapies, and innovation functions. Dr. Kuntz joined Medtronic in 2005 as senior vice president and president of medtronic neuromodulation, which encompasses the company's products and therapies used in the treatment of chronic pain, movement disorders, spasticity, overactive bladder and urinary retention, benign prostatic hyperplasia,

and gastroparesis. In this role, he was responsible for the research, development, operations, and product sales and marketing for each of these therapeutic areas worldwide. Dr. Kuntz brings to Medtronic a broad background and expertise in many areas of health care. Prior to Medtronic he was the founder and chief scientific officer of the Harvard Clinical Research Institute, a university-based contract research organization that coordinates the National Institutes of Health and industry clinical trials with the Food and Drug Administration. He has directed more than 100 multicenter clinical trials and authored more than 200 original publications. His major interests are traditional and alternative clinical trial design and biostatistics. He also served as associate professor of medicine at Harvard Medical School, chief of the Division of Clinical Biometrics, and an interventional cardiologist in the division of cardiovascular diseases at the Brigham and Women's Hospital. Dr. Kuntz graduated from Miami University, and received his M.D. from Case Western Reserve University School of Medicine. He completed his residency in internal medicine at the University of Texas Southwestern Medical School, and then completed fellowships in cardiovascular diseases and interventional cardiology at the Beth Israel Hospital and Harvard Medical School. Dr. Kuntz received his M.S. in biostatistics from the Harvard School of Public Health.

Rebecca D. Kush, Ph.D., is founder, president, and CEO of the Clinical Data Interchange Standards Consortium (CDISC). CDISC is a nonprofit standards-developing organization with a mission to develop and support global, platform-independent standards that enable information system interoperability to improve medical research and related areas of health care and a vision of "informing patient care and safety through higher quality medical research." She has more than 25 years of experience in the area of clinical research, including positions with the National Institutes of Health, academia, a global contract research organization, and biopharmaceutical companies in the United States and Japan. She earned her doctorate in physiology and pharmacology from the University of California, San Diego, School of Medicine. She is lead author of the book *eClinical Trials: Planning and Implementation* and has authored numerous publications for journals including *New England Journal of Medicine* and *Science Translational Medicine*. She has developed a Prescription Education Program for elementary and middle schools and was named in *PharmaVoice* in 2008 as one of the 100 most inspiring individuals in the life sciences industry. Dr. Kush has served on the board of

directors for the U.S. Health Information Technology Standards Panel, Drug Information Association, and currently HL7. She was a member of the advisory committee for the World Health Organization International Clinical Trials Registry Platform. Dr. Kush served on the appointed Planning Committee for the Department of Health and Human Services Office of the National Coordinator for Health Information Technology sponsored Workshop Series on the "Digital Infrastructure for the Learning Health System" for the Institute of Medicine. She is a member of the National Cancer Advisory Board IT Workgroup and was invited to represent research as an appointed member of the U.S. Health Information Technology Standards Committee. Dr. Kush has developed a course, "A Global Approach to Accelerating Medical Research," and has been a keynote speaker at numerous conferences in this arena in Australia, Brazil, China, Europe, Japan, Korea, and the United States.

Elizabeth Loder, M.D., M.P.H., is a senior research editor at *BMJ*. She is also chief of the Division of Headache and Pain in the Department of Neurology at Brigham and Women's Hospital, and an associate professor of neurology at Harvard Medical School. She is board certified in internal medicine and headache medicine. Dr. Loder is the current president of the American Headache Society and has held leadership positions in national and international pain organizations and served as a member of the recent Institute of Medicine Committee on Advancing Pain Care, Research, and Education. Dr. Loder has worked as a clinician and researcher in the headache field for many years, and has extensive experience in the design and conduct of clinical trials. As a result of these experiences, Dr. Loder is deeply interested in problems associated with clinical trial reporting and interpretation, especially the consequences of missing clinical trial information for guidelines and treatment decisions. She has devoted much of her recent career to the development, implementation and dissemination of standards for research reporting. She was responsible for a recent *BMJ* theme issue on unpublished clinical trial data. Research published in that issue of the journal has led several U.S. lawmakers to question whether the National Institutes of Health and Food and Drug Administration are doing enough to enforce research reporting requirements.

