



Discussion Framework for Clinical Trial Data Sharing: Guiding Principles, Elements, and Activities

ISBN
978-0-309-29779-0

72 pages
6 x 9
2014

Committee on Strategies for Responsible Sharing of Clinical Trial Data;
Board on Health Sciences Policy; Institute of Medicine

 [More information](#)

 [Find similar titles](#)

 [Share this PDF](#)



Visit the National Academies Press online and register for...

- ✓ Instant access to free PDF downloads of titles from the
 - NATIONAL ACADEMY OF SCIENCES
 - NATIONAL ACADEMY OF ENGINEERING
 - INSTITUTE OF MEDICINE
 - NATIONAL RESEARCH COUNCIL
- ✓ 10% off print titles
- ✓ Custom notification of new releases in your field of interest
- ✓ Special offers and discounts

Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences. Request reprint permission for this book

DISCUSSION FRAMEWORK FOR CLINICAL TRIAL DATA SHARING

GUIDING PRINCIPLES, ELEMENTS, AND ACTIVITIES

Committee on Strategies for Responsible
Sharing of Clinical Trial Data

Board on Health Sciences Policy

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. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by contracts between the National Academy of Sciences and the U.S. National Institutes of Health (HHSN263201200074I), U.S. Food and Drug Administration, AbbVie Inc., Amgen Inc., AstraZeneca Pharmaceuticals, Bayer, Biogen Idec, Bristol-Myers Squibb, Burroughs Wellcome Fund, Doris Duke Charitable Foundation, Eli Lilly and Company, EMD Serono, Genentech, GlaxoSmithKline, Johnson & Johnson, Medical Research Council (UK), Merck & Co., Inc., Novartis Pharmaceuticals Corporation, Novo Nordisk, Pfizer Inc., Sanofi-Aventis, Takeda, and Wellcome Trust. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

Additional copies of this report 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**.

Copyright 2014 by the National Academy of Sciences. All rights reserved.

Printed in the United States of America

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

Suggested citation: IOM (Institute of Medicine). 2014. *Discussion framework for clinical trial data sharing: Guiding principles, elements, and activities*. 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. C. D. Mote, Jr., 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. C. D. Mote, Jr., are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

COMMITTEE ON STRATEGIES FOR RESPONSIBLE SHARING OF CLINICAL TRIAL DATA

BERNARD LO (*Chair*), President, The Greenwall Foundation, New York City, New York

TIMOTHY COETZEE, Chief Advocacy, Services, and Research Officer, National Multiple Sclerosis Society

DAVE DEMETS, Professor and Chair, Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison

JEFFREY DRAZEN, Editor-in-Chief, *New England Journal of Medicine*

STEVEN GOODMAN, Professor, Medicine & Health Research & Policy, Stanford University School of Medicine

PATRICIA KING, Carmack Waterhouse Professor of Law, Medicine, Ethics, and Public Policy, Georgetown University Law Center

TRUDIE LANG, Principal Investigator, Global Health Network, Nuffield Department of Medicine, University of Oxford

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

ELIZABETH NABEL, President, Brigham and Women's Hospital

ARTI RAI, Elvin R. Latty Professor of Law, Duke University School of Law

IDA SIM, Professor of Medicine and Co-Director of Biomedical Informatics of the Clinical and Translational Science Institute, University of California at San Francisco

SHARON TERRY, President and CEO, Genetic Alliance

JOANNE WALDSTREICHER, Chief Medical Officer, Johnson & Johnson

IOM Staff

ANNE B. CLAIBORNE, Senior Program Officer

LEIGHANNE OLSEN, Interim Study Director (until November 2013)

REBECCA N. LENZI, Study Director (from November 2013)

MICHELLE MANCHER, Associate Program Officer

RACHEL KIRKLAND, Senior Program Assistant (until October 2013)

BARRET J. ZIMMERMANN, Senior Program Assistant (from October 2013)

ANDREW M. POPE, Director, Board on Health Sciences Policy

Consultant

THERESA WIZEMANN, Wizemann Scientific Communications, LLC

Reviewers

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

Lawrence J. Appel, Johns Hopkins Medical Institutions

Marc Boutin, National Health Council

Kay Dickersin, Johns Hopkins Bloomberg School of Public Health

David Harrington, Harvard School of Public Health

Peter S. Kim, Stanford University School of Medicine

Bartha Knoppers, McGill University

David Korn, Harvard University

Harlan M. Krumholz, Yale University School of Medicine

Christine Laine, *Annals of Internal Medicine*

Sir Michael Rawlins, Royal Society of Medicine

Jerome P. Reiter, Duke University

Donald M. Steinwachs, Johns Hopkins University

Alan M. Zaslavsky, Harvard Medical School

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by **Enriqueta C. Bond**, QE Philanthropic Advisors, and **Stephen E. Fienberg**, Carnegie Mellon University. Appointed by the Institute of Medicine and the National Research Council, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

Preface

Sharing data generated through the conduct of clinical trials offers the promise of placing evidence about the safety and efficacy of therapies and clinical interventions on a firmer basis and enhancing the benefits of clinical trials. Ultimately, such data sharing—if carried out appropriately—could lead to improved clinical care and greater public trust in clinical research and health care. However, recommending strategies for how clinical trial data could be shared responsibly is a difficult process—many stakeholders have valid interests that sometimes are not aligned; different models and approaches have been proposed or tried; and proposing an approach that works for the wide range of clinical trials may be daunting. Undergirding any discussion of possibilities and needs for clinical trial data sharing is a fundamental philosophical question: To whom ought the benefits of the data belong, and, accordingly, where lies the presumption as to whether data ought to be shared?

The Institute of Medicine (IOM) has assembled a committee with interdisciplinary expertise and a wide range of backgrounds to study how data from clinical trials might best be shared. The fulfillment of the committee's task requires careful gathering of evidence and listening to and understanding different viewpoints. We plan to learn from what has already been proposed and tried.

As the committee is charged in its statement of task, this document is designed as a framework for discussion. This framework is being released to stimulate reactions and comments from stakeholders and the public. In addition to seeking comments generally, we also have posed several specific questions for which feedback and suggestions would be particularly useful as we move our deliberations forward. Although we welcome comments at any time throughout the study process, comments

on the issues raised in this framework document would be most helpful if received by March 24, 2014. At this early stage in our work, we have reached no conclusions. As all IOM committees are called to do, we approach our charge with minds open to sound evidence and convincing arguments, and we look forward to hearing them.

This document would not have been possible without the dedicated and skilled work of IOM staff, including Anne B. Claiborne, LeighAnne Olsen, Rebecca N. Lenzi, Michelle Mancher, Rachel Kirkland, Barret J. Zimmermann, and Andrew M. Pope, whom the committee gratefully acknowledges and thanks.

Bernard Lo, *Chair*
Committee on Strategies for Responsible
Sharing of Clinical Trial Data

Contents

| | |
|--|----------|
| CONTEXT OF THE STUDY | 1 |
| Charge to the Committee and Scope of the Study, 3 | |
| A FRAMEWORK FOR DISCUSSION AND CALL FOR COMMENTS | 5 |
| Invitation for Public Comments, 7 | |
| GUIDING PRINCIPLES FOR RESPONSIBLE SHARING OF CLINICAL TRIAL DATA | 8 |
| Respect the Individual Participants Whose Data Are Shared, 9 | |
| Respect Through Research Participant Protections, 9 | |
| Respect Through Engagement, 10 | |
| Maximize Benefits to Participants in Clinical Trials and to Society, While Minimizing Harm, 10 | |
| Potential Benefits, 11 | |
| Potential Harms, 13 | |
| Increase Public Trust in Clinical Trials, 14 | |
| Trust in Clinical Trial Data, 14 | |
| Trust in the Data Sharing Process, 15 | |
| Carry Out Sharing of Clinical Trial Data in a Manner That Enhances Fairness, 15 | |
| Fairness in Exposure to Potential Risks of Data Sharing, 16 | |
| Fairness in the Distribution of the Potential Benefits of Research, 16 | |

| | |
|---|-----------|
| ADDITIONAL OBSERVATIONS RELATING TO PRINCIPLES FOR DATA SHARING | 17 |
| Operational Strategies Following from Guiding Principles, 17 | |
| Framing Rights and Responsibilities: To Whom Do the Benefits of Clinical Trial Data Belong?, 18 | |
| DATA SHARING ELEMENTS AND ACTIVITIES | 19 |
| What Types of Data Could Be Shared?, 20 | |
| Current Practices in Data Disclosure, 20 | |
| Data, 21 | |
| Reports Generated from Data, 23 | |
| Metadata and Additional Documentation, 25 | |
| Who Are the Providers of Shared Data?, 27 | |
| Who Are the Recipients of Shared Data?, 28 | |
| When Might Clinical Trial Data Be Shared?, 29 | |
| How Might Data Be Shared?, 29 | |
| Open Access, 33 | |
| Controlled Access, 33 | |
| Selected Set of Clinical Trial Data Sharing Activities, 34 | |
| REFERENCES | 37 |
| APPENDIXES | |
| A Statement of Task | 43 |
| B Committee Biographies | 45 |
| C Figure 1: Data Flow from Participant to Analyzed Data and Reporting | 57 |

Boxes and Figure

BOXES

- 1 Specific Topics for Public Feedback, 6
- 2 Provisional Guiding Principles for Responsible Sharing of Clinical Trial Data, 8
- 3 Types of Summary Reports, 23
- 4 Metadata and Additional Documentation to Support the Use of Shared Clinical Trial Data, 26
- 5 Set of Clinical Trial Data Sharing Activities, 3

FIGURE

- 1 Data flow from raw data to analyzable data, 21

Discussion Framework for Clinical Trial Data Sharing

CONTEXT OF STUDY

Clinical trials are crucial to determining the safety of medical interventions and their ability to achieve particular health outcomes. Clinical trials are required by regulatory authorities around the world before a new medical product can be brought to market, or before a new indication, formulation, or target population can be approved for an intervention already on the market (ICH, 1995). After a product's introduction, additional clinical trials are commonly conducted by industry, government, and academia to further define the relative safety and efficacy (or effectiveness) of the product. Clinical trials are also used to study interventions that do not involve regulated medical products, for example, surgical techniques, behavioral interventions, or studies designed to improve disease management practice (Califf, 2013).

Vast amounts of data are generated over the course of a clinical trial. These data are held by the sponsors conducting the clinical trial, and in some instances, by participants or their advocates (Drazen, 2002; Terry and Terry, 2011). Depending on the regulatory jurisdiction, data might or might not be shared or made available to the public for secondary uses. Shared data might include both summary data and individual patient data. In the United States, if a sponsor is seeking regulatory approval, data are shared in confidence with regulators.¹ Select study data might also be made available to individual researchers on a case by case basis upon request, or could be made publicly available, usually at the

¹However, in the European Union, the European Medicines Agency (EMA) has undertaken regulatory action to share anonymized clinical trial data with external requestors. The EMA's data sharing initiative is described on p. 32.

summary level, for example, through publication in a peer-reviewed journal or through publicly accessible clinical trial registration sites (e.g., ClinicalTrials.gov). Other data, however, remain largely unavailable to outside researchers and the public, including data found in analyzable data sets, clinical study reports (CSRs), and individual participant data (IPD) that, if accessible, could facilitate new analyses and a deeper understanding of a particular therapy or condition (Doshi et al., 2013; Goldacre, 2012; Gordon et al., 2013; Rawlins, 2012). Increased sharing of IPD, in particular, could facilitate activities such as independent reanalysis of trial results, addressing concerns about publication bias,² characterizing trial outcomes by subgroups, considering additional questions beyond the original trial hypotheses, carrying out meta-analysis for systematic reviews, and facilitating hypothesis generation and additional research to develop new therapies (Doshi et al., 2012; IOM, 2013; McGauran et al., 2010; Ross et al., 2012).

The data sharing movement has gained substantial momentum during the last decade, in both the clinical trial and larger scientific communities (Boulton et al., 2011; Royal Society, 2012). A cultural change has occurred in which the conversation around data sharing has moved from whether it should happen to how it can be carried out (IOM, 2013). To that end, a large number of people and organizations involved in clinical trials have endorsed principles promoting, in their view, responsible sharing of clinical trial data (Loder, 2013). Prominent examples include the joint statement from funders of health research data, the AllTrials campaign, and the Pharmaceutical Research and Manufacturers of America (PhRMA) and European Federation of Pharmaceutical Industries and Associations (EFPIA) principles (AllTrials, 2013; PhRMA and EFPIA, 2013; Wellcome Trust, 2011). Some organizations have gone beyond general statements and principles and begun to adopt data sharing policies as well. The U.S. National Institutes of Health (NIH), as well as international groups of researchers and nonprofit funders, has been involved in data sharing activities for years (e.g., BioLINCC, MalariaGen). More recently, European regulators and some pharmaceutical and device companies (e.g., GlaxoSmithKline [GSK], Medtronic) have begun to plan and implement their own data sharing policies (EMA, 2013; Nisen and Rockhold, 2013; YODA, 2013). In addition, medical journals have begun to require authors to include data sharing statements

²The tendency for positive results of trials of health care interventions to be reported or published, with corresponding underreporting/non-publication of negative or inconclusive results (Cochrane Collaboration, 2002).

and, in the case of the *British Medical Journal* (BMJ), to agree to make de-identified patient-level data available upon “reasonable request” as a condition for publication (Godlee and Groves, 2012, p. 1).

In October 2012, the Institute of Medicine (IOM) convened individuals with a broad range of expertise and perspectives to discuss sharing of clinical research data in a public workshop. As highlighted by workshop participants, data sharing can provide important potential benefits to industry, nonprofit and government funders of research, academic investigators, patient advocacy groups, and, ultimately, patients and the public, including, for example, speeding medical innovation by reducing redundancies, facilitating the identification and validation of new drug targets, identifying new indications for use, and improving the understanding of the safety and efficacy of therapies (IOM, 2013). In follow-up to the 2012 IOM workshop, the IOM was asked by a group of federal, industry, and U.S. and international foundation sponsors³ to conduct a consensus study to recommend guiding principles and a framework for the responsible sharing of clinical trial data.

