Sharing Clinical Trial Data: Challenges and a Way Forward

November 18–19, 2019

National Academy of Sciences Building, Great Hall
2101 Constitution Ave. NW, Washington, DC 20418
Washington, DC

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Sharing Clinical Trial Data: Challenges and a Way Forward – A Workshop

November 18-19, 2019

Briefing Book Table of Contents

Meeting Materials

Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk – A Consensus Study (Background Materials)

Literature

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Forum on Drug Discovery, Development, and Translation
Forum on Neuroscience and Nervous System Disorders
National Cancer Policy Forum
Roundtable on Genomics and Precision Health

Sharing Clinical Trial Data: Challenges and a Way Forward

A Workshop

November 18-19 2019 ▪ Washington, DC

Statement of Task

An ad hoc planning committee under the auspices of the National Academies of Sciences, Engineering, and Medicine, will plan and conduct a two-day public workshop to discuss advances, challenges, and opportunities in clinical trial data sharing efforts since release of the 2015 Institute of Medicine report, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk.* This workshop is co-sponsored by the Wellcome Trust.

The public workshop will feature invited presentations and discussions to:

- Consider the value and the potential risks and costs of sharing clinical trial data for key stakeholders, including clinical trialists, sponsors, primary and secondary researchers, and patients;
- Review the current landscape of clinical trial data sharing and reuse across public and private sectors (e.g. policies, platforms, collaborations, data sharing culture, published research output);
- Examine use cases and trends from across public and private sectors when it comes to success, failure, lessons learned, and value;
- Consider the perspectives and expectations of primary and secondary researchers, clinical trial participants, patient organizations, research sponsors (pharmaceutical companies and nonprofit organizations), journals, institutions, and federal agencies; and
- Discuss next step opportunities for stakeholders to better harmonize incentives, policy, data standards, and governance to encourage the sharing and reuse of clinical trial data.

The planning committee will organize the workshop, develop the agenda, select and invite speakers and discussants, and moderate or identify moderators for the discussions. A proceedings of the presentations and discussions at the workshop will be prepared by a designated rapporteur in accordance with institutional guidelines.

Planning Committee

Jeffrey Drazen (Co-Chair), New England Journal of Medicine

Joanne Waldstreicher (Co-Chair), Johnson & Johnson

Monica Bertagnolli, Harvard Medical School; Dana Farber Cancer Institute and Brigham and Women's Hospital

Timothy Coetzee, National Multiple Sclerosis Society

Patrick Cullinan, Bluebird Bio

Sonali Kochhar, Global Healthcare Consulting; University of Washington; Erasmus MC, University Medical Center

Bernard Lo, The Greenwall Foundation

Deven McGraw, Ciitizen

Dina Paltoo, National Library of Medicine, NIH

Liz Roberts, UCB Biosciences, Inc

Frank Rockhold, Duke University Medical Center

Sharon Terry, Genetic Alliance

Deborah Zarin, Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Forum on Drug Discovery, Development, and Translation
Forum on Neuroscience and Nervous System Disorders
National Cancer Policy Forum
Roundtable on Genomics and Precision Health

Sharing Clinical Trial Data: Challenges and a Way Forward

A Workshop

November 18 – 19, 2019
National Academy of Sciences Building, Great Hall
2101 Constitution Ave. NW, Washington, DC 20418

WORKSHOP OBJECTIVES:

Following release of the 2015 Institute of Medicine report, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*:

- Consider the value and potential risks/costs of sharing clinical trial data for key stakeholders, including clinical trialists, sponsors, primary and secondary researchers, and patients;
- Review the current landscape of clinical trial data sharing and reuse across public and private sectors (e.g. policies, platforms, collaborations, data sharing culture, published research output);
- Examine use cases and trends from across public and private sectors when it comes to success, failure, lessons learned, and value;
- Consider the perspectives and expectations of primary and secondary researchers, clinical trial participants, patient organizations, research sponsors (pharmaceutical companies and nonprofit organizations), journals, institutions, and federal agencies; and
- Discuss next step opportunities for stakeholders to better harmonize incentives, policy, data standards, and governance to encourage the sharing and reuse of clinical trial data.

DAY 1: November 18, 2019

8:00 a.m. Breakfast available in the Great Hall

8:30 a.m. **Welcome and opening remarks**

JOANNE WALDSTREICHER, *Workshop Co-Chair*
Chief Medical Officer
Johnson & Johnson

JEFFREY DRAZEN, *Workshop Co-Chair*
Group Editor
New England Journal of Medicine

SESSION I: THE CURRENT LANDSCAPE OF CLINICAL TRIAL DATA SHARING AND REUSE

Session Objectives:

- Consider the value and potential risks/costs of sharing clinical trial data for key stakeholders, including clinical trialists, sponsors, primary and secondary researchers, and patients;
- Review the current landscape of clinical trial data sharing and reuse across public and private sectors (e.g., policies, platforms, collaborations, data sharing culture);
- Examine case studies and trends from across public and private sectors when it comes to success, failure, lessons learned, and value.

9:00 a.m. **Landscape overview by session moderators**

BERNARD LO
President and CEO
The Greenwall Foundation

DEBORAH ZARIN
Director, Program for the Advancement of the Clinical Trials Enterprise
Multi-Regional Clinical Trials Center,
Brigham and Women's Hospital and Harvard

9:20 a.m. **Policies in practice: Lessons learned**

A researcher perspective
HARLAN KRUMHOLZ
Harold H. Hines, Jr. Professor of Medicine
Yale School of Medicine

A funder perspective
LYRIC JORGENSON
Deputy Director
Office of Science Policy, National Institutes of Health

An independent review panel perspective
SONALI KOCHHAR
Medical Director
Global Healthcare Consulting

A patient perspective
MOSES TAYLOR JR.
Participant
SPRINT Trial

10:20 a.m. **Data sharing platforms: Use cases**

Vivli
REBECCA LI
Executive Director
Vivli

The YODA Project
JOSEPH ROSS
Professor of Medicine and Public Health
Yale School of Medicine