Deven McGraw, J.D., is the director of the Health Privacy Project at Center for Democracy and Technology (CDT). The Project focuses on developing and promoting workable privacy and security protections for electronic personal health information. Ms. McGraw is active in efforts

to advance the adoption and implementation of health information technology and electronic health information exchange to improve health care. She was one of three persons appointed by the Department of Health and Human Services (HHS) Secretary Kathleen Sebelius to serve on the Health Information Technology Policy Committee, a federal advisory committee established by the American Recovery and Reinvestment Act of 2009. She co-chairs the committee's Privacy and Security "Tiger Team" and serves as a member of its Meaningful Use, Information Exchange, and Strategic Plan Workgroups. She also served on two key workgroups of the American Health Information Community, the federal advisory body established by HHS to develop recommendations on how to facilitate use of health IT to improve health. Specifically, she co-chaired the Confidentiality, Privacy and Security Workgroup and was a member of the Personalized Health Care Workgroup. She also served on the Policy Steering Committee of the eHealth Initiative and now serves on its Leadership Council. She is also on the Steering Group of the Markle Foundation's Connecting for Health multistakeholder initiative. Ms. McGraw has a strong background in health care policy. Prior to joining CDT, Ms. McGraw was the chief operating officer of the National Partnership for Women & Families, providing strategic direction and oversight for all of the organization's core program areas, including the promotion of initiatives to improve health care quality. Ms. McGraw also was an associate in the public policy group at Patton Boggs, LLP, and in the health care group at Ropes & Gray. She also served as deputy legal counsel to the Governor of Massachusetts and taught in the Federal Legislation Clinic at the Georgetown University Law Center. Ms. McGraw graduated magna cum laude from the University of Maryland. She earned her J.D., magna cum laude, and her L.L.M. from Georgetown University Law Center and was executive editor of the *Georgetown Law Journal*. She also has an M.P.H. from Johns Hopkins School of Hygiene and Public Health.

Meredith Nahm, Ph.D., provides oversight and coordination of clinical research informatics projects undertaken by the Core, including Duke Translational Medicine Institute's provision of infrastructure for clinical research data collection and management. She also oversees informatics for the MURDOCK community registry and the Duke Clinical Research Unit, as well as involvement in national efforts to develop and implement data standards. Additionally, Dr. Nahm supports efforts to advance the application of informatics capabilities to enhance medical practice.

Through her work and associated collaborations, Dr. Nahm helps to shape Duke's Clinical Research Informatics strategy and direction. Prior to her appointment as the associate director for clinical research informatics, Dr. Nahm served as the director of clinical data integration at the Duke Clinical Research Institute in 2001 and has been there since 1998. She has more than 15 years of experience in clinical research informatics and quality control. Dr. Nahm has held numerous industry leadership roles, including serving as a member of the Board of Trustees of the Society for Clinical Data Management (SCDM), chair of the SCDM Good Clinical Data Management Practices (GCDMP) Committee, chair of the Clinical Data Interchange Standards Consortium Industry Advisory Board, and the data management content expert for the National Institutes of Health (NIH)-sponsored National Leadership Forum for Clinical Research Networks. Dr. Nahm authored the GCDMP sections on Measuring and Assuring Data Quality and has played a leadership role in data standards development and implementation efforts. She is currently working with the data standards development efforts in cardiology and tuberculosis on two NIH Roadmap projects, and the Data and Statistical Center for the National Institute on Drug Abuse Clinical Trials Network. She is a frequent presenter at industry meetings, and publishes in the clinical trial operational literature. Her research interests include quantitative evaluation of common clinical data management practices and data quality. She received her master's in nuclear engineering from North Carolina State University, and her Ph.D. at the University of Texas at Houston School of Health Information Sciences.