Charge to the Committee and Scope of the Study

As described in the committee’s charge from the sponsors (see Appendix A), over a 17-month period of deliberations, the committee will release two documents:

1. This document, which is a framework for discussion (“framework”), to be released in January 2014 for public comment. The framework will summarize the committee’s initial thoughts on guiding principles that underpin responsible sharing of clinical trial data, define key elements of clinical trial data and data sharing, and describe a selected set of clinical trial data sharing activities. The charge to the committee excludes evidence-based findings and conclusions and recommendations from this document.
2. A final report with findings and recommendations related to the committee’s full charge.

³U.S. National Institutes of Health, U.S. Food and Drug Administration, AbbVie Inc., Amgen Inc., AstraZeneca Pharmaceuticals, Bayer, Biogen Idec, Bristol-Myers Squibb, Burroughs Wellcome Fund, Doris Duke Charitable Foundation, Eli Lilly and Company, EMD Serono, Genentech, GlaxoSmithKline, Johnson & Johnson, Medical Research Council (UK), Merck & Co., Inc., Novartis Pharmaceuticals Corporation, Novo Nordisk, Pfizer Inc., Sanofi-Aventis, Takeda, and Wellcome Trust.

To respond to the charge, the IOM convened a committee composed of persons with expertise in key scientific and research-related areas, including academia, industry, funding bodies, regulatory activities, scientific publications, clinicians, and patients. Individual committee member expertise spans academic clinical trial design, performance and dissemination; pharmaceutical product development; statistics, informatics, and data security; ethics of human subjects research; and law and regulatory requirements (including privacy, security, and intellectual property). Committee members also have insight into the global context of data sharing; the concerns of research participants, patients, and their families; and other relevant issues.

As charged in its statement of task, in the next phase of the study the committee will identify the key benefits, challenges, and risks of sharing, as well as key risks of not sharing. This analysis will take into consideration the full range of perspectives of research sponsors and investigators, study participants, regulatory agencies, patient groups, and the public. The committee is also charged, in the final report, with suggesting strategies and practical approaches for responsible data sharing. As part of its recommendations, the committee will offer guiding principles for and characteristics of the optimal infrastructure and governance for data sharing. In particular, the committee has been called to consider, among other issues, resource constraints, implementation, disincentives in the academic research model, changing norms, protection of human subjects and patient privacy, intellectual property and other legal issues, and preservation of scientific standards and data quality.

As outlined in the statement of task (Appendix A), many terms are defined for the purposes of this study. “Data sharing” is defined as the responsible entity (“data generator”) making the data available via open or restricted access, or exchanged among parties. For the purposes of this study, data generator may include industry sponsors, data repositories, and researchers conducting clinical trials. The committee has adopted a working definition of “data holder” to mean the entity or entities that have access to data, including regulatory agencies, journals to which manuscripts are submitted, and other repositories such as ClinicalTrials.gov in the United States. Data holders—for example, persons who collect source data, develop an analyzable data set, or carry out statistical analyses—may or may not have legal authority to provide third-party access to the data. The scope of the study is limited to interventional clinical trials. For the purposes of this study, “interventional clinical trials” are defined as “research in which participants are assigned to receive one

or more interventions (or no intervention) so that the effects of the interventions on biomedical or health-related outcomes can be evaluated. Assignments to [intervention]⁴ groups are determined by the study protocol” (ClinicalTrials.gov, 2012). Assignments could be at the individual or group level. An intervention is a “process or action that is the focus of a clinical trial. This can include giving participants drugs, medical devices, procedures, vaccines, and other products that are either investigational or already available” (ClinicalTrials.gov, 2012). For the purposes of this study, intervention types are limited to drugs, devices, biologics, other treatments or therapies (such as radiation), surgical procedures, behavioral interventions, and changes in the administration or delivery of clinical care.

A FRAMEWORK FOR DISCUSSION AND CALL FOR COMMENTS

This framework for discussion articulates the committee’s preliminary thoughts on guiding principles that underpin the responsible sharing of clinical trial data, defines key elements of data and data sharing activities, and describes a selected set of data sharing activities. One goal of this framework is to set the stage for identification of the numerous complicated issues that recommended strategies and practical approaches to sharing of clinical trial data might need to take into account. As with any complex policy problem, there are advantages and disadvantages to be weighed and potential for competing interests and incentives among and within various stakeholder groups. The guiding principles as defined below potentially could be in conflict or might be interpreted or prioritized differently, depending on one’s perspective.

As a first step in fulfilling the committee’s charge, the framework identifies key issues so that the public can point out omissions and begin to suggest benefits, interests, risks, and burdens of options that should be considered. The framework also identifies issues about which the committee will gather additional information through public meetings and the submission of written materials by interested parties (see Box 1 and the section below). Comparing the various public responses to the framework will help clarify the countervailing interests and values the commit-

⁴To maintain consistency in its terminology the committee has substituted the term “intervention” here for the term “treatment,” which is used in the ClinicalTrials.gov definition.

tee will take into account in making its recommendations to enhance responsible sharing of clinical trial data.

BOX 1
Specific Topics for Public Feedback

Global Implementation and Practical Consideration

- Because most large clinical trials are global in nature, how can clinical trial data be shared in that global context? How can different national regulations for research participants' privacy protections, approval of drugs and devices, data exclusivity and intellectual property laws, resources, and health priorities be taken into account?
- How might strategies and approaches regarding data sharing take into account clinical trials conducted in resource-poor settings; trials designed by citizen-scientists using data they contribute directly; and trials designed through participatory research?

Timing and Prioritization

- How might different types of clinical trial data, and different uses of shared data, be prioritized for sharing? What would be the rationale for placing a higher priority on certain types of data or analyses? What might be the advantages and disadvantages of distinguishing highest priority sharing of clinical trial data from other sharing activities?
- What might be the advantages and disadvantages to various stakeholders of sharing different types of datasets, at different points in time after the completion of a clinical trial?
- Should programs or approaches calling for or requiring new data sharing apply only to new trials undertaken from the date of a new program forward, or retroactively apply to clinical trials started before the data sharing program was initiated?

Mitigating Risks

- What might be done to minimize the risks to patients and to public health from the dissemination of findings from invalid analyses of shared clinical trial data?
- What measures should be deployed to minimize the privacy and confidentiality risks to trial participants? For example, are current anonymization or de-identification methodologies sufficient?
- Under what circumstances are identifiable data needed to fulfill articulated purposes of a data sharing activity? Under what circumstances might re-identification of trial participants be beneficial (for the participants or the public)? Have there been there examples of instances of re-identification of trial participants (e.g., for safety reasons to warn a patient of a potential risk, or for questionable and potentially unethical reasons) and what were the impacts?

Enhancing Incentives

- What incentives and protections might be established to encourage clinical trial sponsors and clinical investigators to continue to conduct clinical trials in the future, without unduly restricting the sharing of certain types of data? How do we protect or provide incentives for researchers to share data?
- What is the appropriate responsibility of the primary investigator(s) or research institution(s) to support secondary users in their interpretation of shared data, and what infrastructure or resources are needed to enable such ongoing support? For those with experience in data sharing, what is the burden of providing such support to help others understand and use the provided information?

Measuring Impact

- What would be appropriate outcome measures to assess the usefulness of different models of clinical trial data sharing, and how can they be used to guide improvements in data sharing practices?

Invitation for Public Comments

The issues identified and the options and observations described in this framework are preliminary and do not represent a comprehensive review of the subject. This framework does not assess the benefits and risks of different options and, consistent with its charge, does not contain conclusions or recommendations. Instead, this framework serves to identify areas of interest and concern that will be pursued in greater detail during the second phase of the project and addressed in the final report. The final report will also analyze the risks and benefits of options and make conclusions and recommendations. As required in the charge to the committee, the framework is being released for public comment. The committee welcomes comments from interested parties to help ensure that major concerns and issues are not overlooked, and particularly invites comments on the difficult or complex issues outlined in Box 1 below. Comments may be submitted to the committee at either of two forthcoming public workshops, or via the committee's project website, <http://www8.nationalacademies.org/cp/projectview.aspx?key=49578>.

During the course of the study, members of the public are encouraged to provide comments on this framework, as well as general comments on issues within the scope of the charge to the committee. The committee is also interested in receiving testimony and suggestions on

topics that are likely to be particularly complex, and where differing perspectives are likely to reside. At future meetings, the public will be invited to discuss these issues or to submit written statements.

GUIDING PRINCIPLES FOR RESPONSIBLE SHARING OF CLINICAL TRIAL DATA

In this framework for discussion, the committee proposes four high-level principles as a starting point for developing a framework for responsible sharing of clinical trial data (see Box 2; the principles are discussed individually below). By explicitly articulating these guiding principles and the rationale behind them, the committee hopes to bring into sharper relief the values of concern for different stakeholders that need to be acknowledged and balanced in data sharing policies and procedures.

In developing this provisional set of guiding principles, the committee drew on recent proposals for principles of data sharing in clinical trials from scholars and working groups (AllTrials, 2013; EMA, 2013; FDA, 2013; Godlee and Groves, 2012; Mello, 2013; PhRMA and EFPIA, 2013; Wellcome Trust, 2011; YODA, 2013), as well as widely accepted guiding principles articulated in official statements of research ethics and international standards, such as the Declaration of Helsinki, the *Belmont Report* (which presents the ethical rationale for current U.S. regulations for human subjects research), and others (Childress et al., 2005; CIOMS, 2002; ICH, 1996; National Commission, 1979; WMA, 2013). Policies regarding clinical trial data sharing will have a stronger intellectual foundation and practical applicability if they take into account policies on related topics. A framework for sharing clinical trial

BOX 2

Provisional Guiding Principles for Responsible Sharing of Clinical Trial Data

- Respect the individual participants whose data are shared.
- Maximize benefits to participants in clinical trials and to society, while minimizing harms.
- Increase public trust in clinical trials.
- Carry out sharing of clinical trial data in a manner that enhances fairness.

data should therefore be consistent with the principles guiding related issues, including the protection of human research participants, regulation of drugs and medical devices, scientific publications, and intellectual property protections.

In this section the committee presents potential consequences of clinical trial data sharing, which may or may not occur in any particular data sharing activity. In its further deliberations and final report, the committee will assess these potential consequences of clinical trials data sharing.

Respect the Individual Participants Whose Data Are Shared

The committee's first provisional guiding principle stems from the broadly articulated concept that respect for research participants is a fundamental principle of research ethics (ICH, 1996; National Commission, 1979).

Respect Through Research Participant Protections

Respect for research participants requires protecting their dignity, integrity, and right to self-determination; this includes, at a minimum, compliance with applicable regulations and ethical standards for the conduct of clinical trials and handling of the resulting data. Respect for research participants has historically been understood to require specific informed consent from participants (including consent for how their data will be used) before they enroll in a clinical trial in which the intervention is carried out at the individual participant level (Childress et al., 2005; CIOMS, 2002; WMA, 2013).⁵ For existing trials, data sharing (particularly sharing beyond other investigators in the trial) might not have been explicitly discussed with participants during the consent process. Sharing of data without specific participant consent might be ethically acceptable and legally permitted in certain instances. For example, if the shared data are de-identified, current U.S. federal regulations on human research protections and U.S. health information privacy regulations (e.g., the Health Insurance Portability and Accountability Act

⁵Specific informed consent is not necessarily required for trials at the group level, such as in certain cluster-randomized trials (Weijer and Emanuel, 2000) or for certain comparative effectiveness trials (Faden et al., 2013).

[HIPAA] of 1996)⁶ allow other researchers to use the data for research under certain conditions without consent from the original participants.⁷

Respect also suggests a need to protect the confidentiality and privacy of trial participants when data are shared. Questions have been raised about the sufficiency of commonly used de-identification methodologies (Benitez and Malin, 2010; El Emam, 2013; McGraw, 2012); consequently, additional protections may be needed.

Respect Through Engagement

Respect can also be demonstrated and advanced through efforts to engage participants and their representatives in the development of processes for sharing of clinical trial data (CTSA, 2011). For example, new policies and procedures regarding data sharing and subsequent additional analyses (particularly for specific trials or classes of trials as relevant) could be developed with input and feedback from representatives of research participants, disease advocacy groups, community advisory boards, and the public (Jiang et al., 2013). Such an approach would also include dissemination of information and calls for input about data sharing policies and procedures and a rationale for data sharing that is accessible and understandable to the public. The act of seeking and obtaining such input would not in itself constitute surrogate consent or authorization for data sharing. Rather, it would respect participants by actively seeking to identify concerns about and potential unappreciated benefits of data sharing that were not previously taken into account, and allow participants or their advocates to suggest how the process of data sharing might be improved (Stiles and Petrila, 2011).

Maximize Benefits to Participants in Clinical Trials and to Society, While Minimizing Harm

Understanding and balancing the potential benefits and harms of health interventions is a significant component of health care and of health intervention research and development. Similarly, there are poten-

⁶Public Law 104-191, 104th Cong. (August 21, 1996).

⁷The U.S. example has been described here for illustrative purposes. Privacy protections with respect to sharing anonymized data without re-consent vary across jurisdictions. For example, the European Union has strong data privacy protections that need to be observed when clinical trial data are shared by its member states (Article 29 Data Protection Working Party, 2013).

tial benefits and harms associated with the sharing of clinical trial data. Strategies for data sharing should maximize benefits to those who give of themselves to participate—and to society as a whole—while minimizing potential harms for various stakeholders. This provisional guiding principle for responsible sharing of clinical trial data is derived from the ethical concept of beneficence.