SOAR
FRANK ROCKHOLD
Professor of Biostatistics and Bioinformatics
Duke University Medical Center

ClinicalStudyDataRequest.com
SCOTT SHAUNESSY
Chair
ClinicalStudyDataRequest.com

11:20 a.m. **BREAK**

11:50 a.m. **Panel discussion: Striking a balance between benefit/value and risk/cost – is the juice worth the squeeze?**

A data analyst perspective
DAVID DEMETS
Professor Emeritus, Department of Biostatistics and Medical Informatics
University of Wisconsin, Madison

A publisher perspective
JEFFREY DRAZEN
Group Editor
New England Journal of Medicine

A patient perspective
DEBORAH C. PEEL
Founder and President
Patient Privacy Rights

12:35 p.m. *Moderated audience discussion with the panel* (*25 mins*)

1:00 p.m. **BREAK** (Lunch available in the Great Hall)

SESSION II: KEY CHALLENGES IN CLINICAL TRIAL DATA SHARING AND REUSE

Session Objectives:

- Discuss key challenges to clinical trial data sharing and reuse by including use cases;
- Consider the perspectives and expectations of clinical trial participants, patients, research sponsors (pharmaceutical companies and nonprofit organizations), journals, academic researchers and institutions, institutional review boards, and federal agencies.

1:30 p.m. **Opening remarks by session moderators**

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

DINA PALTOO
Assistant Director, Policy Development
National Library of Medicine, National Institutes of Health

1:45 p.m. **Panel Discussion: Data interoperability and platform usability**

Use case: population data
ERNEST HAWK
Vice President and Division Head, Cancer Prevention and Population Sciences
MD Anderson

Use case: meta-analysis
TIANJING LI
Associate Professor
University of Colorado Denver

A platform perspective
BILL LOUV
President
Project Data Sphere

2:15 p.m. *Moderated audience discussion with the panel*

2:45 p.m. **Panel discussion: Infrastructure sustainability**

A funder perspective
GEORGINA HUMPHREYS
Clinical Data Sharing Manager
Wellcome Trust

A business perspective
PANDURANG KULKARNI
Chief Analytics Officer R&D
VP, Biometrics and Advanced Analytics
Eli Lilly and Company

A platform perspective
SEAN COADY
Program Officer
National Heart, Lung, and Blood Institute
National Institutes of Health

3:15 p.m. *Moderated audience discussion with the panel*

3:45 p.m. **BREAK**

4:15 p.m. **Panel discussion: Challenges and disincentives for sharing and using data**

Use case: Statistical replication

A researcher perspective
MATTHEW SYDES
Professor of Clinical Trials and Methodology
MRC Clinical Trials Unit at UCL
University College London

A sponsor perspective
RAMIN DARON
Vice President, Data Architecture and Technology
Takeda Pharmaceuticals

A patient perspective
SHARON TERRY
President and CEO
Genetic Alliance

4:45 p.m. *Moderated audience discussion with the panel*

DAY 1: REFLECTIONS

5:15 p.m. **Finding value as we move forward**

JOANNE WALDSTREICHER, *Workshop Co-Chair*
Chief Medical Officer
Johnson & Johnson

JEFFREY DRAZEN, *Workshop Co-Chair*
Group Editor
New England Journal of Medicine

5:30 p.m. **ADJOURN WORKSHOP DAY 1**

DAY 2: November 19, 2019

8:00 a.m. Breakfast Available in the Great Hall

8:30 a.m. **Welcome and overview of Day 1**
JOANNE WALDSTREICHER, *Workshop Co-Chair*
Chief Medical Officer
Johnson & Johnson

JEFFREY DRAZEN, *Workshop Co-Chair*
Group Editor
New England Journal of Medicine

SESSION III: FINDING VALUE IN SHARING CLINICAL TRIAL DATA

Session Objectives:

- Discuss next step opportunities for stakeholders to better align incentives, and implement policy, technology, and governance that encourage the sharing and reuse of clinical trial data;
  - Explore opportunities for overcoming technical barriers;
  - Explore opportunities for addressing cultural barriers; and
  - Discuss possible solutions/next steps going forward.

8:45 a.m. **Panel discussion: Overcoming usability and sustainability challenges**

Moderator
IDA SIM
Professor of Medicine
UCSF School of Medicine

A data user perspective
MARK HELFAND
Professor of Medicine, Medical Informatics, and Clinical Epidemiology
Oregon Health & Science University

A data generator/sharer perspective
MONICA BERTAGNOLLI
Professor of Surgery
Harvard Medical School

A platform perspective
REBECCA KUSH
Chief Scientific Officer
Elligo Health Research

9:30 a.m. *Moderated audience discussion with the panel*

10:00 a.m. **BREAK**

10:30 a.m. **Panel discussion: Looking forward: incentivizing data sharing and reuse**

Moderator and sponsor perspective
LIZ ROBERTS
Senior Director and Global Public Policy Lead
UCB Biosciences

A researcher perspective
MARTIN HO
Associate Director for Quantitative Patient Inputs & RWE
Center for Biologics Evaluation and Research
U.S. Food and Drug Administration

A researcher perspective
COLIN BAIGENT
Deputy Director, Clinical Trial Service Unit and Epidemiological Studies Unit
Oxford University

An institutional perspective
AMY NURNBERGER
Program Head, Data Management Services
Massachusetts Institute of Technology

A funder perspective
GEORGINA HUMPHREYS
Clinical Data Sharing Manager
Wellcome Trust

11:45 a.m. *Moderated audience discussion with the panel*

12:15 p.m. **Next Steps**

JOANNE WALDSTREICHER, *Workshop Co-Chair*
Chief Medical Officer
Johnson & Johnson