Jeffrey S. Nye, M.D., Ph.D., is vice president, neuroscience innovation and partnership strategy, at Janssen Research & Development LLC, a Johnson & Johnson (J&J) pharmaceutical company. He is responsible for the strategy and implementation of external R&D in the neurosciences, aiming to expand the pipeline with novel business relationships, venture, academic and public-private partnerships, and risk-shared outsourcing. Previously, Dr. Nye was chief medical officer and co-head of the East Coast Research and Early Development unit at J&J pharma. Prior roles include vice president of experimental medicine, vice president of compound development for Galantamine (Razadyne/Reminyl), a therapy for Alzheimer's disease, and for Topamax, an epilepsy and migraine medicine. He previously served as a director of CNS Discovery Genomics and Biotechnology at Pharmacia. Prior to joining the pharmaceutical industry, Dr. Nye was a tenured associate professor of

molecular pharmacology, biological chemistry, and pediatrics (neurology) at Northwestern University and Children's Memorial Hospital. He received his bachelor's degree in biochemistry and master's degree in pharmacology from Harvard, and his M.D. and Ph.D. (with Solomon Snyder) from the Johns Hopkins University School of Medicine. He served as a pediatrics resident, postdoctoral fellow (with Richard Axel), and assistant professor of pediatrics in neurology at the College of Physicians and Surgeons of Columbia University. Dr. Nye has published more than 60 peer-reviewed papers on topics spanning basic and clinical research, and has led pharma programs in discovery through clinical development. He has served on numerous study sections and advisory panels, and currently serves as a councilor of the British Association of Psychopharmacology. He is the co-chair of the New York Academy Alzheimer's Leadership council.

Sally Okun, R.N., M.M.H.S., is a member of the Research, Data and Analytics Team at PatientsLikeMe, Inc. As the head of Health Data Integrity and Patient Safety, she is responsible for the site's medical ontology and the integrity of patient-reported health data. She also developed and oversees the PatientsLikeMe Drug Safety and Pharmacovigilance Platform. Prior to joining PatientsLikeMe, Ms. Okun practiced as a palliative care specialist. In addition to working with patients and families, she was an independent consultant supporting multiyear clinical, research, and education projects focused on palliative and end-of-life care for numerous clients, including Brown University, Harvard Medical School, the Massachusetts Department of Mental Health, and the Robert Wood Johnson Foundation. Ms. Okun is a member of the Institute of Medicine Clinical Effectiveness Research Innovation Collaborative, the Evidence Communication Innovation Collaborative, and the Best Practices Innovation Collaborative. She is a contributing author on two Institute of Medicine discussion papers: "Principles and Values for Team-Based Health Care" and "Communicating Evidence in Health Care: Engaging Patients for Improved Health Care Decisions." Ms. Okun received her master's degree from The Heller School for Social Policy & Management at Brandeis University; completed study of Palliative Care and Ethics at Memorial Sloan-Kettering Cancer Center; and was a Naional Library of Medicine Fellow in biomedical informatics.

Eric D. Perakslis, Ph.D., is chief information officer and chief scientist (informatics) at the Food and Drug Administration (FDA). In this role, Dr. Perakslis is responsible for modernizing and enhancing the IT capabilities as well as the in silico scientific capabilities at FDA. Prior to FDA, he was senior vice president of R&D information technology at Johnson & Johnson (J&J) Pharmaceuticals R&D and a member of the Corporate Office of Science and Technology. During his 13 years at J&J, Dr. Perakslis also held the posts of vice president R&D informatics, vice president and chief information officer, director of research information technology as well as assistant director and director of drug discovery research prior to his current role. Before joining J&J, he was the group leader of Scientific Computing at ArQule Inc. and he began his professional career with the Army Corps of Engineers. Dr. Perakslis has a Ph.D. in chemical and biochemical engineering from Drexel University and also holds B.S. and M.S. degrees in chemical engineering. His current research interests are enterprise knowledge management, patient stratification, health care IT and translational informatics with the specific focus on precompetitive data sharing, and open-source systems globalization. Dr. Perakslis is a late-stage kidney cancer survivor and an avid patient advocate. He has served as the chair of the Survivor Advisory Board at the Cancer Institute of New Jersey and as the chief information officer of the King Hussein Institute for Biotechnology and Cancer in Jordan. He has also worked extensively with Lance Armstrong's LiveStrong Foundation, the Kidney Cancer Association, the Scientist ↔ Survivor Program of the American Association for Cancer Research, OneMind4Research and several other prominent disease-based organizations to further their agendas.