Potential Benefits

International ethical standards identify beneficence as a basic ethical principle and obligation of research involving human subjects. The International Conference on Harmonisation (ICH) Guideline for Good Clinical Practice declares: “Before a trial is initiated, foreseeable risks and inconveniences should be weighed against the anticipated benefit for the individual trial subject and society. A trial should be initiated and continued only if the anticipated benefits justify the risks” and “the rights, safety, and well-being of the trial subjects are the most important considerations and should prevail over interests of science and society” (ICH, 1996). Benefits include both the immediate knowledge gained from answering the hypothesis of a particular clinical trial and the broader utility of the study data in informing development of effective new interventions. As discussed in the *Belmont Report*, practitioners are faced with deciding “when it is justifiable to seek certain benefits despite the risks involved” (National Commission, 1979, p. 5). The potential utility of data is a component in the balance of potential benefits and risks when making the decision to expose individual participants to risk in order to seek benefits to society as a whole.

Clinical trials are designed and carried out to address research questions about the safety and efficacy (or effectiveness) of one or more health interventions. The interventions that participants receive are determined by the study protocol, not by what their personal physicians consider best for them as individuals. In consenting to participate, clinical trial participants also accept that complying with the study protocol potentially entails inconvenience and risks (Lidz et al., 2004). Although participation in clinical trials, on the whole, might not be significantly riskier than ordinary clinical care or receiving the same intervention outside the trial (Gross et al., 2006), in a specific trial the benefits and risks of the study arms are not known at the outset. In some instances the intervention arm of a trial will be shown to have significantly worse out-

comes than the control arm, a finding that cannot be predicted at the time of enrollment.

From the perspective of clinical trial participants, data sharing increases their contributions to generalizable human health knowledge by potentially facilitating additional findings beyond the original, predefined clinical trial outcomes. Thus, if data are not shared, opportunities are missed to generate additional knowledge from their contributions (Califf, 2013; IOM, 2013; Mello et al., 2013).

From the perspective of society as a whole, sharing data from clinical trials, if it were to yield additional results that contribute to the scientific knowledge base, could provide a more accurate, less biased, and more comprehensive picture of the benefits and risks of an intervention. Sharing clinical trial data could potentially lead to enhanced efficiency and safety of the clinical research process by, for example,

- reducing duplication of efforts and costs of future studies;
- reducing exposure of participants in future trials to avoidable harms that can be identified through data sharing;
- providing a deeper knowledge base for regulatory decisions;
- supporting the development of clinical practice guidelines; and
- allowing health care professionals and patients to make more informed decisions about clinical care.

The usefulness of clinical trial data would be enhanced if, in addition to sharing the protocol, manual of operations, statistical analysis plan, a copy of the case report forms, and metadata about the analyzable data file, the data user could ask the data generator questions regarding logistical and practical issues such as the conduct of the clinical trial or data analyses. The usefulness of shared clinical trial data might also be enhanced if data were collected in a standardized manner, where scientifically indicated (Califf, 2013; IOM, 2013; Mello et al., 2013).

In the long run, proponents have suggested that sharing clinical trial data could improve public health and patient outcomes, reduce the incidence of adverse effects from therapies, and decrease expenditures for medical interventions that are ineffective or less effective than alternatives. In addition, data sharing could open opportunities for exploratory research that might lead to hypotheses about the mechanisms of disease, more effective therapies, or alternative uses of existing or abandoned therapies that could be tested in additional research (Califf, 2013; IOM, 2013; Mello, 2013).

Potential Harms

These potential benefits of clinical trial data sharing need to be balanced against any potential harms. Data sharing could place clinical trial participants at increased risk for invasions of privacy or breaches of confidentiality, which could then lead to social or economic harms to the participants (IOM, 2013; Mello et al., 2013).

Data sharing could also result in potential harms to society. For example, shared clinical trial data might be analyzed in a manner that leads to biased effect estimates or invalid conclusions (although this might also occur with the original analyses) (Krumholz and Ross, 2011). Incorrect conclusions or treatment recommendations for either whole-patient populations or subgroups could produce suboptimal care and unnecessary anxiety, and result in possible discrimination (IOM, 2013; Spertus, 2012). Concerns about such future uses of their clinical trial data might also deter some individuals and/or communities from participating in clinical trials.

Clinical trialists and sponsors of clinical trials could also suffer potential harms due to the administrative and financial burden of data sharing, which they might regard as an unfunded mandate. The manner in which data are shared might undermine the incentives of clinical trial sponsors, clinical investigators, coordinating centers, researchers, and other essential stakeholders to invest time and resources in the development and clinical testing of potential new treatment practices (Dickersin, 2013; Rathi et al., 2012). For example, data sharing might allow confidential commercial information (CCI) to be discerned from the data (Teden, 2013). Competitors might use shared data to seek regulatory approval of competing products in countries that do not recognize data exclusivity periods or that do not grant patents for certain types of research. The manner in which clinical trial data are shared might also harm the intellectual capital and professional recognition of academic clinical investigators, who devote considerable effort and time to designing a clinical trial, recruiting and retaining participants, and collecting the primary data. If subsequent independent analyses fail to give appropriate recognition to the original investigators, there will not be incentives for investigators to share clinical trial data.

Increase Public Trust in Clinical Trials

Public trust is an intrinsic value undergirding the biomedical science and health research enterprise, which is fundamentally aimed at improving human health. Trust is also essential for ensuring continued public support for clinical research and for fostering participation in clinical trials. The concept of public trust in clinical trials encompasses both trust in the scientific process of generating data (i.e., that there is accountability for how the trial is carried out) and trust in the validity of the reported trial findings (i.e., that the reported findings are an accurate representation of the underlying data) (IOM, 2013). The sharing of clinical trial data could be carried out in ways that enhance or undermine public trust in clinical research. The process of data sharing should therefore be undertaken in a manner that enhances, rather than undermines, public trust in both the clinical trial process and the data sharing process.

Trust in Clinical Trial Data

By increasing the transparency of trial design and conduct and the pathway to trial conclusions, sharing clinical trial data might increase public trust in the outcomes of specific trials and of clinical trials generally (Loder, 2013). Data sharing might also increase the usefulness and trustworthiness of clinical trial data and analyses because clinical researchers who know that others will be using their data might be more thorough and more careful in their methodology and its documentation. Such additional attention to detail might also help to reduce bias in the data and findings (Mello et al., 2013).

Sharing clinical trial data could enhance public trust by facilitating confirmatory analyses that could determine whether the final conclusions and summaries of clinical trials are robust and valid inferences from the original evidence, although this must be done in a credible and fair manner (Laine et al., 2007). Whether the inferences from a particular trial are strong or weak, activities seeking to demonstrate widespread applicability of the findings could enhance overall trust in the scientific process and in the resulting evidence-based recommendations for clinical care.

Trust in clinical research could be enhanced further if sharing clinical trial data were accompanied by outreach and engagement to help the public understand that numerous judgments are needed to transform source data into analyzable data (CTSA, 2011), and that highly trained researchers might take different approaches to answering a research

question or to analyzing a given dataset. Different analytical approaches and interpretations of data by researchers are an expected part of the scientific process and discussions. Such outreach could also help the public better understand that findings from individual clinical trials are often not definitive, and that attempts to replicate original analyses or the conduct of meta-analyses using pooled data from multiple clinical trials can strengthen, modify, refute, or extend the original reports from a trial. Moreover, trust would be enhanced if additional analyses using data from a clinical trial were easily connected with the publications from the original trial.

Trust in the Data Sharing Process

Sharing clinical trial data could carry the risk of undermining public trust in clinical trials under certain circumstances, for example, if multiple analyses yield conflicting or invalid conclusions (Califf, 2013). Public trust in clinical trials where data are shared might be undermined unless there are clear, transparent, and accountable processes for the data sharing, including established criteria for clinical trial data sharing, procedures for fairly adjudicating requests for data against those criteria, mechanisms for ensuring or confirming the integrity of data when they are collected and analyzed by different parties, and accountability for both data holders and requesters in adhering to those standards. Clear, transparent, and accountable processes for data sharing would also include protection of patient privacy and respectful handling of individual patient data. Further, public trust could be increased if the public sees evidence of the incorporation of its perspectives (whether through the mechanisms described above or by addressing specific community concerns). If analyses of shared data use inappropriate methods and statistics and lead to biased conclusions, they can undermine patient trust in valid conclusions about the study intervention. Such mistrust might ultimately lead to seriously flawed clinical care decisions, unwarranted patient concerns about the quality of care, and avoidable patient anxiety.

Carry Out Sharing of Clinical Trial Data in a Manner That Enhances Fairness

Fairness, broadly articulated, is a core ethical principle that is applicable in several ways to sharing clinical trial data. In very general terms, fairness entails persons receiving what is due to them or what they de-

serve (Beauchamp and Childress, 2009). Fairness requires similar treatment of all people (as individuals or as part of groups, entities, processes, etc.) unless there are justifiable reasons to treat certain people differently. Disagreements about fairness arise in specifying what an individual or group is due or deserves, what are sufficient reasons for different treatment under what might be perceived by some as similar circumstances, and whether inequity, unfairness, or injustice has occurred. All individuals and organizations involved in the clinical trial enterprise have a stake in the fairness of data sharing. In the context of this discussion framework, fairness addresses the distribution of the risks and potential benefits of clinical trial data sharing, whereas beneficence generally refers to the aggregate risks and potential benefits.

Fairness in Exposure to Potential Risks of Data Sharing

Clinical trial data sharing should be carried out in a fair manner so that it does not repeat, in the data sharing context, well-documented historical examples of placing disproportionate risks of clinical research on vulnerable groups (Bioethics Commission, 2011; Emanuel et al., 2008; Jones, 2008; Wertheimer, 2008). For example, data sharing should include protections for participant subgroups that are particularly vulnerable to breaches of confidentiality or other adverse consequences of data sharing. Clinical trial participants could be particularly vulnerable to harm if they have conditions, or are members of groups, that are often stigmatized. In this regard, there might be justifiable and ethical reasons for handling some types of clinical trial data sharing differently to reduce the potential for unfair treatment of participants. For example, whole-genome sequencing data could be identifiable (Gymrek et al., 2013), which might be considered to put participants at heightened risk, and additional safeguards or protections for participants whose genomic data could be shared might be warranted.

Fairness in the Distribution of the Potential Benefits of Research

Clinical trial data sharing could increase fairness in clinical research by helping to distribute the benefits of clinical research more equitably across different groups of participants and communities. Pooling shared data from several clinical trials could, for example, benefit groups who have been enrolled in clinical trials in such small numbers that the statistical power in any single trial to draw valid inferences about risks and

benefits is limited. Among the underserved groups for whom data sharing might accelerate research are individuals with rare conditions, rare subtypes of common conditions, and members of certain ethnic groups that historically have low enrollment in clinical trials. Underrepresentation of these groups in clinical trials can lead to a weaker evidence base for clinical care decisions, as well as health disparities and discrimination (IOM, 2002).

Clinical trial sponsors, clinical investigators, and coordinating centers that design and carry out clinical trials might consider fairness to include appropriate recognition, protection, and reward. Investigators who make substantial investments of intellectual capital, time, and resources have an interest in carrying out additional analyses of the data they have collected and in receiving due credit when others take advantage of these data. In addition, there are administrative burdens when sharing data with others that need to be fairly distributed among those who collect and control clinical trial data, those who seek access to shared data for additional research, and society. Appropriate protection of these interests could help provide incentives (or reduce disincentives) to share data and to conduct future clinical trials.

ADDITIONAL OBSERVATIONS RELATING TO PRINCIPLES FOR DATA SHARING

Operational Strategies Following from Guiding Principles

The principles described in this framework suggest several possible operational strategies that approaches for responsible data sharing might take into account. The committee will consider such operational strategies in its further deliberations and in formulating recommendations in the final report and welcomes comments about operational strategies. Implementation of these principles might take into account operational considerations, such as

- timing of when data are shared;
- proportional consideration of benefits, burdens, and risks to various parties in adopting a data sharing program or activity;
- opportunities to embed “learning” in a clinical trial data sharing system, permitting tracking of the beneficial outcomes of different approaches to data sharing, identification of unanticipated

adverse consequences of approaches, dissemination of best practices for data sharing, and promotion of ongoing quality improvement in data sharing systems and approaches; and

- need for data sharing systems and approaches to be globally applicable and practically achievable.

Framing Rights and Responsibilities: To Whom Do the Benefits of Clinical Trial Data Belong?

Foundational to any discussion of possibilities and needs for clinical trial data sharing is a fundamental philosophical question: To whom do the benefits of the data belong? Are data primarily for the benefit of the public or for private parties? This is a separate question from a technical or legal analysis of which party owns the data or which parties have certain property interests in the data, which includes control over access to and use of the data.

With respect to ownership of clinical trial data, academic institutions that receive research grants might claim ownership of the data collected during the research in order to comply with regulatory requirements (Drazen, 2002). Private funders of clinical trials might claim that they own the research data, particularly if the data will form part of a submission to the U.S. Food and Drug Administration (FDA) for regulatory approval. The language of research grants and contracts and the wording of informed consent forms that participants sign in clinical trials could also delineate what party owns or has rights of disposition of data. For example, the NIH includes data sharing requirements in the terms and conditions of its research grants with direct costs greater than \$500,000 (NIH, 2003). A property owner does not always have absolute dominion over the property in question; others may have legal access to it under certain conditions for certain purposes.⁸ Ultimately, the question of who owns the data is less important than asking what are the rights and responsibilities of data holders. The committee intends to engage the question of to whom ought the benefits of clinical trial data belong as a thought experiment to help inform its analysis.

On one hand, the benefit of clinical trial data might be regarded as primarily belonging to the public—in order to benefit patients and society through the advancement of science and clinical knowledge that leads to improved patient care. From this perspective, some might argue that

⁸Moreover, property may be taken for public use without consent of the owner, subject to Constitutional requirements for due process and fair compensation (Evans, 2011).

sharing clinical trial data ought to be a *prima facie* obligation. That is, the default policy—the presumption—should be to share data, with justification needed to restrict sharing or recognize an exception to sharing.

On the other hand, the benefit of clinical trial data might be regarded as primarily belonging to organizations and individuals who invested resources and time to plan and carry out the clinical trial and analyze the data. The rationale for this perspective might be the provision of fair rewards for investment and work in carrying out the trial, or it might be the risk that organizations and individuals might otherwise lack appropriate incentives to develop new tests and therapies. From this perspective, the policy presumption might be that clinical trial data sharing should only be undertaken if the interests and rights of those who carried out the trial are appropriately protected and incentivized. Some might argue that clinical trial data sharing should be optional and voluntary, at the discretion of the organization and individuals who invested resources and time to conduct the clinical trial.