JEFFREY DRAZEN, *Workshop Co-Chair*
Group Editor
New England Journal of Medicine

12:30 p.m. **ADJOURN WORKSHOP DAY 2**

ABOUT THE FORUM

The Forum on Drug Discovery, Development, and Translation of the National Academies of Sciences, Engineering, and Medicine was created in 2005 by the Board on Health Sciences Policy to provide a unique platform for dialogue and collaboration among thought leaders and stakeholders in government, academia, industry, foundations, and patient advocacy with an interest in improving the system of drug discovery, development, and translation. The Forum brings together leaders from private sector sponsors of biomedical and clinical research, federal agencies sponsoring and regulating biomedical and clinical research, the academic community, and patients, and in doing so serves to educate the policy community about issues where science and policy intersect. The Forum convenes several times each year to identify, discuss, and act on key problems and strategies in the discovery, development, and translation of drugs. To supplement the perspectives and expertise of its members, the Forum also holds public workshops to engage a wide range of experts, members of the public, and the policy community. The Forum also fosters collaborations among its members and constituencies. The activities of the Forum are determined by its members, focusing on the major themes outlined below.

INNOVATION AND THE DRUG DEVELOPMENT ENTERPRISE

Despite exciting scientific advances, the pathway from basic science to new therapeutics faces challenges on many fronts. New paradigms for discovering and developing drugs are being sought to bridge the ever-widening gap between scientific discoveries and translation of those discoveries into life-changing medications. There is also increasing recognition of the need for new models and methods for drug development and translational science, and "precompetitive collaborations" and other partnerships, including public–private partnerships, are proliferating. The Forum offers a venue to discuss effective collaboration in the drug discovery and development enterprise and also hosts discussions that could help chart a course through the turbulent forces of disruptive innovation in the drug discovery and development "ecosystem."

Key gaps remain in our knowledge about science, technology, and methods needed to support drug discovery and development. Recent rapid advances in innovative drug development science present opportunity for revolutionary developments of new scientific techniques, therapeutic products, and applications. The Forum provides a venue to focus ongoing attention and visibility to these important drug development needs and facilitates exploration of new approaches across the drug development lifecycle. The Forum has held workshops that have contributed to the defining and establishment of regulatory science and have helped inform aspects of drug regulatory evaluation.

CLINICAL TRIALS AND CLINICAL PRODUCT DEVELOPMENT

Clinical research is the critical link between bench and bedside in developing new therapeutics. Significant infrastructural, cultural, and regulatory impediments challenge efforts to integrate clinical trials into the health care delivery system. Collaborative, cross-sector approaches can help articulate and address these key challenges and foster systemic responses. The Forum has convened a multiyear initiative to examine the state of clinical trials in the United States, identify areas of strength and weakness in our current clinical trial enterprise, and consider transformative strategies for enhancing the ways in which clinical trials are organized and conducted. In addition to sponsoring multiple symposia and workshops, under this initiative, the Forum is fostering innovative, collaborative efforts to facilitate needed change in areas such as improvement of clinical trial site performance.

INFRASTRUCTURE AND WORKFORCE FOR DRUG DISCOVERY, DEVELOPMENT, AND TRANSLATION

Considerable opportunities remain for enhancement and improvement of the infrastructure that supports the drug development enterprise. That infrastructure, which includes the organizational structure, framework, systems, and resources that facilitate the conduct of biomedical science for drug development, faces significant challenges. The science of drug discovery and development, and its translation into clinical practice, is cross-cutting and multidisciplinary. Career paths can be opaque or lack incentives such as recognition, career advancement, or financial security. The Forum has considered workforce needs as foundational to the advancement of drug discovery, development, and translation. It has convened workshops examining these issues, including consideration of strategies for developing a discipline of innovative regulatory science through the development of a robust workforce. The Forum will also host an initiative that will address needs for a workforce across the translational science lifecycle.

Forum on Drug Discovery, Development, and Translation

Robert Califf (Co-Chair)
Duke University and
Verily Life Sciences

Gregory Simon (Co-Chair)
Kaiser Permanente Washington
Health Research Institute and
University of Washington

Christopher Austin
National Center for Advancing
Translational Sciences, NIH

Linda Brady
National Institute of Mental Health,
NIH

Barry Coller
The Rockefeller University

Thomas Curran
Children's Mercy, Kansas City

Richard Davey
National Institute of Allergy and
Infectious Diseases, NIH

James Doroshow
National Cancer Institute, NIH

Jeffrey Drazen
New England Journal of Medicine

Steven Galson
Amgen Inc.

Carlos Garner
Eli Lilly and Company

Julie Gerberding
Merck & Co., Inc.

Deborah Hung
Harvard Medical School

Esther Krofah
Milken Institute

Ross McKinney
Association of American Medical
Colleges

Joseph Menetski
Foundation for the NIH

Bernard Munos
InnoThink Center for Research in
Biomedical Innovation

Kelly Rose
Burroughs Wellcome Fund

Joseph Scheeren
Critical Path Institute

Rob Scott
AbbVie, Inc.

Anantha Shekhar
Indiana University School of
Medicine

Ellen Sigal
Friends of Cancer Research

Lana Skirboll
Sanofi

Amir Tamiz
National Institute of Neurological
Disorders and Stroke, NIH

Ann Taylor
AstraZeneca

Pamela Tenaerts
Clinical Trials Transformation
Initiative

Joanne Waldstreicher
Johnson & Johnson

Carrie Wolinetz
National Institutes of Health, Office
of Science Policy

Alastair Wood
Vanderbilt University

Janet Woodcock
Center for Drug Evaluation and
Research, U.S. FDA

Project Staff

Carolyn Shore, Ph.D.
Forum Director
cshore@nas.edu
202-334-2669

Jennifer Hinners, M.D., M.P.H.
Program Officer

Amanda Wagner Gee, M.S.
Program Officer

Sylvia Ncha, M.P.H.
Associate Program Officer

Eeshan Khandekar, M.Sc.
Associate Program Officer

Melvin Joppy
Senior Program Assistant

For more information, please visit:

NATIONALACADEMIES.ORG/DRUGFORUM

Health and Medicine Division
Board on Health Sciences Policy

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The National Academies of SCIENCES ENGINEERING MEDICINE

Forum on NEUROSCIENCE and NERVOUS SYSTEM DISORDERS

The Forum on Neuroscience and Nervous System Disorders was established in 2006 to provide a venue for building partnerships, addressing challenges, and highlighting emerging issues related to brain disorders, which are common, major causes of premature mortality, and, in aggregate, the largest cause of disability worldwide. The Forum's meetings bring together leaders from government, industry, academia, disease advocacy organizations, and other interested parties to examine significant—and sometimes contentious—issues concerning scientific opportunities, priority setting, and policies related to research on neuroscience and brain disorders; the development, regulation, and use of interventions for the nervous system; and related ethical, legal, and social implications.

Forum members meet several times a year to exchange information, ideas, and differing perspectives. The Forum also sponsors workshops (symposia), workshop proceedings, and commissioned papers as additional mechanisms for informing its membership, other stakeholders, and the public about emerging issues and matters deserving scrutiny. Additional information is available at www.nas.edu/NeuroForum.

Recent Workshops

Biomarkers of Neuroinflammation (2017)

Enabling Novel Treatments for Nervous System Disorders by Improving Methods for Traversing the Blood-Brain Barrier (2017)

Accelerating Therapeutic Development for Pain and Opioid Use Disorders through Public-Private Partnerships (2017)

Neuroforensics: Exploring the Legal Implications of Emerging Neurotechnologies (2018)

Harnessing Digital Technology for Brain Disorders (2018)

Transgenic and Chimeric Neuroscience Research: Exploring the Scientific Opportunities Afforded by New Nonhuman Primate Models (2018)

The Role of Nonpharmacological Approaches to Pain Management (2018)

Advancing Gene-Targeted Therapies for Nervous System Disorders (2019)

Upcoming Workshop

Neuroscience Data in the Cloud (Sept 24, 2019)

Enhancing Scientific Reproducibility through Transparent Reporting (Sept 25-26, 2019)*

Challenges and a Way Forward in Sharing Clinical Trial Data (Nov 18-19, 2019)*

**Co-hosted with the Forum on Drug Discovery, Development, and Translation; National Cancer Policy Forum; and Roundtable on Genomics and Precision Health*

Staff

Clare Stroud, Forum Director
Sheena Posey Norris, Program Officer
Phoenix Wilson, Senior Program Assistant
Andrew Pope, Director, Board on Health Sciences Policy

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Roundtable on GENOMICS and PRECISION HEALTH

The sequencing of the human genome is rapidly opening new doors to research and progress in biology, medicine, and health care. At the same time, these developments have produced a diversity of new issues to be addressed.

The National Academies of Sciences, Engineering, and Medicine has convened a Roundtable on Genomics and Precision Health (previously the Roundtable on Translating Genomic-Based Research for Health) that brings together leaders from academia, industry, government, foundations and associations, and representatives of patient and consumer interests who have a mutual concern and interest in addressing the issues surrounding the translation of genome-based research for use in maintaining and improving health. The mission of the Roundtable is to advance the field of genomics and improve the translation of research findings to health care, education, and policy. The Roundtable will discuss the translation process, identify challenges at various points in the process, and discuss approaches to address those challenges.

The field of genomics and its translation involves many disciplines, and takes place within different economic, social, and cultural contexts, necessitating a need for increased communication and understanding across these fields. As a convening mechanism for interested parties from diverse perspectives to meet and discuss complex issues of mutual concern in a neutral setting, the Roundtable: fosters dialogue across sectors and institutions; illuminates issues, but does not necessarily resolve them; and fosters collaboration among stakeholders.

To achieve its objectives, the Roundtable conducts structured discussions, workshops, and symposia. Workshop summaries will be published and collaborative efforts among members are encouraged (e.g., journal articles). Specific issues and agenda topics are determined by the Roundtable membership, and span a broad range of issues relevant to the translation process.

Issues may include the integration and coordination of genomic information into health care and public health including encompassing standards for genetic screening and testing, improving information technology for use in clinical decision making, ensuring access while protecting privacy, and using genomic information to reduce health disparities. The patient and family perspective on the use of genomic information for translation includes social and behavioral issues for target populations. There are evolving requirements for the health professional community, and the need to be able to understand and responsibly apply genomics to medicine and public health.

Of increasing importance is the need to identify the economic implications of using genome-based research for health. Such issues include incentives, cost-effectiveness, and sustainability.

Issues related to the developing science base are also important in the translation process. Such issues could include studies of gene-environment interactions, as well as the implications of genomics for complex disorders such as addiction, mental illness, and chronic diseases.

Roundtable sponsors include federal agencies, pharmaceutical companies, medical and scientific associations, foundations, and patient/public representatives. For more information about the Roundtable on Genomics and Precision Health, please visit our website at nationalacademies.org/GenomicsRT or contact Sarah Beachy at 202-334-2217, or by e-mail at sbeachy@nas.edu.

500 Fifth Street, NW, Washington, DC 20001
Phone 202.334.2217 E-mail sbeachy@nas.edu nationalacademies.org/GenomicsRT

Roundtable on Genomics and Precision Health Membership

Geoffrey Ginsburg, M.D., Ph.D. (Co-Chair) Duke University
Michelle Penny, Ph.D. (Co-Chair) Biogen

Naomi Aronson, Ph.D.
BlueCross/BlueShield Association

Aris Baras, M.D., M.B.A.
Regeneron Pharmaceuticals

John Belmont, M.D., Ph.D.
Illumina

Karina Bienfait, Ph.D.
Merck and Co., Inc.