Richard Platt, M.D., M.Sc., is a professor and chair of the Department of Population Medicine, and executive director of the Harvard Pilgrim Health Care Institute. He is an internist trained in infectious diseases and epidemiology. His research focuses on developing multi-institution automated record linkage systems for use in pharmacoepidemiology, and for population-based surveillance, reporting, and control of both hospital- and community-acquired infections. He is principal investigator of the Food and Drug Administration's (FDA's) Mini-Sentinel program, of a Centers for Disease Control and Prevention (CDC) Center of Excellence in Public Health Informatics, of the Agency for Healthcare Research and Quality HMO Research Network DEcIDE center, and of a CDC Prevention Epicenter. He is a member of the Institute of Medicine Roundtable

on Value & Science-Driven Health Care, and of the Association of American Medical Colleges' Advisory Panel on Research. Dr. Platt formerly chaired the FDA Drug Safety and Risk Management Advisory Committee, a National Institutes of Health study section (Epidemiology and Disease Control 2), and the CDC Office of Health Care Partnerships steering committee, and co-chaired the Board of Scientific Counselors of the CDC National Center for Infectious Diseases.

William Z. Potter, M.D., Ph.D., earned his B.A., M.S., M.D., and Ph.D. at Indiana University, after which he held positions of increasing responsibility and seniority in translational neuroscience over the next 25 years at the National Institutes of Health (NIH). While there, Dr. Potter was widely published and appointed to many societies, committees, and boards, which enabled him to develop a wide reputation as an expert in psychopharmacological sciences and champion the development of novel treatments for central nervous system (CNS) disorders. Dr. Potter left the NIH in 1996 to accept a position as executive director and Research Fellow at Lilly Research Labs, specializing in the Neuroscience Therapeutic Area. In 2004 he joined Merck Research Labs (MRL) as vice president of clinical neuroscience, and in 2006, he took on the newly created position of vice president of Translational Neuroscience, from which he retired in 2012. His experience at Lilly and MRL in identifying, expanding, and developing methods of evaluating CNS effects of compounds in human brains cover state-of-the-art approaches across multiple modalities. These include brain imaging and cerebrospinal fluid proteomics (plus metabolomics) as well as development of more sensitive clinical, psychophysiological, and performance measures allowing a range of novel targets to be tested in a manner that actually addresses the underlying hypotheses. He has become a widely recognized champion for the position that more disciplined hypothesis testing of targets in humans is the best near-term approach to moving CNS drug development forward for important neurologic and psychiatric illnesses.

Jonathan Rabinowitz, Ph.D., is the Elie Wiesel Professor and a former department chair at Bar Ilan University. He also holds a visiting professorship in the Department of Psychiatry at Mount Sinai School of Medicine in New York. His current research focuses on understanding and minimizing placebo response in antipsychotic and antidepressant trials and developing statistical and methodological ways to improve efficiency of the central nervous system drug trials. His research has been sup-

ported by the National Institutes of Health, Federal Ministry of Education and Research (BMBF), Innovative Medicines Initiative, European Commission of the European Union, Brain & Behavior Research Foundation, formerly NARSAD, and major pharmaceutical companies. He is a member of the editorial board of *European Neuropsychopharmacology* and a member of the International Biometric Society, International Society for Schizophrenia Research, and European College of Neuropsychopharmacology. He serves on the Food and Drug Administration Clinical Expert Review Committee data standards development initiative, which will standardize clinical data elements for regulatory submission. He is a member of the Drug Information Association scientific working group on missing data in clinical trials. He has published more than 150 scientific papers.

Frank W. Rockhold, Ph.D., is senior vice president, global clinical safety and pharmacovigilance at GlaxoSmithKline (GSK) Pharmaceuticals Research and Development. This includes case management, signal detection, safety evaluation, and risk management for all development and marketed products. He is also co-chair of the GSK Global Safety Board. In his 20 years at GSK, he has also held management positions within the Statistics & Epidemiology Department and Clinical Operations, both in R&D and in the U.S. pharmaceutical business, and most recently as interim head of the Cardiovascular and Metabolic Medicines Development Center. Dr. Rockhold previously held positions of research statistician, Lilly Research Laboratories, and executive director of biostatistics, data management, and health economics, Merck Research Laboratories. He has a B.A. in statistics from the University of Connecticut, an Sc.M. in biostatistics from Johns Hopkins University, and a Ph.D. in biostatistics from the Medical College of Virginia. He has held academic appointments at Butler University; Indiana University; adjunct professor of health evaluation sciences, Penn State University; and adjunct scholar in the Department of Epidemiology and Biostatistics at the University of Pennsylvania. Dr. Rockhold is currently affiliate professor of biostatistics at the Virginia Commonwealth University Medical Center. He is a past chair of the board of directors of the Clinical Data Interchange Standards Consortium and a member of the National Library of Medicine Advisory Group for ClinicalTrials.Gov. He is past president, Society for Clinical Trials; past chair, Pharmaceutical Research and Manufacturers Association (PhRMA) Biostatistics Steering Committee, and a member of the ICH E-9 and E-10 Expert Working Groups. He also served as associate