While we have framed these positions as a dichotomy for the sake of discussion, another approach would be to balance both perspectives in a policy of clinical trial data sharing, in addition to adhering to the other guiding principles discussed above, as well as applicable existing, modified, or new legal requirements. Any attempt to balance competing interests would require careful attention to the conditions of sharing, responsibilities, limitations, and exceptions.

DATA SHARING ELEMENTS AND ACTIVITIES

The following sections define key elements of data and data sharing activities, such as the type(s) of data to be shared; provider(s) and recipient(s) of shared data; and whether and when data are disclosed publicly, with or without restrictions, or exchanged privately among parties. This section then describes a selected set of data sharing activities. The purpose of outlining potential data sharing activities is to provide an heuristic approach to organizing the work of the committee throughout the course of the study, including information gathering and discussions in public sessions. In its final report, the committee will, with respect to each of these data sharing activities, present findings relating to benefits, risks, and burdens associated with these data sharing activities and suggest strategies and practical approaches to facilitate responsible data sharing. The data sharing activities noted in this discussion should not be

interpreted as a conceptual framework that would necessarily lead to particular findings, conclusions, or recommendations.

What Types of Data Could Be Shared?

Current Practices in Data Disclosure

During the course of a clinical trial, different types of data are collected, transformed into analyzable datasets to address specific research questions, and used to generate various publications and reports for different audiences (Drazen, 2002) (see Figure 1).

Publication in peer-reviewed scientific journals is currently the primary method for sharing clinical trial data with the scientific and medical communities, as well as the public (often through media coverage of published findings). These publications, however, contain only a small subset of the data collected, produced, and analyzed in the course of a trial (Doshi et al., 2013; Zarin, 2013). Scientific journal articles generally contain a brief summary of the trial background, research question(s), methodology, results, figures and tables, and discussion.

Clinical trial sponsors seeking regulatory approval from authorities such as the FDA and the EMA must submit detailed CSRs (discussed below) and IPD as required, which forms the basis of the marketing application for a product. In trials that are not conducted for regulatory approval of a product, detailed CSRs may or may not be prepared (Doshi et al., 2012; Teden, 2013).

Beyond the selected clinical trial data that are disclosed in journal publications, IPD and more detailed clinical datasets have not been routinely shared with the broader scientific community or the public. Some sponsors in both industry and academia have shared IPD and summary data reports upon request and on a case-by-case basis (Rathi et al., 2012). There are a number of initiatives at the NIH to share clinical trial data from the time of publication (Immune Tolerance Network, 2013; NHLBI, 2007). Recently, proposals have been put forth for more proactive sharing of both CSRs and IPD, and several plans have been announced or implemented (EMA, 2013; Krumholz and Ross, 2011; Kuntz, 2013; Loder, 2013; Nisen and Rockhold, 2013; *PhRMA* and *EFPIA*, 2013; *YODA*, 2013).

Many discussions of clinical trial data sharing thus far have not been specific regarding which of many possible clinical trial data elements or datasets might be shared. This framework articulates more specific defi-

nitions and descriptions of data that might be shared to help facilitate more focused discussions among the public and various stakeholders.

Data

The committee has identified various types of clinical trial data (containing differing levels of detail) that might be included in a data sharing activity (see Figure 1).

Raw data Sometimes called source data, raw data are observations about individual participants. These data might be collected specifically for the study protocol, or as part of routine care and used by the investigators. At the source, data might be in the form of measurements of participant characteristics such as weight, blood pressure, or heart rate, and can be associated with the baseline (or initial) visit or subsequent followup visits.

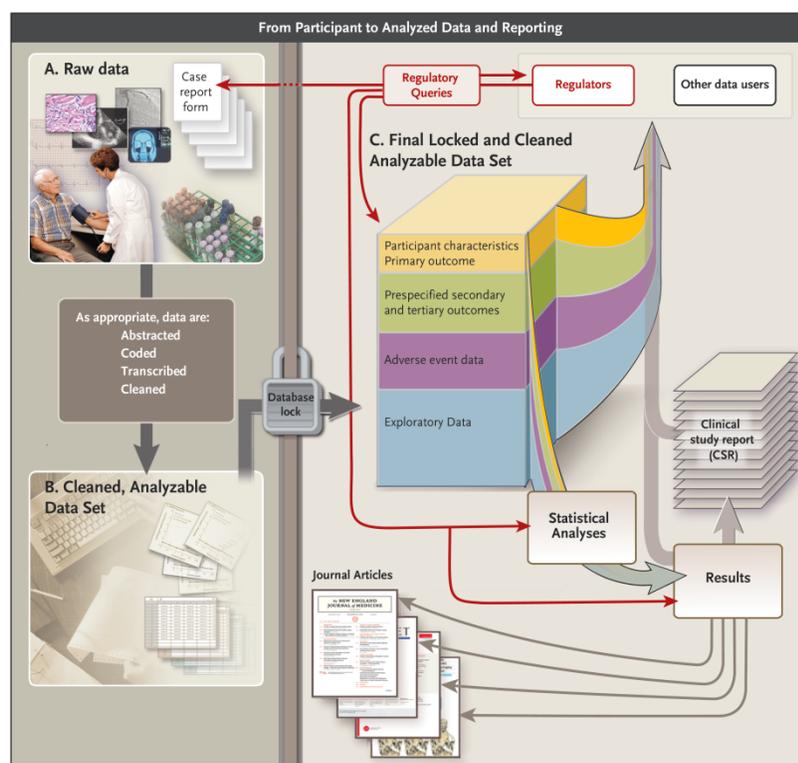


FIGURE 1 Data flow from participant to analyzed data and reporting.
NOTE: See Appendix C for more detail about each element of the figure.

Raw data might also include a baseline description of the participant's medical history, physical exam information, clinical laboratory results (e.g., serum lipid values, hemoglobin levels), whole exome or genome sequences, or imaging (e.g., X-ray, magnetic resonance imaging). Depending on the trial, demographics, clinical data, and other appropriate raw source information are entered into case report forms. Some data must be abstracted and/or interpreted for the purposes of the protocol, for example, reading the X-ray for tumor size or evaluating the electrocardiogram for evidence of a heart attack. Data might also include assessments by clinical study staff or adjudication committees to determine whether specific clinical end points or adverse events in the participant's profile (e.g., heart attack, death) meet protocol-specified criteria. In addition to "traditional" clinical trial data, other types of health data are increasingly being collected, including self-reported measures (e.g., quality of life), quantified sensor data (e.g., readings from remote monitoring devices, including smartphone apps and geolocation data), consumer genomics data (e.g., from companies like 23andMe), and community-level self-reported data (e.g., from sites like PatientsLikeMe).

Data entry into the database After collection (and assessment, abstraction, or adjudication as appropriate), source data typically must be entered into an organized data management system (i.e., database) for further evaluation and processing. Data typically undergo a process of cleaning, quality assurance, and quality control to detect inconsistent, incomplete, or inaccurate entries, and to confirm that the data were collected and evaluated according to the protocol and that they match the source data. This process continues throughout the course of the trial as data are collected.

Analyzable dataset Typically, after data are entered in computerized form, new variables are mathematically generated to serve as the basis for later analyses. These variables are sometimes called "derived" variables. For example, patient age might not be entered directly, but calculated by subtracting the birthdate from the date of a given clinic visit. "Treatment response" could be entered as a mathematical comparison of lesion sizes recorded from two images. After the trial is declared complete, the edited and cleaned data are moved into an analyzable data file and locked (i.e., no further changes may be made). If the study is blinded (or masked), then after the database is locked, the treatment code file is merged with the analyzable data file, and the data are unblinded to the investigators. Some or all of this now-unblinded analyzable dataset will then be used for data analyses.

A statistical analysis plan (SAP) is finalized before the trial is completed and unblinded; the SAP drives the initial analyses of the analyzable dataset. The trial protocol should contain both a basic SAP and a more detailed SAP that includes the analyses and any interim analyses to be conducted, as well as the statistical methods that will be used, as determined by the protocol. The full SAP includes, for example, plans for analysis of baseline descriptive data, adherence to the intervention, primary and secondary outcomes, definition of adverse events and serious adverse events, as well as the comparison of these measures across interventions for pre-specified subgroups. The analysis may be very extensive and can result in several dozen tables and figures.

For many clinical trials, the SAP-defined analysis might not use all of the data available in the analyzable dataset. Moreover, peer-reviewed journal publications of clinical trials generally draw on only part of the analyzable dataset. Supplemental data are often collected to permit exploration of ancillary questions not directly related to the primary purpose of the protocol, and researchers might conduct exploratory and post hoc analysis not defined in the SAP to answer additional questions.

Reports Generated from Data

Box 3 lists some of the types of reports that are commonly generated from the analyzable dataset. In addition to preparing peer-reviewed journal publications describing the primary and major secondary outcomes specified in the protocol, trial sponsors prepare various additional reports, including results summaries for registries, lay summaries, and clinical study reports.

Publications Several scientific journal publications are commonly derived from the analyses driven by the SAP and from post hoc analyses. Typically, a primary publication will address the primary and possibly the leading

BOX 3 **Types of Summary Reports**

- Lay-language summary
- Registry results summary (e.g., for ClinicalTrials.gov)
- Publications (including peer-reviewed scientific journals)
- Clinical Study Report (CSR)
 - Full CSR, with or without appendixes
 - CSR synopsis (executive summary)
 - Redacted CSR
 - Abbreviated CSR

secondary outcome measures specified in the protocol. The primary publication would also include the baseline measures to demonstrate participant comparability between intervention arms and comparisons of any adverse events of major interest or frequency. Subsequent journal publications might address in more detail a specific aspect of the primary analysis that was not included in the primary publication or analyze outcomes in particular prespecified subgroups of participants. Ideally, each journal publication should have a specific dataset corresponding exclusively to the data used to generate the tables and figures in the publication (which would be a subset of the full analyzable dataset). Each specific dataset is typically stored in separate sets of data files to document the data used for each journal publication.

Registry results summary and lay-language summary Many clinical trials are subject to a requirement that their results be reported to one or more registries, in formats specified by the particular registry (for example, results of trials of FDA-regulated products must be reported to ClinicalTrials.gov in the United States). These summaries are publicly available on the registry website and are generally limited to major outcomes and adverse events.

A lay-language summary is a brief, nontechnical overview written for the general public and trial participants. Lay summaries of the clinical trial protocol are often required by institutional review boards (IRBs) to assist nonscientist members of the IRB in the protocol review and approval process; however, the preparation of lay-language summaries of clinical trial results is uncommon. A recent study suggests that trial participants would value such summaries and that provision of lay-language results summaries to participants is feasible (Getz et al., 2012).

Clinical Study Report (CSR) When a clinical trial is submitted to regulatory agencies as part of an application for marketing approval of an intervention or approval of a new indication, trial sponsors usually submit a detailed CSR. Specifications for CSRs were defined by the ICH and adopted by the FDA, the EMA, and the Japanese Ministry of Health, Labor, and Welfare in an effort to simplify the application process for new interventions globally (ICH, 1995). According to the FDA guidance, the CSR is an

integrated full report of an individual study of any therapeutic, prophylactic or diagnostic agent ... conducted in patients. The clinical and statistical description, presen-

tation, and analyses are integrated into a single report incorporating tables and figures into the main text of the report or at the end of the text, with appendices containing such information as the protocol, sample case report forms, investigator-related information, information related to the test drugs/investigational products including active control/comparators, technical statistical documentation, related publications, patient data listings, and technical statistical details such as derivations, computations, analyses, and computer output. (FDA, 1996, p. 1)

Although a CSR contains mainly summary data and summary tables and graphs, it also usually contains considerable additional information (often thousands of pages), including, as described in the definition above, numerous large appendixes. Supplemental information can include detailed narratives describing individual participants. In some instances, the CSR and/or its appendixes might include identifiable participant or commercially confidential information or other protected health information or intellectual property. A CSR synopsis (i.e., executive summary) is sometimes drafted to accompany a full CSR. Some of the supporting clinical trials included in a regulatory submission do not directly contribute to the evaluation of effectiveness of the intervention; for these studies, sponsors may be permitted to submit an abbreviated CSR (FDA, 1999). Some CSRs may also be redacted before they are shared to remove any commercially confidential information and personally identifiable information.

Metadata and Additional Documentation

In order for researchers to make use of clinical trial data that is shared with them (e.g., to perform confirmatory analyses or carry out exploratory analyses), they need further information or metadata (i.e., “data about the data”) in addition to the data elements or datasets described above. Box 4 summarizes some of the metadata that might be needed to facilitate full use of shared data. Supporting documentation critical to interpretation of shared clinical trial data include the full protocol, manual of operations, consent form, case report forms, and the SAP.

The trial protocol describes the trial rationale; the eligibility and exclusion criteria for participants; the primary and secondary hypotheses

and the corresponding primary and secondary outcome measures; the methods used to gather and adjudicate adverse events; other measures intended to evaluate the intervention; and a full description of the intervention and how it is administered. While a trial protocol provides the overall experimental design, a detailed manual of operations describes how the trial was conducted. A copy of the template for the informed consent form describes what participants agreed to, what hypotheses were included, and the additional purposes for which their data might be used. Case report forms capture precisely what measures were made, and at what time points during the trial, as defined in the protocol. The SAP sets out how each data element was analyzed, what specific statistical method was used for each analysis, and how adjustments were made for testing multiple variables. If some analysis methods required critical assumptions, data users would need to understand how those assumptions were verified. For key analyses, full use of the shared data would be aided by providing the computer software and version used, as well as the statistical programming code for the statistical software used for each analysis.