Vence Bonham, Jr., J.D.
National Human Genome Research Institute

Robert B. Darnell, M.D. Ph.D.
NY Genome Center / The Rockefeller University

Katherine Donigan, Ph.D.
U.S. Food and Drug Administration

W. Gregory Feero, M.D., Ph.D.
Journal of the American Medical Association

Jessica M. Gill, Ph.D., R.N., FAAN
National Institute of Nursing Research

Marc Grodman, M.D.
Genosity

Emily Harris, Ph.D., M.P.H.
National Cancer Institute

Richard Hodes, M.D.
National Institute on Aging

Praduman Jain, M.S.
Vibrent Health

Sekar Kathiresan, M.D.
Massachusetts General Hospital

Muin Khoury, M.D., Ph.D.
Centers for Disease Control and Prevention

Thomas Lehner, Ph.D., M.P.H.
National Institute of Mental Health

Patrick Loerch, Ph.D.
Johnson & Johnson

Sean McConnell, Ph.D.
American Medical Association

Mona Miller, M.P.P.
American Society of Human Genetics

Jennifer Moser, Ph.D.
U.S. Department of Veterans Affairs

Anna Pettersson, Ph.D.
Pfizer Inc.

Victoria M. Pratt, Ph.D., FACMG
Association for Molecular Pathology

Nadeem Sarwar, Ph.D.
Eisai Inc.

Sheri Schully, Ph.D.
NIH Office of Disease Prevention

Joan A. Scott, M.S., C.G.C.
Health Resources and Services Administration

Nikoletta Sidiropoulos, M.D.
University of Vermont Health Network Medical Group

Katherine Johansen Taber, Ph.D.
Myriad Women's Health

Jacquelyn Taylor, Ph.D.
New York University

Sharon Terry, M.A.
Genetic Alliance

Joyce Tung, Ph.D.
23andMe, Inc.

Jameson Voss, M.D.
U.S. Air Force

Michael S. Watson, Ph.D.
American College of Medical Genetics and Genomics

Karen Weck, M.D.
College of American Pathologists

Catherine A. Wicklund, M.S., C.G.C.
National Society of Genetic Counselors

Huntington F. Willard, Ph.D.
Geisinger Health

Janet K. Williams, Ph.D., R.N., FAAN
American Academy of Nursing

Sarah Wordsworth, Ph.D.
University of Oxford

Alicia Zhou, Ph.D.
Color Genomics

Project Staff
Sarah H. Beachy, Ph.D., *Roundtable Director*
Siobhan Addie, Ph.D., *Program Officer*
Meredith Hackmann, *Associate Program Officer*
Michael Berrios, M.A., *Senior Program Assistant*

NATIONAL CANCER POLICY FORUM

ABOUT THE FORUM

The National Cancer Policy Forum serves as a trusted venue in which experts can identify emerging high-priority policy issues in cancer research and care and work collaboratively to examine those issues through convening activities focused on opportunities for action. The Forum provides a continual focus within the National Academies on cancer, addressing issues in science, clinical medicine, public health, and public policy that are relevant to the goal of reducing the cancer burden, through prevention and by improving the care and outcomes for those diagnosed with cancer. Forum activities inform stakeholders about critical policy issues through published reports and often inform consensus committee studies. The Forum has members with a broad range of expertise in cancer, including patient advocates, clinicians, and basic, translational, and clinical scientists. Members represent patients, federal agencies, academia, professional organizations, nonprofits, and industry.

The Forum has addressed a wide array of topics, including:

- enhancing collaborations to accelerate research and development;
- improving the quality and value of care for patients who have been diagnosed with or are at risk for cancer;
- developing tools and technologies to enhance cancer research and care; and
- examining factors that influence cancer incidence, mortality, and disparities.

nationalacademies.org/NCPF

To receive updates on the National Cancer Policy Forum, visit **nationalacademies.org/HMDmail**

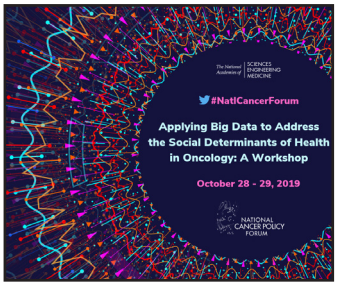
#NatlCancerForum

The National Academies of SCIENCES ENGINEERING MEDICINE

Upcoming Workshops

Applying Big Data to Address the Social Determinants of Health in Oncology

October 28-29, 2019



The collection and analysis of big data is expected to transform the field of cancer research and improve cancer care. Analyses of big data have the potential to elucidate ways in which the social determinants of health contribute to cancer incidence and outcomes, and may also identify promising avenues for intervention. However, not all individuals and communities may benefit equally from these advances: concerns remain about whether applications of big data research will reduce existing health disparities in oncology, or whether they might inadvertently exacerbate these disparities.

This workshop will examine the social determinants of health in the context of cancer, and consider opportunities to effectively leverage big data to improve health equity and reduce disparities. The workshop will feature invited presentations and discussion on topics such as:

- The impact of social determinants of health on cancer incidence and outcomes
- Opportunities to leverage big data and analytic methods in oncology
- Examples of novel data sources and methodologies
- Data policy and ethical considerations
- Priorities and opportunities for collaboration to improve health equity in oncology
- Identifying research gaps and setting a research agenda
- Participant recommendations for the path forward

Workshop website: http://nationalacademies.org/hmd/Activities/Disease/NCPF/2019-OCT-28.aspx

Save the Date: Future Workshops

July 13-14, 2020

November 9-10, 2020

Advancing Progress in Cancer Prevention and Risk Reduction

March 2-3, 2020

Evidence-based screening approaches have been found to reduce cancer morbidity and mortality, and ongoing research continues to evaluate the potential of new technologies and approaches for the early detection of cancer. However, there are a number of challenges related to the development and implementation of high-quality screening programs.

Even effective cancer screening tests have associated risks of harm, including the potential for false positive results that lead to unnecessary diagnostic testing, as well as overdiagnosis—the identification of abnormalities that would never result in harm—and overtreatment. Thus, communicating the risks and benefits of screening to patients, and engaging them in shared decision making are critical aspects of screening.