editor for *Controlled Clinical Trials*. He is a fellow of the American Statistical Association and a fellow of the Society for Clinical Trials. He is also a recipient of the PhRMA Career Achievement award. He has more than 150 publications and abstracts.

Laura Lyman Rodriguez, Ph.D., is director of the Office of Policy, Communications, and Education at the National Institutes of Health (NIH) National Human Genome Research Institute (NHGRI). She works to develop and implement policy for research initiatives at the NHGRI, design communication and outreach strategies to engage the public in genomic science, and prepare health care professionals for the integration of genomic medicine into clinical care. Dr. Rodriguez is particularly interested in the policy and ethics questions related to the inclusion of human research participants in genomics and genetics research and sharing human genomic data through broadly used research resources (e.g., databases). Among other activities, Dr. Rodriguez has provided leadership for many policy development activities pertaining to genomic data sharing and the creation of the database for Genotypes and Phenotypes (dbGaP) at the NIH. Dr. Rodriguez received her B.S. with honors in biology from Washington and Lee University and earned a doctorate in cell biology from Baylor College of Medicine.

Michael Rosenblatt, M.D., is executive vice president and chief medical officer of Merck & Co., Inc. He is the first person to serve in this role for Merck. Previously he served as dean of Tufts University School of Medicine. Prior to that, he held the appointment of George R. Minot Professor of Medicine at Harvard Medical School and chief of the Division of Bone and Mineral Metabolism Research at Beth Israel Deaconess Medical Center (BIDMC). He served as the president of BIDMC from 1999 to 2001. Previously, he was the Harvard faculty dean and senior vice president for academic programs at CareGroup and BIDMC, and a founder of the Carl J. Shapiro Institute for Education and Research at Harvard Medical School and BIDMC, a joint venture whose mission is to manage the academic enterprise and promote academic innovation. Earlier, he served as director of the Harvard–Massachusetts Institute of Technology (MIT) Division of Health Sciences and Technology, during which time he led a medical education organization for M.D., Ph.D., and M.D.–Ph.D. training jointly sponsored by Harvard and MIT. Previously, he was senior vice president for research at Merck Sharp & Dohme Research Laboratories, where he co-led the worldwide development team for alendronate

(FOSAMAX), Merck's bisphosphonate for osteoporosis and bone disorders. In addition, he directed drug discovery efforts in molecular biology, bone biology and calcium metabolism, virology, cancer research, lipid metabolism, and cardiovascular research in the United States, Japan, and Italy. In leading most of Merck's international research efforts, he established two major basic research institutes, one in Tsukuba, Japan, and one near Rome, Italy. He also headed Merck Research's worldwide University and Industry Relations Department. He is the recipient of the Fuller Albright Award for his work on parathyroid hormone and the Vincent du Vigneaud Award in peptide chemistry and biology, and the Chairman's Award from Merck. His research is in the field of hormonal regulation of calcium metabolism, osteoporosis, and cancer metastasis to bone. His major research projects are in the design of peptide hormone antagonists for parathyroid hormone and the tumor-secreted parathyroid hormone-like protein; isolation/characterization of receptors and mapping hormone–receptor interactions; elucidation of the mechanisms by which breast cancer "homes" to bone; and osteoporosis and bone biology. He has been an active participant in the biotechnology industry, serving on the board of directors and scientific advisory boards of several biotech companies. He was a scientific founder of ProScript, the company that discovered bortezomib (Velcade), now Millennium Pharmaceutical's drug for multiple myeloma and other cancers. He was a member of the Board of Scientific Counselors of the National Institute of Diabetes and Digestive and Kidney Diseases at the National Institutes of Health. He has been elected to the American Society of Clinical Investigation and the Association of American Physicians; to fellowship in the American Association for the Advancement of Science and the American College of Physicians; and as president of the American Society of Bone and Mineral Research. He has testified before the U.S. Senate on U.S. biomedical research priorities, and served as a consultant to the U.S. President's Council of Advisors on Science and Technology. He previously served as Chief of the Endocrine Unit, Massachusetts General Hospital. He received his undergraduate degree summa cum laude from Columbia University and his M.D. magna cum laude from Harvard. His internship, residency, and endocrinology training were all at Massachusetts General Hospital.