BOX 4**Metadata and Additional Documentation to Support the Use of Shared Clinical Trial Data**

- Clinical trial registration number and dataset (available through ClinicalTrials.gov and other registries)
- Full protocol (e.g., all outcomes, study structure), including first version, last version, and all amendments
- Manual of operations (e.g., assay method) and standard operating procedures, including names of parties involved, specifically:
 - names of persons on the clinical trial team, trial sponsor team, data management team, and data analysis team; and
 - names of members of the steering committee, clinical events committee (CEC, which adjudicates end points), data and safety monitoring board (DSMB) or data monitoring committee (DMC), as well as committee charters
- Details of study execution (e.g., participant flow, deviations from protocol)
- Case report forms, informed consent forms, biospecimen information
- SAP, all amendments, and all documentation for additional work processes (including codes, software, and audit of the statistical workflow)
- Publications and report documents (see Box 3)

Summaries and reports (e.g., publications from the trial, lay-language summary of trial results, trial registry results summary, CSR synopsis) would also help shared-data recipients understand and make the most efficient use of shared data.

There are variations in how these data elements are defined and in the terminology used to describe them. In its future deliberations, the committee will seek clarity and consistency in use of the terms. The final report will include discussion of how the information in the analyzable dataset, CSR, and IPD differ and which types of analyses, either confirmatory or exploratory, require which level of data sharing.

Who Are the Providers of Shared Data?

Data are generated at almost every step in the clinical research process, from the initial collection of baseline participant data to the analysis of the analyzable dataset. Different individuals or organizations hold or control the data at different times during the course of the trial (e.g., laboratory technicians, investigators, database administrators, statisticians, DSMBs, sponsors, and regulatory agencies). A data holder may or may not have legal authority to share the data with others. In contrast, an individual or organization with the authority to share the data might not have physical possession of the data at a particular time. Thus, responsibility for providing data for sharing might need to be coordinated among several entities. Potential entities that are likely to be data providers include (but are not limited to) the following:

- Individual participants in a clinical trial (the initial “providers” of data to researchers). Some participants might hold data to the extent that they self-generate and transmit the data (from self-quantifying devices), retain copies of the data, or receive information from investigators. Participants could, in turn, share their information with organizations that aggregate data from many participants (e.g., disease advocacy groups, research platforms such as PatientsLikeMe, Reg4ALL, or Sage Bionetwork’s Bridge).
- Clinical trial funders (e.g., government, industry, foundations, or advocacy organizations).
- Contract research organizations that collect source data from participants on behalf of sponsors.
- Principal investigators or their institutions.

- Site principal investigators of a multisite trial or their institutions.
- The data or biostatistics coordinating center or the institution hosting the center.
- Regulatory agencies to which data are submitted.
- Systematic reviewers and guideline developers.

Potential users of shared data will need to know whom to contact to obtain the data they seek, who owns the data, who controls the data, and where they can get answers to questions that will inevitably arise about the dataset. As such, data providers might have an ongoing resource role, beyond simply sending data and metadata to recipients. The final report will include analysis of the advantages and disadvantages of various actors having responsibility for providing data to be shared.

Who Are the Recipients of Shared Data?

Many individuals and entities could be recipients of shared data from clinical trials. These include (but are not limited to) the following:

- Individuals participating in the trial, as a part of the agreement for participation, for a variety of reasons described above (e.g., trust, transparency, respect, engagement).
- Researchers seeking to reanalyze a study or explore new scientific questions.
- The institutions supporting the researchers.
- Funding agencies (e.g., government, private sector).
- IRBs or scientific peer review committees reviewing a new study of the same or a similar intervention in order to have a more comprehensive safety profile of the intervention.
- The DSMB or DMC for another clinical trial, whose decision to recommend continuing or stopping that trial can be informed by the results of a completed trial that has not yet been or will never be published.
- Educators requesting to use a dataset for teaching purposes (e.g., in a biostatistics class).
- A disease advocacy group seeking to advance research.
- Prospective plaintiffs or attorneys seeking information that could be used in current or future litigation.

- Competitors of the industry sponsor of the intervention studied in the trial.
- Members of the media.
- Interested members of the public.

Potential data recipients are interested in different types of data for potentially very different purposes.

When Might Clinical Trial Data Be Shared?

Data might be shared at various points in the timeline of a clinical trial, for instance,

- After publication of the primary results of the clinical trial in a peer-reviewed journal.
- After discontinuation of development of an intervention by a sponsor.
- After completion or early termination of a clinical trial.
- After regulatory approval of a new intervention or a new indication for the intervention.
- Following the occurrence of serious adverse events.
- Earlier, at the discretion of the data provider or generator.

The final report will include consideration of the advantages and disadvantages associated with disclosure of clinical trial data at various time points. The timing of data sharing could have very important implications that need serious consideration. For example, timing of release of the data will have consequences for the interests of the clinical trial team (e.g., the impact of data sharing on the timing of further publications from the data). Timing is also of great interest to the trial sponsor (e.g., relative to the timing of securing intellectual property rights or regulatory approval).

How Might Data Be Shared?

A variety of models for clinical trial data sharing have been proposed, planned, or implemented (select examples are summarized in Box 5). The types of data that are shared differ across the models. Proposed models of data sharing have generally imposed some sort of restriction on the sharing of data that could directly or potentially identify trial partici-

pants, as well as data that reveal CCI or trade secrets or might result in inaccurate analysis. Access to clinical trial data in current models ranges from essentially full access to de-identified data to fully restricted or no access.

Open Access

In an open or public access model, data are made available, at a defined time, to any party who seeks them, for any purpose. For example, the EMA has announced that it will release, to any data requester that is a known entity to the agency, both summary and participant-level data (excluding, for example, personally identifiable data and information the EMA deems to be CCI) immediately after a regulatory decision about a new drug (Eichler et al., 2013; EMA 2013; Immport, 2013; Immune Tolerance Network, 2013; NHLBI, 2007).

Controlled Access

In some models of data sharing, access is restricted to specific classes of user or for specific purposes. Requestors might need to demonstrate that they meet specified eligibility criteria. Some models require only the name and contact information of the requestor, while others require information about the proposed use of the requested data or how the data will be analyzed. Some models might also impose conditions relating to whether the data generators would receive credit in publications.

In some cases, the actual data are not provided to the requestor. Instead, data holders might run specific data analyses for approved requestors and deliver to the requestors only the results of the requested analyses. In another model, recipients receive credentials to access and run queries on the data, but are not able to download or obtain copies of the data.

Data sharing can also take place indirectly, through a “trusted intermediary” or “honest broker,” who either negotiates the conditions for data sharing (with the data provider retaining control over the data and its release) or takes full control of the data and brokers both the conditions for data release and the delivery to recipients (Mello et al., 2013). Trusted intermediaries might also accept and facilitate data analysis queries from secondary investigators, as mentioned above. The use of a trusted intermediary raises a number of issues, including selection, administration, funding, and compensation.

A controlled-access data sharing model could implicitly or explicitly address issues such as specification of data that will be shared; any categories of data that are to be specifically restricted; and how to address risks of breaches of confidentiality, attempts to re-identify trial participants, data retention periods, scientifically inappropriate analyses, and the need for timely responses to data requests.

To obtain shared data under a controlled-access model, a recipient typically must execute a data use agreement (DUA). Conditions in the DUA might include

- prohibitions on re-identification or contact of individual trial participants;
- requirements to acknowledge the providers of the data in any publications resulting from the shared data;
- requirements to send copies of submitted manuscripts and publications to the trial investigators or study sponsor;
- restrictions on further sharing of the data with additional parties or using the data for purposes other than originally proposed;
- assignment of intellectual property rights for discoveries from the shared data;
- requirements to publish or post findings from the data; and
- requirements to notify industry sponsors of the trial of any findings that raise safety concerns.

There might also be limits on the length of time that a recipient may use or access shared data (ADNI, 2013; Harvard University MRCT, 2013; Nisen and Rockhold, 2013; P4RMA and EFPIA, 2013; YODA, 2013).

Selected Set of Clinical Trial Data Sharing Activities

The possible models and approaches to clinical trial data sharing, and the purposes motivating that sharing, are extensive. The rationale, benefits, risks, and burdens associated with one particular data sharing activity could differ from those of another data sharing activity, depending on the data elements and parties involved.

To stimulate public comments and to provide heuristic organizational structure to its work, the committee has described a selected set of data sharing activities as examples of the types of arrangements or approaches under which clinical trial data might be shared (see Box 5). To derive this selected set, the committee reviewed a range of existing and pro-

posed data sharing activities and distilled them into four conceptual categories that represent broad “families” of activities that have key features in common. No single proposed or enacted data sharing activity has all of the characteristics in the familial description, and different data sharing activities in the same category can have important differences. The categories therefore are derived from, but do not describe any specific, data sharing activities currently underway or proposed. To the extent there are redundancies in characteristics within the models, the committee will address them or take them into account in its forthcoming analysis.

Each activity is described according to the type(s) of data that could be shared; provider(s) of data; recipients of data; timing of data sharing; conditions or qualifications for access; and conditions of data use. The descriptive characteristics are an illustrative but not exhaustive list. Detailed descriptions and, particularly, conclusions and recommendations regarding those descriptive characteristics or strategies and approaches for sharing will be included in the final report.

BOX 5

Set of Clinical Trial Data Sharing Activities

1. Open Access

A data sharing program or system in which data are made broadly available to the public through an open-access website. Data might be aggregated from multiple sources (i.e., more than one institution, company, or researcher), or the website might provide open access to data from one trial or one institutional or individual researcher. Aggregated data might require trial- or institution-specific queries. An **open-access data sharing activity** could have some or many of the following descriptive characteristics:

- a. **Type of data:** summary data or anonymized IPD
- b. **Providers of data:** as determined by the organizational structure of the data sharing activity. Providers could include sponsors of applications for regulatory approval of products, researchers funded by a government source, or other data generators or aggregators (such as regulatory agencies) required or agreeing to share data
- c. **Recipients of data:** all members of the public via a website
- d. **Timing of sharing:** upon publication or after regulatory decision or abandonment of drug development
- e. **Conditions or qualifications for access:** none; no or minimal identification/log-in requirements; no governing body to make access determinations
- f. **Conditions of data use:** no conditions of use or requirement for signed data use agreement; expectation to abide by applicable

- law (e.g., intellectual property protections, privacy protections)
- a. **Example activities:**^a EMA Category 2, National Institute of Allergy and Infectious Diseases' Trial Share, Immunology Database and Analysis Portal (ImmPort); National Heart, Lung, and Blood Institute's BioLINCC
2. **Controlled Access to Individual Company, Institution, or Researcher Data**
A data sharing program or system in which data are made available to requestors on a controlled-access basis, pursuant to defined restrictions or conditions. Data are from one institution or researcher. Data might require or permit trial- or product/intervention-specific queries. A **controlled access–individual entity data sharing activity** could have some or many of the following descriptive characteristics:
 - a. **Type of data:** summary data, anonymized, or limited dataset
 - b. **Providers of data:** an individual company, institution, researcher, or other entity generating or assembling data (e.g., a patient advocacy organization)
 - c. **Recipients of data:** members of the public via a website
 - d. **Timing of sharing:** upon publication or after regulatory decision or abandonment of drug development
 - e. **Conditions or qualifications for access:** access controlled by the institution/data generator or by a specified governing body such as a learned intermediary. Controls could include requirements to register, have specified expertise (e.g., statistics), successfully complete a test to demonstrate expertise, disclose funding sources, provide a purpose for the data request that meets specified criteria (e.g., related to public health or patient care), or provide an analysis plan (which might or might not need to meet specified criteria)
 - f. **Conditions of data use:** restrictions or conditions on data use might include a requirement to sign a data use agreement, requirement to provide and adhere to a publication plan (which might need to meet specified criteria), prohibition on use of data for commercial purposes, prohibition on attempts to re-identify data, requirement to inform data generator about any safety concerns identified, agreement that the data generator retains exclusive rights to inventions or other intellectual property generated by the data recipient, or requirement to acknowledge the source of the shared data and the investigators in the original clinical trial
 - g. **Example activities:** PhRMA/EFPIA commitments, Harvard Multi-Regional Clinical Trials Center proposal, Yale Open Data Access
 3. **Controlled Access to Pooled or Multiple Data Sources**
A data sharing program or system in which data are made available to requestors on a controlled-access basis pursuant to defined restrictions or conditions. Data are aggregated from multiple sources (i.e., more than one institution, company, or researcher). Aggregated data might require or permit trial- or institution-specific queries. A **controlled access–multiple entity data sharing activity** could have some or many of the following descriptive characteristics:

- a. **Type of data:** summary data or anonymized IPD
 - b. **Providers of data:** as determined by the organizational structure of the data sharing activity. Providers could include sponsors of applications for regulatory approval of products, researchers funded by a government source, or other data generators as required or agreed on
 - c. **Recipients of data:** members of the public via a website
 - d. **Timing of sharing:** upon publication or after regulatory decision or abandonment of drug development
 - e. **Conditions or qualifications for access:** access controlled by the institution/data generator or by a specified governing body, such as a learned intermediary. Controls could include requirements to register, have specified expertise (e.g., statistics), successfully complete a test to demonstrate expertise, disclose funding sources, provide a purpose for the data request that meets specified criteria (e.g., related to public health or patient care), or provide an analysis plan (which might or might not need to meet specified criteria)
 - f. **Conditions of data use:** restrictions or conditions on data use might include a requirement to sign a data use agreement, requirement to provide and adhere to a publication plan (which might need to meet specified criteria), prohibition on use of data for commercial purposes, prohibition on attempts to re-identify data, requirement to inform that data generator about any safety concerns identified, agreement that the data generator retains exclusive rights to inventions or other intellectual property generated by the data recipient, requirement to acknowledge the source of the shared data and the investigators in the original clinical trial
 - g. **Example activities:** GSK, EMA Category 3 (clinical trial data with Protection of Personal Data concerns), Alzheimer's Disease Neuroimaging Initiative
4. **Closed Partnership/Consortium**
A data sharing program in which data are shared between or among two or more parties. Arrangements could be between or among public, nonprofit, or for-profit entities. Data are not intended to be shared with individuals or others who are not party to the partnership or consortium arrangement. A **closed partnership/consortium data sharing activity** could have some or many of the following descriptive characteristics:
- a. **Type of data:** summary data, IPD (anonymized or not)
 - b. **Providers of data:** as determined by the parties to the arrangement or initiators of the data sharing activity. Providers could include companies or other product development organizations, individual researchers or institutions, or other entities generating data (e.g., a patient advocacy organization) as agreed upon
 - c. **Recipients of data:** members of the partnership or consortium
 - d. **Timing of sharing:** as agreed upon by the terms of the partnership or consortium arrangement

- e. **Conditions or qualifications for access:** access controls determined by partnership/consortium arrangement
- f. **Conditions of data use:** restrictions or conditions on data use might include a requirement to sign a partnership/consortium agreement, an agreement governing rights to inventions or other intellectual property generated by the data sharing activity, or an agreement dictating whether and how future publications will be undertaken
- g. **Example activities:** PatientsLikeMe

^aActivities listed here are considered examples of the described data sharing activity because they have some or many of the listed descriptive characteristics. No one activity would necessarily have all of the listed characteristics (and some of those characteristics might be mutually exclusive). The activities cited as examples could vary substantially from each other depending on their descriptive characteristics.