In addition, there are methodologic and implementation challenges in screening. These can include determining optimal screening intervals and age ranges, assessing the incremental benefit of early detection among different populations, and ensuring access to high-quality screening and, if warranted, cancer diagnosis and treatment.

This workshop will examine current issues in the development and implementation of effective, high-quality cancer screening. The workshop will feature invited presentations and panel discussions on topics that may include:

- Key gaps in the evidence base for cancer screening, including assessing potential benefits and risks
- Opportunities and challenges in developing, validating, and implementing new technologies for cancer screening tests
- Strategies to help patients understand the benefits, risks, and costs of cancer screening and participate in shared decision-making with their care team.
- Challenges in the clinical management of patients with premalignant lesions detected by screening.
- Opportunities to reduce disparities in cancer morbidity and mortality by facilitating patient access to high-quality screening and diagnosis in low-resource areas and among vulnerable populations.

Workshop website forthcoming

Recent Workshops and Publications

Health Literacy and Communication Strategies in Oncology

This workshop, held in collaboration with the Roundtable on Health Literacy, examined opportunities, methods, and strategies to improve the communication of cancer information in a clinic visit, across a health care organization, and among the broader community. Workshop presentations and discussion addressed procedures, policies, and programs to support health literacy needs of patients and families; best practices to improve communication about cancer prevention, detection, treatment, and survivorship; and communication strategies to build public trust and counter inaccurate information about cancer.

Workshop videos and presentation files:
http://nationalacademies.org/hmd/Activities/Disease/NCPF/2019-JULY-15.aspx

Updating Labels for Generic Oncology Drugs

In March 2019, participants examined the challenges and opportunities to update oncology drugs labels that are inconsistent with the current evidence base and use in clinical practice. Discussions focused on what information sources should be considered for labeling updates, evidentiary standards for labeling updates, and evidence considerations for special populations like pediatric oncology. This project was sponsored by the FDA and held in collaboration with the Forum on Drug Discovery, Development, and Translation.

Workshop presentation files:
http://nationalacademies.org/hmd/Activities/Disease/NCPF/2019-MARCH-26.aspx

Developing and Sustaining an Effective and Resilient Oncology Careforce

The landscape of cancer care is undergoing rapid change. Advances in cancer research, screening and diagnostic practices, and cancer treatment have led to improved outcomes for patients with cancer and a growing population of cancer survivors, but they have also increased the complexity of cancer care. Demographic trends, new care delivery and payment models, the widespread adoption of technologies in clinical practice, and increasing family caregiving responsibilities have had a profound effect on the cancer careforce. This workshop examined opportunities to better support the oncology careforce and improve the delivery of high-quality cancer care.

Workshop videos and presentation files:
http://nationalacademies.org/hmd/Activities/Disease/NCPF/2019-FEB-11.aspx

The Clinical Application of Computational Methods in Precision Oncology

Precision oncology therapies, which target specific abnormalities in a patient's cancer, are changing the nature of cancer treatment by enabling clinicians to select therapies that are most likely to benefit individual patients. Increasingly, oncologists are formulating cancer treatment plans using results from complex tests that characterize the molecular underpinnings of an individual patient's cancer. These advances depend on the use of computational methods to analyze large-scale datasets derived from genomic tests and other omics technologies. This workshop examined the challenges and opportunities in the development of computational methods for precision medicine to improve cancer diagnosis and care.

Workshop videos and presentation files:
http://nationalacademies.org/hmd/Activities/Disease/NCPF/2018-OCT-29.aspx

Proceedings:
http://nationalacademies.org/hmd/Reports/2019/improving-cancer-diagnosis-care-clinical-application-of-computational-methods-precision-oncology-pw.aspx

Advancing Progress in the Development of Combination Cancer Therapies with Immune Checkpoint Inhibitors

Immune checkpoint inhibitors, like those that target PD-1 and PD-L1 proteins, have changed the standard of care for multiple types of cancer and represent a majority share of new cancer drug applications to the FDA. There has been growing interest in combining checkpoint inhibitors with other therapies to further improve efficacy. Several challenges impede optimal development of combination therapies with checkpoint inhibitors, such as prioritizing combinations for testing, identifying patients who are most likely to benefit, assessing endpoints for safety and clinical benefit, overcoming resistance to therapy, and developing cancer site-agnostic indications. This workshop examined the opportunities to improve the development of combination cancer therapies that include immune checkpoint inhibitors.

Workshop videos and presentation files:
http://nationalacademies.org/hmd/Activities/Disease/NCPF/2018-JUL-16.aspx

Proceedings:
http://nationalacademies.org/hmd/Reports/2019/advancing-progress-development-of-combination-cancer-therapies-with-immune-checkpoint-inhibitors-pw.aspx

FORUM SPONSORS

- Centers for Disease Control and Prevention
- National Institutes of Health/National Cancer Institute
- American Association for Cancer Research
- American Cancer Society
- American College of Radiology
- American Society of Clinical Oncology
- Association of American Cancer Institutes
- Association of Community Cancer Centers
- Bristol-Myers Squibb
- Cancer Support Community
- CEO Roundtable on Cancer
- Flatiron Health
- Helsinn Therapeutics (U.S.), Inc.
- LIVE**STRONG** Foundation
- Merck
- National Comprehensive Cancer Network
- Novartis Oncology
- Oncology Nursing Society
- Pfizer Inc.

MEMBERSHIP OF THE FORUM

Edward J. Benz, Jr., M.D. *(Chair)*
Dana-Farber Cancer Institute

Garnet Anderson, Ph.D.
Fred Hutchinson Cancer Research Center

Kenneth Anderson, M.D.
Dana-Farber Cancer Institute

William Bailey, Pharm.D.
Helsinn Therapeutics (U.S.), Inc.