Vicki L. Seyfert-Margolis, Ph.D., is senior advisor for science innovation and policy in the Food and Drug Administration Commissioner's Office. She focuses on initiatives in innovation, regulatory science, per-

sonalized medicine, scientific computing, and health informatics. Previously, she served as chief scientific officer at Immune Tolerance Network (ITN), a nonprofit consortium of researchers seeking new treatments for diseases of the immune system. At ITN, she oversaw the development of more than 20 centralized laboratory facilities, and the design and execution of biomarker discovery studies for more than 25 Phase II clinical trials. As part of the biomarker efforts, she established construction of a primer library of 1,000 genes that may be involved in establishing and maintaining immunologic tolerance and codiscovered genes that may mark kidney transplant tolerance. Dr. Seyfert-Margolis was also an adjunct associate professor with the Department of Medicine at the University of California, San Francisco. She also served as director of the Office of Innovative Scientific Research Technologies at the National Institute of Allergy and Infectious Diseases at the National Institutes of Health, where she worked to integrate emerging technologies into existing immunology and infectious disease programs. Dr. Seyfert-Margolis completed her Ph.D. in immunology at the University of Pennsylvania's School of Medicine.

Jay M. Tenenbaum, Ph.D., is the founder and chair of Cancer Commons, a nonprofit, open science community that compiles and continually refines information about cancer subtypes and treatments, based on the literature and actual patient outcomes. Dr. Tenenbaum's background brings a unique perspective of a world-renowned Internet commerce pioneer and visionary. He was founder and CEO of Enterprise Integration Technologies, the first company to conduct a commercial Internet transaction, secure Web transaction, and Internet auction. In 1994, he founded CommerceNet to accelerate business use of the Internet. In 1997, he cofounded Veo Systems, the company that pioneered the use of XML for automating business-to-business transactions. Dr. Tenenbaum joined Commerce One in 1999, when it acquired Veo Systems. As chief scientist, he was instrumental in shaping the company's business and technology strategies for the Global Trading Web. After Commerce One, Dr. Tenenbaum was an officer and a director of Webify Solutions, which was sold to IBM in 2006, and Medstory, which was sold to Microsoft in 2007. Dr. Tenenbaum was also the Founder and Chair of CollabRx, a provider of Web-based applications and services that help cancer patients and their physicians select optimal treatments and trials; it was acquired by Tegal in 2012. Earlier in his career, Dr. Tenenbaum was a prominent artificial intelligence (AI) researcher and led AI research groups at SRI

International and Schlumberger Ltd. Dr. Tenenbaum is a Fellow and former board member of the American Association for Artificial Intelligence, and a former consulting professor of computer science at Stanford. He currently serves as a director of Efficient Finance, PatientsLikeMe, and the Public Library of Science, and is a consulting professor of information technology at Carnegie Mellon's new West Coast campus. Dr. Tenenbaum holds a B.S. and an M.S. in electrical engineering from the Massachusetts Institute of Technology, and a Ph.D. from Stanford University.

Andrew Vickers, Ph.D., is attending research methodologist in the Department of Epidemiology and Biostatistics at Memorial Sloan-Kettering Cancer Center (MSKCC). He obtained a first in natural science from the University of Cambridge and has a doctorate in clinical medicine from the University of Oxford. Dr. Vickers' clinical research falls into three broad areas: randomized trials, surgical outcomes research, and molecular marker studies. A particular focus of his work is the detection and initial treatment of prostate cancer. Dr. Vickers has analyzed the "learning curve" for radical prostatectomy, is working on a series of studies demonstrating that a single measure of prostate-specific antigen taken in middle age can predict prostate cancer up to 25 years later, and has developed a statistical model to predict the result. His work on randomized trials focuses on methods for integrating randomized trials into routine surgical practice so as to compare different approaches to surgery. As part of this work, he has pioneered the use of Web interfaces for obtaining quality-of-life data from patients recovering from radical prostatectomy. Dr. Vickers' methodological research centers primarily on novel methods for assessing the clinical value of predictive tools. In particular, he has developed decision-analytic tools that can be directly applied to a dataset, without the need for data gathering on patient preferences or utilities. Dr. Vickers has a strong interest in teaching statistics. He is course leader for the MSKCC biostatistics course, teaches on the undergraduate curriculum at Weill Medical College of Cornell University, and writes the statistics column for Medscape. He is author of the introductory statistics textbook *What Is a P-Value Anyway? 34 Stories to Help You Actually Understand Statistics.*