References

- AllTrials. 2013. *The AllTrials campaign calls for all past and present clinical trials to be registered and their results reported*. <http://www.alltrials.net/all-trials> (accessed October 1, 2013).
- ADNI (Alzheimers Disease Neuroimaging Initiative). 2013. *ADNI overview*. <http://www.adni-info.org/Scientists/ADNIOverview.aspx> (accessed October 12, 2013).
- Article 29 Data Protection Working Party. 2013. *Opinion 03/2013 on purpose limitation, 00569/13/EN WP 203, at 29 (2 April 2013)*. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf (accessed January 3, 2014).
- Beauchamp, T. L., and J. F. Childress. 2009. *Principles of biomedical ethics*. 6th ed. New York: Oxford University Press.
- Benitez, K., and B. Malin. 2010. Evaluating re-identification risks with respect to the HIPAA privacy rule. *Journal of the American Medical Informatics Association* 17(2):169-177.
- Bioethics Commission (Presidential Commission for the Study of Bioethical Issues). 2011. *“Ethically Impossible” STD Research in Guatemala from 1946 to 1948*. <http://bioethics.gov/node/654> (accessed November 6, 2013).
- Boulton, G., M. Rawlins, P. Vallance, and M. Walport. 2011. Science as a public enterprise: The case for open data. *Lancet* 377(9778):1633-1635.
- Califf, R. 2013. *Clinical trial data and challenges to data sharing: Current and future*. Presentation at Committee on Strategies for Responsible Sharing of Clinical Trial Data: Meeting 1, Open Sessions: Institute of Medicine Workshop, October 23, Washington, DC.
- Childress, J. F., E. M. Meslin, and H. T. Shapiro. 2005. *Belmont revisited: Ethical principles for research with human subjects*. Washington, DC: Georgetown University Press.

- CIOMS (Council for International Organizations of Medical Sciences). 2002. *International ethical guidelines for biomedical research involving human subjects*. <http://www.cioms.ch/index.php/publications/printable-publications?task=view&id=48&catid=57> (accessed November 5, 2013).
- ClinicalTrials.gov. 2012. *Glossary of commons site terms*. <http://clinicaltrials.gov/ct2/about-studies/glossary#I> (accessed November 5, 2013).
- Cochrane Collaboration. 2002. *The Cochrane Collaboration open learning material: Publication bias*. <http://www.cochrane-net.org/openlearning/html/mod15-2.htm> (accessed November 20, 2013).
- CTSA (Clinical and Translational Science Awards) Community Engagement Key Function Committee Task Force. 2011. *Principles of community engagement: Second edition*. Washington, DC: NIH.
- Dickersin, K. 2013. *Principles of responsible data sharing*. Paper read at Committee on Strategies for Responsible Sharing of Clinical Trial Data: Meeting 1, Open Sessions: Institute of Medicine Workshop, October 23, Washington, DC.
- Doshi, P., T. Jefferson, and C. Del Mar. 2012. The imperative to share clinical study reports: Recommendations from the Tamiflu experience. *PLoS Medicine* 9(4):e1001201.
- Doshi, P., K. Dickersin, D. Healy, S. S. Vedula, and T. Jefferson. 2013. Restoring invisible and abandoned trials: A call for people to publish the findings. *BMJ (Online)* 346(7913).
- Drazen, J. M. 2002. Who owns the data in clinical trial? *Science and Engineering Ethics* 8(3):407-411.
- Eichler, H. G., F. Pétavy, F. Pignatti, and G. Rasi G. 2013. Access to patient-level trial data: A boon to drug developers. *New England Journal of Medicine* 369(17):1577-1579.
- El Emam, K., F. Dankar, A. Neisa, and E. Jonker. 2013. Evaluating the risk of patient re-identification from adverse drug event reports. *BMC Medical Informatics and Decision Making* 13(1):114.
- Emanuel, E. J., D. Wendler, and C. Grady. 2008. An ethical framework for biomedical research. In *Oxford textbook of research ethics*, edited by E. J. Emanuel, C. Grady, R. A. Crouch, R. K. Lie, F. G. Miller and D. Wendler. New York: Oxford University Press. Pp. 123-135.
- EMA (European Medicines Agency). 2013. *Publication and access to clinical-trial data*. http://www.ema.europa.eu/docs/en_GB/document_library/Other/2013/06/WC500144730.pdf (accessed December 1, 2013).
- European Commission. 2013. *Protection of personal data*. <http://ec.europa.eu/justice/data-protection> (accessed November 22, 2013).
- Evans, B. J. 2011. Much ado about data ownership. *Harvard Journal of Law and Technology* 25. <http://ssrn.com/abstract=1857986> (accessed December 6, 2013).
- Faden, R., N. Kass, D. Whicher, W. Stewart, and S. Tunis. 2013. Ethics and informed consent for comparative effectiveness research with prospective electronic clinical data. *Medical Care* 51(8 Suppl 3):S53-S57.

- FDA (U.S. Food and Drug Administration). 1996. *Guideline for industry: Structure and content of clinical study reports*. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073113.pdf> (accessed November 7, 2013).
- FDA. 1999. *Guidance for industry: Submission of abbreviated reports and synopses in support of marketing applications*. <http://www.fda.gov/downloads/Drugs/Guidances/ucm072053.pdf> (accessed November 7, 2013).
- FDA. 2013. Availability of masked and deidentified non-summary safety and efficacy data; request for comments. *78 Federal Register* 33421 (June 4), p. 60290.
- Getz, K., Z. Hallinan, D. Simmons, M. J. Brickman, Z. Jumadilova, L. Pauer, M. Wilenzick, and B. Morrison. 2012. Meeting the obligation to communicate clinical trial results to study volunteers. *Expert Reviews in Clinical Pharmacology* 5(2):149-156.
- Gibson, L. 2004. GlaxoSmithKline to publish clinical trials after U.S. lawsuit. *BMJ* 328(7455):1513.
- Godlee, F., and T. Groves. 2012. The new BMJ policy on sharing data from drug and device trials. *BMJ (Online)* 345(7884):1-3.
- Goldacre, B. 2012. *Bad pharma: How medicine is broken and how we can fix it*. Revised and updated edition. London: Faber & Faber.
- Gordon, D., W. Taddei-Peters, A. Mascette, M. Antman, P. G. Kaufmann, and M. S. Lauer. 2013. Publication of trials funded by the National Heart, Lung, and Blood Institute. *New England Journal of Medicine* 369(20):1926-1934.
- Gross, C. P., H. M. Krumholz, G. Van Wye, E. J. Emanuel, D. Wendler. 2006. Does random treatment assignment cause harm to research participants? *PLoS Medicine* 3(6):e188.
- Gymrek, M., A. L. McGuire, D. Golan, E. Halperin, and Y. Erlich. 2013. Identifying personal genomes by surname inference. *Science* 339(6117):321-324.
- Harvard University MRCT (Multi-Regional Clinical Trials Center Data Sharing Workgroup). 2013. Paper presented at Issues and Case Studies in Clinical Trial Data Sharing: Lessons and Solutions Conference, May 17, Cambridge, MA.
- ICH (International Conference on Harmonisation). 1995. *Structure and content of clinical study reports (E3)*. <http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/structure-and-content-of-clinical-study-reports.html> (accessed November 7, 2013).
- ICH. 1996. *Guideline for good clinical practice E6(R1)*. <http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/good-clinical-practice.html> (accessed November 5, 2013).
- Immport (Immunology Database and Analysis Portal). 2013. *Immport: Bioinformatics for the future of immunology*. <https://immport.niaid.nih.gov/immportWeb/home/home.do?loginType=full> (accessed November 15, 2013).
- Immune Tolerance Network. 2013. *Trial share: Clinical trials research portal*. <https://www.itntrialshare.org> (accessed September 25, 2013).

- IOM (Institute of Medicine). 2002. *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: The National Academies Press.
- IOM. 2013. *Sharing clinical research data: Workshop summary*. Washington, DC: The National Academies Press.
- Jiang, X. P., A. D. P. Sarwate, and L. P. M. D. Ohno-Machado. 2013. Privacy technology to support data sharing for comparative effectiveness research: A systematic review. *Medical Care* 51(8):S58-S65.
- Jones, J. H. 2008. Tuskegee syphilis experiment. In *The Oxford textbook of clinical research ethics*, edited by E. J. Emanuel, C. Grady, R. A. Crouch, R. K. Lie, F. G. Miller and D. Wendler. New York: Oxford University Press. Pp. 86-96.
- Kaiser, L., C. Wat, T. Mills, P. Mahoney, P. Ward, and F. Hayden. 2003. Impact of oseltamivir treatment on influenza-related lower respiratory tract complications and hospitalizations. *Archives of Internal Medicine* 163(14): 1667-1672.
- Krumholz, H. M., and J. S. Ross. 2011. A model for dissemination and independent analysis of industry data. *JAMA* 306(14):1593-1594.
- Kuntz, R. E. 2013. The changing structure of industry-sponsored clinical research: Pioneering data sharing and transparency. *Annals of Internal Medicine* 158(12):914-915.
- Laine, C., S. N. Goodman, M. E. Griswold, and H. C. Sox. 2007. Reproducible research: Moving toward research the public can really trust. *Annals of Internal Medicine* 146(6):450-453.
- Laine, C., E. Guallar, C. Mulrow, D. B. Taichman, J. E. Cornell, D. Cotton, M. E. Griswold, A. R. Localio, A. R. Meibohm, C. B. Stack, S. V. Williams, and S. N. Goodman. 2013. Closing in on the truth about recombinant human bone morphogenetic protein-2: Evidence synthesis, data sharing, peer review, and reproducible research. *Annals of Internal Medicine* 158(12):916-918.
- Lidz, C. W., P. S. Appelbaum, T. Grisso, and M. Renaud. 2004. Therapeutic misconception and the appreciation of risks in clinical trials. *Social Science & Medicine* 58(9):1689-1697.
- Loder, E. 2013. Sharing data from clinical trials: Where we are and what lies ahead. *BMJ* 347:f4794.
- McGauran, N., B. Wieseler, J. Kreis, Y.-B. Schuler, H. Kolsch, and T. Kaiser. 2010. Reporting bias in medical research—a narrative review. *Trials* 11(1):37.
- McGraw, D., K. Rosati, and B. Evans. 2012. A policy framework for public health uses of electronic health data. *Pharmacoepidemiology and Drug Safety* 21:18-22.
- Mello, M. 2013. *Preparing for responsible sharing of clinical trial data*. Presentation at Committee on Strategies for Responsible Sharing of Clinical Trial Data: Meeting 1, Open Sessions: Institute of Medicine Workshop, October 23, Washington, DC.

- Mello, M. M., J. K. Francer, M. Wilenzick, P. Teden, B. E. Bierer, and M. Barnes. 2013. Preparing for responsible sharing of clinical trial data. *New England Journal of Medicine* 369(17):1651-1658.
- National Commission (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research). 1979. *The Belmont report: Ethical principles and guidelines for the protection of human subjects of biomedical and behavioral research*. <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html> (accessed December 1, 2013).
- NIH (National Institutes of Health). 2003. *Final NIH statement on sharing research data*. http://grants.nih.gov/grants/policy/data_sharing (accessed January 3, 2014).
- NHLBI (National Heart, Lung, and Blood Institute). 2007. *Biologic specimen and data repository information coordinating center*. <https://biolincc.nhlbi.nih.gov/home> (accessed November 16, 2013).
- Nisen, P., and F. Rockhold. 2013. Access to patient-level data from GlaxoSmithKline clinical trials. *New England Journal of Medicine* 369(5):475-478.
- Nissen, S. E., and K. Wolski. 2007. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine* 356(24):2457-2471.
- PhRMA (Pharmaceutical Research and Manufacturers of America) and EFPIA (European Federation of Pharmaceutical Industries and Associations). 2013. *Principles for responsible clinical trial data sharing: Our commitment to patients and researchers*. [http://phrma.org/sites/default/files/pdf/PhRMA PrinciplesForResponsibleClinicalTrialDataSharing.pdf](http://phrma.org/sites/default/files/pdf/PhRMA%20PrinciplesForResponsibleClinicalTrialDataSharing.pdf) (accessed September 20, 2013).
- Rathi, V., K. Dzara, C. P. Gross, I. Hrynaskiewicz, S. Joffe, H. M. Krumholz, K. M. Strait, and J.S. Ross. 2012. Sharing of clinical trial data among trialists: a cross sectional survey. *BMJ* 345:e7570.
- Rawlins, M. 2012. Saints or sinners. *Lancet* 380(9852):1462-1463.
- Ross, J. S., R. Lehman, and C. P. Gross. 2012. The importance of clinical trial data sharing: Toward more open science. *Circulation: Cardiovascular Quality and Outcomes* 5(2):238-240.
- Royal Society. 2012. *Science as an open enterprise*. London: Royal Society.
- Silverstein, F. E., G. Faich, J. L. Goldstein, and et al. 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.
- Simmonds, M. C., J. V. E. Brown, M. K. Heirs, J. P. T. Higgins, R. J. Mannion, M. A. Rodgers, and L. A. Stewart. 2013. Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual-participant data. *Annals of Internal Medicine* 158(12):877-889.