Karen Basen-Engquist, Ph.D., M.P.H.
The University of Texas MD Anderson Cancer Center

Chris Boshoff, M.D., Ph.D., F.MedSci.
Pfizer Inc.

Cathy Bradley, Ph.D.
University of Colorado Cancer Center

Otis W. Brawley, M.D.
Johns Hopkins University

Robert W. Carlson, M.D.
National Comprehensive Cancer Network

Nancy Davidson, M.D.
Fred Hutchinson Cancer Research Center/Seattle Cancer Care Alliance

George D. Demetri, M.D.
Dana-Farber Cancer Institute

James H. Doroshow, M.D.
National Cancer Institute

Nicole F. Dowling, Ph.D.
Centers for Disease Control and Prevention

Scot W. Ebbinghaus, M.D.
Merck

Kojo S.J. Elenitoba-Johnson, M.D.
Perelman School of Medicine, University of Pennsylvania

Awny Farajallah, M.D., FACP
Bristol-Myers Squibb

Stanton L. Gerson, M.D.
Case Comprehensive Cancer Center/ University Hospitals Seidman Cancer Center

Lori Hoffman Högg, M.S., R.N., CNS, AOCN
Department of Veterans Affairs

Linda House, R.N., BSN, MSM
Cancer Support Community

Hedvig Hricak, M.D., Ph.D.
Memorial Sloan Kettering Cancer Center

Roy A. Jensen, M.D.
The University of Kansas Cancer Center

Lisa Kennedy Sheldon, Ph.D., APRN-BC, AOCNP, FAAN
Oncology Nursing Society

Samir Khleif, M.D.
Georgetown University

Ronald M. Kline, M.D., FAAP
Centers for Medicare & Medicaid Services

Michelle M. Le Beau, Ph.D.
The University of Chicago Comprehensive Cancer Center

Mia Levy, M.D., Ph.D.
Rush University Cancer Center

J. Leonard Lichtenfeld, M.D., MACP
American Cancer Society

Neal Meropol, M.D.
Flatiron Health

Martin J. Murphy, Ph.D., D.Med.Sc.
CEO Roundtable on Cancer

Randall Oyer, M.D.
Lancaster General Health

Richard Pazdur, M.D.
Food and Drug Administration

Richard L. Schilsky, M.D., FASCO
American Society of Clinical Oncology

Deborah Schrag, M.D., M.P.H.
Dana-Farber Cancer Institute

Lawrence N. Shulman, M.D.
University of Pennsylvania Abramson Cancer Center

Dan Theodorescu, M.D., Ph.D.
Samuel Oschin Comprehensive Cancer Institute at Cedars-Sinai

Verena Voelter, M.D.
Novartis Oncology

George J. Weiner, M.D.
University of Iowa Holden Comprehensive Cancer Center

Robert A. Winn, M.D.
University of Illinois Health Cancer Center

FORUM STAFF

Sharyl Nass, Ph.D.
Forum Director and Director, Board on Health Care Services

Erin Balogh, M.P.H.
Senior Program Officer

Emily Zevon, Sc.D.
Associate Program Officer

Natalie Lubin, B.A.
Research Assistant

Annalee Gonzales, B.A.
Administrative Assistant

Ruth Cooper, B.A.
Senior Program Assistant

Micah Winograd, M.A.
Financial Officer

The National Academies of
SCIENCES • ENGINEERING • MEDICINE
Forum on Drug Discovery, Development, and Translation
Forum on Neuroscience and Nervous System Disorders
National Cancer Policy Forum
Roundtable on Genomics and Precision Health

Planning Committee Biographies

CO-CHAIRS

JEFFREY DRAZEN (co-chair), M.D., was born in Missouri. He attended Tufts University, with a major in physics, and Harvard Medical School, and served his medical internship at Peter Bent Brigham Hospital in Boston. Thereafter, he joined the Pulmonary Divisions of the Harvard hospitals. He served as Chief of Pulmonary Medicine at the Beth Israel Hospital, Chief of the combined Pulmonary Divisions of the Beth Israel and Brigham and Women's Hospitals, and finally as the Chief of Pulmonary Medicine at Brigham and Women's Hospital. Through his research, he defined the role of novel endogenous chemical agents in asthma. This led to four new licensed pharmaceuticals for asthma, with over 5 million people on treatment worldwide. In 2000, he assumed the post of Editor-in-Chief of the *New England Journal of Medicine* (*NEJM)*. During his tenure, the *NEJM* has published major papers advancing the science of medicine, including the first descriptions of SARS and papers modifying the treatment of cancer, heart disease and lung disease. The *NEJM*, which has over a million readers every week, has the highest impact factor of any journal publishing original research.

JOANNE WALDSTREICHER (co-chair), M.D., is Chief Medical Officer, Johnson & Johnson. In this role, she has cross-sector oversight for safety of all Johnson & Johnson products worldwide. In addition, she also plays a leadership role for epidemiology, internal and external partnerships and collaborations, and development of the corporate science, technology, and R&D policies, including those related to clinical trial transparency. Dr. Waldstreicher also chairs the pharmaceuticals R&D Development Committee, which reviews all late stage development programs in the pharmaceutical pipeline. Under her leadership in a prior role as both Chief Medical Officer of the Pharmaceutical sector and Head of Asia Pacific Medical Sciences, four legacy safety groups were integrated into one independent Global Medical Safety organization. In addition, Dr. Waldstreicher reshaped and realigned the R&D and medical affairs groups across Asia Pacific, resulting in an industry leading drug pipeline in Japan, and the company's first ever international drug approval from a team based in China.

Prior to becoming Chief Medical Officer in the pharmaceutical sector, Dr. Waldstreicher served as Head of Global Drug Development for the Johnson & Johnson Pharmaceutical Research & Development, L.L.C. (J&JPRD) CNS/Internal Medicine business unit. In this role she was responsible for late-stage development of the CNS/Internal Medicine pipeline, 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 of Mevacor®, Zocor®, Proscar® and Propecia®, and for 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 has received both 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. Dr. Waldstreicher combines broad experience in science, medicine, and pharmaceutical development with a passion for advancing transparency and trust in our industry, and a dedication to advancing patient safety.