John A. Wagner, M.D., Ph.D., received his M.D. from Stanford University School of Medicine and his Ph.D. from the Johns Hopkins University School of Medicine. Postgraduate training included an internship and resi-

dency in internal medicine as well as molecular and clinical pharmacology postdoctoral fellowships. Currently, Dr. Wagner is vice president and head, Early Development Pipeline and Projects, and head, Global Project Management, at Merck & Co., Inc. Previously, he was vice president and head, clinical pharmacology, at Merck. Dr. Wagner is also an adjunct assistant professor, Division of Clinical Pharmacology, Department of Medicine, Jefferson Medical College, and visiting clinical scientist, Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology, Center for Experimental Pharmacology and Therapeutics. He is the past chair of the PhRMA Clinical Pharmacology Technical Group, past chair of the adiponectin workgroup for the Biomarkers Consortium, past committee member of the Institute of Medicine (IOM) Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease, current member of the Board of Directors and current chair of the 2012 Scientific Program Committee for the American Society for Clinical Pharmacology and Therapeutics, and current full member of the IOM National Cancer Policy Forum. His more than 150 peer-reviewed publications include work in the areas of biomarkers and surrogate endpoints, experimental medicine, modeling and simulation, pharmacokinetics, pharmacodynamics, precompetitive collaboration, and drug–drug interactions across a variety of therapeutic areas.

John Wilbanks, B.A., is a senior fellow in entrepreneurship with the Ewing Marion Kauffman Foundation. He works on open-content, open-data, and open-innovation systems. Mr. Wilbanks also serves as a research fellow at Lybba. He has worked at Harvard Law School, the Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory, the World Wide Web Consortium, the U.S. House of Representatives, and Creative Commons, as well as starting a bioinformatics company. He sits on the board of directors for Sage Bio networks and iCommons, and the advisory board for Boundless Learning. Mr. Wilbanks holds a degree in philosophy from Tulane University and also studied modern letters at the University of Paris (La Sorbonne).

Janet Woodcock, M.D., is the director, Center for Drug Evaluation and Research (CDER), Food and Drug Administration (FDA). Dr. Woodcock held various leadership positions within the Office of the Commissioner, FDA, including deputy commissioner and chief medical officer, deputy commissioner for operations, and chief operating officer and director,

Critical Path Programs. Previously, Dr. Woodcock served as director, CDER, from 1994 to 2005. She also held other positions at FDA, including director, Office of Therapeutics Research and Review, and acting deputy director, Center for Biologics Evaluation and Research. A prominent FDA scientist and executive, Dr. Woodcock has received numerous awards, including a Presidential Rank Meritorious Executive Award, the American Medical Association's Nathan Davis Award, and Special Citations from FDA Commissioners. Dr. Woodcock received her M.D. from Northwestern Medical School, and completed further training and held teaching appointments at the Pennsylvania State University and the University of California, San Francisco.

Deborah A. Zarin, M.D., is director of ClinicalTrials.gov and assistant director of clinical research projects at the Lister Hill National Center for Biomedical Communications in the National Library of Medicine. In this capacity, Dr. Zarin oversees the development and operation of an international registry of clinical trials. Previous positions include director of the Agency for Healthcare Research and Quality Technology Assessment Program and director of the American Psychiatric Association Practice Guidelines program. In these positions, Dr. Zarin conducted systematic reviews and related analyses in order to develop clinical and policy recommendations. Her academic interests are in the area of evidence-based clinical and policy decision making, and she is the author of more than 70 peer-reviewed articles. Dr. Zarin graduated from Stanford University and received her M.D. from Harvard Medical School. She completed a clinical decision-making fellowship, and is board certified in general psychiatry as well as in child and adolescent psychiatry.