- Spertus, J. A. 2012. The double-edged sword of open access to research data. *Circulation: Cardiovascular Quality and Outcomes* 5(2):143-144.
- Stiles, P. G., and J. Petrila. 2011. Research and confidentiality: Legal issues and risk management strategies. *Psychology, Public Policy, and Law* 17(3):333-356.
- Teden, P. 2013. *Types of clinical trial data to be shared*. Presentation at Committee on Strategies for Responsible Sharing of Clinical Trial Data: Meeting 1, Open Sessions: Institute of Medicine Workshop, October 23, Washington, DC.
- Terry, S. F., and P. F. Terry. 2011. Power to the people: Participant ownership of clinical trial data. *Science Translational Medicine* 3(69):69cm63.
- Weijer, C., and E. J. Emanuel. 2000. Protecting communities in biomedical research. *Science* 289(5482):1142-1144.
- Weil, D., M. Graham, and A. Fung. 2013. Science and regulation. Targeting transparency. *Science* 340(6139):1410-1411.
- Wellcome Trust. 2011. *Sharing research data to improve public health: Full joint statement by funders of health research*. <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm> (accessed October 15, 2013).
- Wendler, D., R. Kington, J. Madans, G. Van Wye, H. Christ-Schmidt, L. A. Pratt, O. W. Brawley, C. P. Gross, and E. Emanuel. 2006. Are racial and ethnic minorities less willing to participate in health research? *PLoS Medicine* 3(2):e19.
- Wertheimer, A. 2008. Exploitation in clinical research. In *The Oxford textbook of clinical research ethics*, edited by E. J. Emanuel, C. Grady, R. A. Crouch, R. K. Lie, F. G. Miller, and D. Wendler. New York: Oxford University Press. Pp. 201-210.
- White House. 2011. *The Obama administration's commitment to open government: A status report*. http://www.whitehouse.gov/sites/default/files/opengov_report.pdf (accessed October 6, 2013).
- WMA (World Medical Association). 2013. *World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects*. <http://www.wma.net/en/30publications/10policies/b3/index.html> (accessed November 5, 2013).
- YODA (Yale Open Data Access). 2013. Yale University Open Data Access (YODA) project policy for the public availability of rhBMP-2 clinical trial data. New Haven, CT: Yale University Center for Outcomes Research and Evaluation.

A

Statement of Task

An ad hoc committee of the Institute of Medicine will conduct a study to develop guiding principles and a framework (activities and strategies) for the responsible sharing of clinical trial data. For the purposes of the study, the scope will be limited to interventional clinical trials and “data sharing” will include the responsible entity (data generator) making the data available via open or restricted access, or exchanged among parties. For the purposes of this study, data generator will include industry sponsors, data repositories, and researchers conducting clinical trials. Specifically, the committee will

- Articulate guiding principles that underpin the responsible sharing of clinical trial data.
- Describe a selected set of data and data sharing activities, including, but not limited to
 - types of data (e.g., summary, participant);
 - provider(s) and recipient(s) of shared data; and
 - whether and when data are disclosed publicly, with or without restrictions, or exchanged privately among parties.
- For each data sharing activity, the committee will
 - identify key benefits of sharing and risks of not sharing to research sponsors and investigators, study participants, regulatory agencies, patient groups, and the public.
 - address key challenges and risks of sharing (e.g., resource constraints, implementation, disincentives in the academic research model, changing norms, protection of human subjects and patient privacy, IP [intellectual property]/legal issues, preservation of scientific standards and data quality).

- Outline strategies and suggest practical approaches to facilitate responsible data sharing.
- Make recommendations to enhance responsible sharing of clinical trial data. The committee will identify guiding principles and characteristics for the optimal infrastructure and governance for sharing clinical trial data, taking into consideration a variety of approaches (e.g., a distributed/federated data system).

In developing the principles and framework and in defining the rights, responsibilities, and limitations underpinning the responsible sharing of clinical trial data, the committee will take into account the benefits of data sharing, the potential adverse consequences of both sharing and not sharing data, and the landscape of regulations and policies under which data-sharing occurs. Focused consideration will also be given to the ethical standards and to integrating core principles and values, including privacy. The committee is not expected to develop or define specific technical data standards.

A framework for discussion will be released for public comment, which will include tentative findings regarding (a) guiding principles and (b) a selected set of data sharing activities. Based on the public comments received and further deliberations, the committee will prepare a final report with its findings and recommendations.

B

Committee Biographies

Bernard Lo, M.D. (*Chair*), is currently president of The Greenwall Foundation. Previously, Dr. Lo was professor of medicine and director of the Program in Medical Ethics at the University of California, San Francisco (UCSF). Currently, he is co-chair of the Standards Working Group of the California Institute of Regenerative Medicine, which recommends regulations for stem cell research funded by the state of California. Dr. Lo serves on the board of directors of the Association for the Accreditation of Human Research Protection Programs and on the Medical Advisory Panel of Blue Cross Blue Shield. Formerly, he was a member of the National Bioethics Advisory Commission under President Clinton, the National Institutes of Health (NIH) Recombinant DNA Advisory Committee and the Ethics Subcommittee, and the Advisory Committee to the director of the Centers for Disease Control and Prevention. He served on a number of Data and Safety Monitoring Committees at NIH for HIV prevention and treatments, diabetes prevention, and oxygen treatment in chronic obstructive pulmonary disease. A member of the Institute of Medicine (IOM), Dr. Lo served on the IOM Council and chaired the Board on Health Sciences Policy. He chaired IOM committees on conflicts of interest in medicine and on confidentiality in health services research and has been a member of several other IOM committees. He currently is a member of the Board on Life Sciences of the National Academy of Sciences. Dr. Lo and his colleagues have published about 200 peer-reviewed articles on ethical issues concerning decision making near the end of life, stem cell research, research with human participants and its oversight, the doctor-patient relationship, conflicts of interest, HIV infection, and public health. With colleagues on the UCSF Stem Cell Research Oversight

Committee, he has written articles on ethical issues in the procurement of embryos for research, oversight of stem cell lines derived in other institutions, informed consent for future research, and prohibiting the use of induced pluripotent stem cells for reproductive cloning. Dr. Lo is the author of *Resolving Ethical Dilemmas: A Guide for Clinicians* (5th ed., 2013) and of *Ethical Issues in Clinical Research* (2010). At UCSF, he directed medical student teaching in ethics, chaired the hospital ethics committee, and served as an attending physician on the medicine inpatient service. He was co-director of the Policy and Ethics Core of the Center for AIDS Prevention Studies. He continues to serve as the primary care physician for a panel of general internal medicine patients.

Timothy Coetzee, Ph.D., is Chief Advocacy, Services, and Research Officer of the National Multiple Sclerosis (MS) Society and is responsible for the society's research program, which funds more than 375 academic and commercial research projects around the world. Most recently, he served as president of Fast Forward, a venture philanthropy project of the National Multiple Sclerosis Society, where he was responsible for the society's strategic funding of biotechnology and pharmaceutical companies, as well as partnerships with the financial and business communities. Prior to Fast Forward, Dr. Coetzee led the society's translational research initiatives on nervous system repair and protection in multiple sclerosis. He is a member of the Institute of Medicine's Forum on Neuroscience and Nervous System Disorders and serves on the board of directors of the American Society of Experimental Neurotherapeutics. He also chairs the Integration Panel for the MS Research Program of the Department of Defense's Congressionally Directed Medical Research Program. Dr. Coetzee received his Ph.D. in molecular biology from Albany Medical College in 1993 and has since been involved in the field of multiple sclerosis research. He has been with the National MS Society since fall 2000.

David L. DeMets, Ph.D., is currently the Max Halperin Professor of Biostatistics and founder/former chair of the Department of Biostatistics and Medical Informatics at the University of Wisconsin (UW)–Madison. Since receiving his Ph.D. in 1970 from the University of Minnesota, he has been very active in the design, conduct, and analysis of clinical trials in several disease areas. Following a postdoctoral appointment at the National Institutes of Health (NIH) (1970-1972), he spent 10 years (1972-1982) at the National Heart, Lung, and Blood Institute at NIH,

where he became chief of the Biostatistics Research Branch. He has co-authored four texts on the topic of clinical trial design, interim monitoring, and analyses. Dr. DeMets is a recognized international leader in statistical research and methods for the analysis of clinical trials. He has collaborated in the development of statistical methods for the sequential analysis of outcome data and the design of clinical trials. He has extensive national and international clinical trial experience and has served on and chaired numerous NIH and industry-sponsored Data Safety and Monitoring Committees for clinical trials in diverse disciplines. He served on the Board of Scientific Counselors of the National Cancer Institute and Board of Directors of the American Statistical Association, as well as serving as president of the Society for Clinical Trials and president of the Eastern North American Region of the Biometric Society. He is a fellow of the American Statistical Association, the International Statistics Institute, the Society of Clinical Trials, the American Medical Informatics Association, and the American Association for the Advancement of Science. Dr. DeMets has served on the Human Subjects Committee (1982-1987) and on the following UW committees since 1990: Ad Hoc Committee on Conflict of Interest (1992-1993); Tenure Track Promotions Committee (1995-1998); and Biomedical Industry Relations Committee (1996-1998). In addition, he has served on the search committees for the UW Cancer Center Director (1994-1997); Associate Dean for Research (Chair) (1995-1996); Associate Dean for Administration (Chair) (1995-1996); Section of Cardiology Chief (1998-1999); and Preventive Medicine Chair (1999-2000). His graduate school committees include the Committee on Training Research Ethics (Chair) (1993-1996; 1998-2000); the Health Sciences Information Technology Committee (1999-2003); and the University Conflict of Interest Committee (2010-present).

Jeffrey M. Drazen, M.D., joined the *New England Journal of Medicine* (NEJM) as editor-in-chief in July 2000. At NEJM, Dr. Drazen's responsibilities include oversight of all editorial content and policies. His editorial background includes service as an associate editor or editorial board member for the *Journal of Clinical Investigation*, the *American Journal of Respiratory Cell and Molecular Biology*, and the *American Journal of Medicine*. A specialist in pulmonology, Dr. Drazen maintains an active research program. Dr. Drazen has published more than 300 articles on topics such as lung physiology and the mechanisms involved in asthma. In 1999, he delivered the Amberson Lecture, the major

research address at the annual meeting of the American Thoracic Society. In 2000, he received the Chadwick Medal from the Massachusetts Thoracic Society for his contributions to the study of lung disease. Dr. Drazen is the Distinguished Parker B. Francis Professor of Medicine at Harvard Medical School, professor of physiology at the Harvard School of Public Health, and a senior physician at Brigham and Women's Hospital. In 2003, he was elected as a member of the Institute of Medicine. Dr. Drazen has served on numerous committees for the National Institutes of Health, including the Respiratory and Applied Physiology Study Section; the Lung Biology and Pathology Study Section; the Pulmonary Disease Advisory Council; the National Heart, Lung, and Blood Institute (NHLBI) Advisory Council; the Public Access Working Group; and the NHLBI Division of Lung Disease Executive Planning Committee. He has also served on the Department of Veterans Affairs National Research Advisory Committee. He currently serves on the Global Initiative for Asthma Science Committee and the World Health Organization's Scientific Advisory Group on Clinical Trials Registration and co-chairs the Institute of Medicine's Forum on Drug Discovery, Development, and Translation. Dr. Drazen earned his bachelor's degree and graduated summa cum laude from Tufts University. He received his medical degree from Harvard Medical School and completed his internship and residency at Peter Bent Brigham Hospital in Boston. Dr. Drazen has received honorary degrees from the University of Ferrara, Italy, and the National and Kapodistrian University of Athens, Greece.

Steven N. Goodman, M.D., M.H.S., Ph.D., is associate dean for clinical and translational research and professor of medicine and health policy and research at Stanford University School of Medicine. Before joining Stanford in 2011, Dr. Goodman was professor of oncology in the division of biostatistics of the Johns Hopkins Kimmel Cancer Center, with appointments in the departments of pediatrics, biostatistics, and epidemiology in the Johns Hopkins Schools of Medicine and Public Health. He was also on the core faculties of the Johns Hopkins Center for Clinical Trials, Berman Bioethics Institute, Graduate Training Program in Clinical Investigation, and co-directed the epidemiology doctoral program. He is the editor of *Clinical Trials: Journal of the Society for Clinical Trials* and has been the statistical and associate editor of *the Annals of Internal Medicine* since 1987. He served on the Institute of Medicine's Committee to Review the Health Effects in Vietnam

Veterans of Exposure to Herbicides: Second Biennial Update; Committee to Review the Evidence Regarding the Link between Exposure to Agent Orange and Diabetes; Immunization Safety Review Committee; Committee on Alternatives to the Daubert Standards; and Committee on Treatment of Post-Traumatic Stress Disorder in Veterans. Dr. Goodman also served on the Surgeon General's committee to write the 2004 report on the *Health Consequences of Smoking*. He is a scientific advisor to the Medical Advisory Panel of the National Blue Cross Blue Shield Technology Evaluation Center and was appointed to the Methodology Committee of the Patient-Centered Outcomes Research Institute. Dr. Goodman received a B.A. from Harvard University and an M.D. from New York University, trained in pediatrics at Washington University in St. Louis, 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.

Patricia A. King, J.D., has expertise in the study of law, medicine, ethics, and public policy. She is an adjunct professor in the Department of Health Policy and Management, School of Hygiene and Public Health, at Johns Hopkins University. She is the co-author of *Cases and Materials on Law, Science and Medicine*. She teaches family law courses and offers a seminar in bioethics and the law. She is a member of the American Law Institute and the Institute of Medicine and a fellow of the Hastings Center. Her work in the field of bioethics has included service on the Health, Education and Welfare-Advisory Recombinant DNA Advisory Committee; the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research; the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research; and the Ethics, Legal and Social Issues Working Group of the Human Genome Project. Ms. King is a former member of the Harvard Corporation and a trustee emeritus of Wheaton College in Massachusetts. Her professional experience before joining the Law Center faculty in 1973 was primarily in the civil rights field; she was the deputy director of the Office of Civil Rights and a special assistant to the chairman of the Equal Employment Opportunity Commission. She also served as a deputy assistant attorney general in the Civil Division of the Department of Justice.

Trudie Lang, Ph.D., is a clinical trials research methodologist with specific expertise in capacity development and trial operations in low-resource settings. She currently leads the Global Health Network (GHN) at the University of Oxford, which is a forum to help clinical researchers with trial design, methods, interpretation of regulations, and general operations. GHN conducts methodology research to identify the real barriers and issues involved in noncommercial trials, with the aim of developing best-practice guidelines. Dr. Lang has worked in the field of clinical trials for 20 years and has experienced the benefit of working in the pharmaceutical industry, the World Health Organization, and in academia. Ten years managing trials in industry gave Dr. Lang strong capabilities in leading teams, regulations, managing complex projects, and conducting effective strategic planning and taught her the rigors of designing and operating clinical trials in varied settings. At Oxford, she has further developed her expertise in the design, operation, and methodology of running trials in developing countries. Dr. Lang set up a clinical trial facility in Kenya with a strong focus on developing local research skills and engagement. More recently, she devised and set up Global Health Trials (www.globalhealthtrials.org). This is an online, research-led facility to support and guide research teams and is being used by more than 85,000 researchers. This evolved into GHN (www.theglobalhealthnetwork.org), a virtual science park that hosts 15 international collaborations across varied diseases areas—all aiming to support research by sharing knowledge, research tools, and methods.

Deven McGraw, J.D., M.P.H., L.L.M., is the director of the Health Privacy Project at the Center for Democracy and Technology (CDT). The project is focused 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 (HIT) and electronic health information exchange to improve health care. She was one of three persons appointed by Kathleen Sebelius, the Secretary of the U.S. Department of Health & Human Services (HHS), to serve on the Health Information Technology Policy Committee, a federal advisory committee established in the American Recovery and Reinvestment Act of 2009. She has chaired that committee's Privacy and Security Workgroup since 2010 and also serves on its Information Exchange and Meaningful Use Workgroups. In addition, she served on two key workgroups of the American Health Information Community, the federal advisory body established by HHS

in the George W. Bush Administration to develop recommendations on how to facilitate use of HIT to improve health. She also serves on the executive committee of the Clinical Trials Transformation Initiative, the steering committee of the Electronic Data Methods Forum, and the leadership council of the eHealth 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, College Park. 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 the Johns Hopkins School of Hygiene and Public Health.

Elizabeth G. Nabel, M.D., has served as president of Harvard University–affiliated Brigham and Women's Hospital (BWH) since 2010. A cardiologist and distinguished biomedical researcher, Dr. Nabel is professor of medicine at Harvard Medical School. She brings a unique perspective to health care based on her experience as a physician, research scientist, academic medicine leader, and wellness advocate. At BWH, she led development of a comprehensive strategic plan that defines a new model of medicine characterized by cross-disciplinary collaboration, patient-inclusive care, and innovation. Initiatives include a new translational medical facility, patient-centered intensive care unit care, and a \$1 billion campaign to advance innovation, patient care, and community health. Dr. Nabel has a long record of advocacy for health and broadening access to care. As director of the National Heart, Lung, and Blood Institute (NHLBI) from 2005 to 2009, she leveraged the \$3 billion research portfolio to establish pioneering scientific programs in genomics, stem cells, and translational research. One of Dr. Nabel's signature advocacy efforts was the Red Dress Heart Truth campaign, which raises heart awareness in women through innovative partnerships. Throughout her career, Nabel has been a champion for global health. At NHLBI, she established Centers of Excellence in developing countries to

combat cardiovascular and lung diseases. At BWH, she helped create a national teaching hospital in Haiti and is advancing training for clinicians in underresourced countries. An accomplished physician-scientist, Dr. Nabel's work on the molecular genetics of cardiovascular diseases has produced 17 patents and more than 250 scientific publications. Her colleagues elected her to the American Academy of the Arts and Sciences and the Institute of Medicine, and she is a fellow of the American Association for the Advancement of Science. Dr. Nabel's honors include the Willem Einthoven Award from Leiden University in the Netherlands; two Distinguished Achievement Awards from the American Heart Association; and six honorary doctorates. A native of St. Paul, Minnesota, Dr. Nabel attended Weill Cornell Medical College and completed her cardiology training at BWH.

Arti K. Rai, J.D., Elvin R. Latty Professor of Law at Duke University School of Law, is an internationally recognized expert in intellectual property (IP) law, administrative law, and health policy. Ms. Rai has also taught at the Harvard University, Yale University, and University of Pennsylvania law schools. Ms. Rai's research on IP law and policy in biotechnology, pharmaceuticals, and software has been funded by the National Institutes of Health, the Kauffman Foundation, and the Woodrow Wilson Center. She has published more than 50 articles, essays, and book chapters on IP law, administrative law, and health policy. Her publications have appeared in both peer-reviewed journals and law reviews, including *Science*, the *New England Journal of Medicine*, the *Journal of Legal Studies*, *Nature Biotechnology*, and the *Columbia*, *Georgetown*, and *Northwestern* university law reviews. She is the editor of *Intellectual Property Law and Biotechnology: Critical Concepts* (Edward Elgar, 2011), the co-author of a 2012 Kauffman Foundation monograph on cost-effective health care innovation, and the co-author of a casebook on law and the mental health system. From 2009 to 2010, Ms. Rai served as the administrator of the Office of External Affairs at the U.S. Patent and Trademark Office (USPTO). As external affairs administrator, she led policy analysis of the patent reform legislation that ultimately became the America Invents Act and worked to establish USPTO's Office of the Chief Economist. Prior to that time, she had served on President-Elect Obama's transition team reviewing USPTO. Prior to entering academia, Ms. Rai clerked for the Honorable Marilyn Hall Patel of the U.S. District Court for the Northern District of California; was a litigation associate at Jenner & Block (doing patent

litigation as well as other litigation); and was a litigator at the Federal Programs Branch of the U.S. Department of Justice's Civil Division. Ms. Rai regularly testifies before Congress and relevant administrative bodies on IP law and policy issues and advises federal agencies on IP policy issues raised by the research that they fund. She is a member of the National Advisory Council for Human Genome Research and of an Expert Advisory Council to the Defense Advanced Projects Research Agency. She is a public member of the Administrative Conference of the United States, a member of the American Law Institute, and co-chair of the IP Committee of the Administrative Law Section of the American Bar Association. In 2011, Ms. Rai won the World Technology Network Award for Law. She graduated from Harvard College, magna cum laude, with a B.A. in biochemistry and history, attended Harvard Medical School for the 1987-1988 academic year, and received her J.D., cum laude, from Harvard Law School in 1991.

Ida Sim, Ph.D., is professor of medicine, co-director of biomedical informatics at the University of California, San Francisco, Clinical and Translational Sciences Institute, and co-founder of Open mHealth, a nonprofit organization that is breaking down barriers to mobile health app and data integration through an open software architecture. Her primary research work is on knowledge-based technologies for evidence-based practice, especially in the ontological representation of clinical trials for data sharing and scientific computation. In 2005, Dr. Sim was the founding project coordinator of the World Health Organization's International Clinical Trials Registry Platform, where she led the establishment of the first global policy on clinical trial registration, including the development of the Trial Registration Data Set, the common 20-item dataset that all registers worldwide adhere to. She has also published on clinical trial reporting bias, new models of scientific e-publication of clinical research, and other policies and practices of trial reporting and registration. In other work, Dr. Sim co-founded Open mHealth (<http://openmhealth.org>), a nonprofit to build an open-source way for the mobile health (mHealth) ecosystem to share data, methods, and software to maximize the potential of mHealth to improve individual and population health. Dr. Sim was a member of the National Research Council Committee on Computational Technology for Effective Health Care. She serves on the editorial board of the *Journal of Biomedical Informatics* and is on the advisory board for *PLoS One*. She is a recipient of the United States Presidential Early Career Award for Scientists and

Engineers, a fellow of the American College of Medical Informatics, and a member of the American Society for Clinical Investigation. She is also a practicing primary care physician.

Sharon Terry, M.A., is president and CEO of Genetic Alliance, a network of more than 10,000 organizations, of which 1,200 are disease advocacy organizations. Genetic Alliance enables individuals, families, and communities to reclaim their health and become full participants in translational research and services. She is the founding CEO of PXE International, a research advocacy organization for the genetic condition pseudoxanthoma elasticum (PXE). As co-discoverer of the gene associated with PXE, she holds the patent for ABCC6 to act as its steward and has assigned her rights to the foundation. She developed a diagnostic test and conducts clinical trials. She is the author of 120 peer-reviewed papers, of which 30 are PXE clinical studies. Ms. Terry is also a co-founder of the Genetic Alliance Registry and Biobank. 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) Science and Policy Board, the IOM Roundtable on Translating Genomic-Based Research for Health, the PubMed Central National Advisory Committee, and the International Rare Disease Research Consortium Executive Committee, and as founding president of EspeRare Foundation. 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. In 2012, she became an honorary professor of Hebei United University in Tangshan, China, and also received the Facing Our Risk of Cancer Empowered Spirit of Empowerment Advocacy Award. She was named one of the Food and Drug Administration's "30 Heroes for the Thirtieth Anniversary of the Orphan Drug Act" in 2013. She is an Ashoka Fellow.

Joanne Waldstreicher, M.D., is chief medical officer, Johnson & Johnson. In this role, she has oversight for epidemiology and safety of all

Johnson & Johnson products worldwide across all sectors, including pharmaceuticals, devices, and consumer products. In addition, she also plays a leadership role for internal and external partnerships and collaborations, including development of the corporate science, technology, regulatory, and government policies. Dr. Waldstreicher also chairs the Pharmaceuticals R&D Development Committee, which reviews all late-stage development programs in the pharmaceutical pipeline. Dr. Waldstreicher's prior role was as chief medical officer of the pharmaceutical sector and head of Asia Pacific Medical Sciences, responsible for medical safety, Asia research and development, epidemiology, adaptive design strategy, pediatrics, established products, and small molecule clinical pharmacology. Prior to that role, she was head of Global Drug Development for the Johnson & Johnson Pharmaceutical Research & Development, LLC (J&JPRD) CNS/Internal Medicine business unit. In this role, she was responsible for late-stage development spanning the areas of psychiatry, neurology, pain, infectious disease, cardiovascular medicine, urology, metabolism, and other emerging areas. Prior to joining J&JPRD in 2002, Dr. Waldstreicher was head of the endocrinology and metabolism clinical research group at Merck Research Laboratories and responsible for overseeing clinical development programs in atherosclerosis, obesity, diabetes, urology, dermatology, and oncology. During that time, she received numerous distinctions, including the Merck Research Laboratory Key Innovator Award. Dr. Waldstreicher received the Jonas Salk and Belle Zeller scholarships from the City University of New York and graduated summa cum laude from Brooklyn College and cum laude from Harvard Medical School. She completed her fellowship in endocrinology and metabolism at Massachusetts General Hospital, has won numerous awards and scholarships, and has authored numerous papers and abstracts.

C

Figure 1: Data Flow from Participant to Analyzed Data and Reporting

A. Raw Data. All clinical trial data originate from patients and healthy volunteers who participate in studies that are carried out according to a detailed research protocol. This protocol is approved by an institutional review board and explained to participants through the informed consent process. Depending on the study under consideration, demographics, clinical outcomes data, and other appropriate raw source information are entered into case report forms. Some data (e.g., imaging studies) are interpreted by study investigators, and these interpretations are entered into the database—a process referred to as abstraction. The data are then coded to meet the study guidelines (e.g., men may be coded as “1” and women as “0”); the coded data are then entered into the case report forms. In addition, narrative data from the case reports are also transcribed into the database. The data are then reviewed (i.e., cleaned) to be sure that entries make sense and are internally consistent. The data are then abstracted, coded, transcribed, and cleaned as appropriate.

B. Cleaned Analyzable Dataset. Once the database is cleaned and all queries are resolved, the database, which consists of both individual participant data and computed/summary-level data, is then analyzable. It is called analyzable because a very large percentage is never used. The next step is to lock the database so that no further changes may be made and the data may be unblinded. However, the cleaned analyzable dataset in its unlocked condition has the potential for subsequent use, because it could be re-analyzed at later time points with the addition of data (e.g., when 1-year, 3-year, and 10-year outcomes measures are added).

C. Cleaned and Locked Analyzable Dataset. The final cleaned and locked analyzable dataset consists of different components (participant characteristics and primary outcome, prespecified secondary and tertiary outcomes, adverse events data and exploratory data). A statistical analysis may involve a composite outcome using any of the various components. In addition, when data are missing, values may be imputed using this dataset. Results are derived from data in the cleaned and locked analyzable dataset, which have undergone statistical analysis. Analyses that were prespecified in the Statistical Analysis Plan form the basis for the Clinical Study Report (CSR) (a detailed analysis of the study efficacy data and the complete adverse event data). The CSR and the supporting cleaned dataset are available to regulators (e.g., the Food and Drug Administration, the European Medicines Agency) and to other data users as appropriate (e.g., ministries of health). Journal articles generally represent slices of the data that make a coherent intellectual whole. For example, the “lead article” usually describes the data on the primary efficacy outcomes, key secondary outcomes, and the relevant adverse event data. Subsequent articles often focus on different aspects of the secondary, tertiary, or exploratory outcomes. Investigators can also use parts of the analyzable dataset to prepare analyses for presentations, for data exploration, and for hypothesis generation. A biostatistics best practice is to freeze a copy of whatever data were used in an analysis so the results can later be repeated if necessary. It would also be desirable to store the code used in the analysis (i.e., the computer program), especially for any derived data.