MEMBERS

MONICA BERTAGNOLLI, M.D, is a Professor of Surgery at Harvard University Medical School and an Associate Surgeon at Dana Farber Cancer Institute and Brigham and Women's Hospital. Dr. Bertagnolli is presently the Chief of the Division of Surgical Oncology at Dana Farber/Brigham and Women's Cancer Center. She is also a member of the Gastrointestinal Cancer and Sarcoma Disease Centers at DF/BWCC, where she collaborates with colleagues in medical oncology, radiation oncology, and pathology to treat cancer patients in a tertiary care setting. Dr. Bertagnolli maintains an active research laboratory focused on understanding the role of the inflammatory response in epithelial tumor formation. In 1999, she extended her basic laboratory observations to the clinical trials setting as the lead Principal Investigator of the Adenoma Prevention with Celecoxib Trial. This pivotal study, reported in 2006, demonstrated dramatic suppression of colorectal adenomas with selective cyclooxygenase-2 inhibition, but also uncovered unanticipated cardiovascular toxicity with these agents. Dr. Bertagnolli was an organizing member of gastrointestinal correlative science initiatives within the NCI-funded Cancer Cooperative Groups, where she facilitated integration of tumor-specific molecular markers of treatment outcome into nation-wide clinical cancer treatment protocols. She has had numerous leadership roles in multi-institutional cancer clinical research consortia, and from 2010-2011 served as Group Chair of Cancer and Leukemia Group B (CALGB). Dr. Bertagnolli was elected in July 2011 to lead the Alliance for Clinical Trials in Oncology, a new NCI-funded cooperative group formed by merger of CALGB, the North Central Cancer Treatment Group, and the American College of Surgeons Oncology Group.

TIMOTHY COETZEE, Ph.D., is the chief advocacy, services and research officer at the National Multiple Sclerosis Society (NMMS) in New York. Dr. Coetzee has been engaged in multiple sclerosis advocacy work throughout his career. He leads the society's federal and state activism programs and manages its investment in basic, clinical and commercial research. He has also helped launch and served as president of Fast Forward, an initiative of the NMMS to speed the commercial development of new treatments for multiple sclerosis. He earned his Ph.D. at Albany Medical College in New York.

PATRICK CULLINAN, Ph.D., is currently Senior Director of Medical and Technical Communications at Bluebird Bio, a leading gene-therapy innovator located in Cambridge Massachusetts. In this role, Dr. Cullinan contributes to Bluebird's commitment to patient-centric clinical development and management of regulatory submissions to support delivering vital new innovative therapies to patients.

Formerly, Dr. Culinan was the Head of Clinical Trial Transparency and the Head of Science Advocacy for Takeda Pharmaceuticals. In this role, Dr. Cullinan was responsible for patient-centric clinical trial transparency, entailing publicly sharing information about Takeda clinical trials to communicate our research programs to the wider patient, scientific and regulatory communities. In this role Dr. Cullinan also managed Takeda's data sharing program whereby independent researchers could apply to Takeda for access to data sets to further leverage these data to further our collective scientific understanding of these medicines and demonstrate Takeda's commitment to transparency. As the Head of Science Advocacy, Dr. Cullinan spearheaded external engagement on scientific and bioethics topics relating of medical research policy.

Dr. Cullinan was originally from Dublin Ireland where he attended University College Dublin and studied Cell Biology and Molecular Genetics. Dr. Cullinan then moved to Chicago and completed is Ph.D. in Pathology from the University of Chicago (Pritzker Medical School), where his research focused on Immunology, and specifically on T cell activation and motility.

SONALI KOCHHAR, M.D., is a clinical associate professor of global health at the University of Washington. She has twenty years of experience in a leadership position for Global Phase I-IV Clinical Research, Epidemiology and Safety Studies for Vaccines and Drugs conducted in the USA, Europe, Asia, Africa and India in Adult and Pediatric populations; Infectious Diseases (HIV/AIDS, Diarrheal Diseases, Influenza, Group B Strep, RSV,

Malaria, Tropical diseases); Vaccines for Pregnant Women; Introduction of New Vaccines; Pandemic Preparedness; Translating research into programs by Healthcare Systems Strengthening (including Governance, Capacity Building, Logistics and Supply Chains) for Immunization and Reproductive, Maternal and Child Health programs; Research Ethics for Vaccines, Epidemics, and New Technologies; HIV/AIDs Prevention, and Treatment; and working with vulnerable and at-risk populations (including women, adolescents, children, MSMs, IDUs and Sex Workers).

Dr. Kochhar has conducted multiple studies – including vaccine trials – in Kenya, Uganda, Zambia, The Gambia, South Africa, the United States, India, Bangladesh and Nepal. She provides expertise for Vaccine Research and Development (including for Viral Vectors (MMV, AAV, Ad-5 etc), DNA, Proteins, Inactivated, Subunit Vaccines, Adjuvants etc).

She has set up global multimillion dollar international strategic partnerships for policy, ethics and regulatory development; communications, institutional strengthening; and strategy and advocacy with government partners (across Asia-Pacific, Africa, and South America); international aid agencies (Gates Foundation, USAID, Basque Government etc.); public health authorities; international and bilateral Organizations (e.g. WHO, NIH, CDC); regulatory bodies (including the FDA, EMEA, African and Asian National Regulatory Authorities); ethical committees; scientific organizations; international pharmaceutical companies; local communities; and the Media.

Dr. Kochhar has co-authored multiple internationally accepted Vaccine Safety Research Standards, Case definitions of Adverse Events for Vaccines Research, Guidelines for Vaccine Safety Research and Template Research Protocols. She is the Lead for the development of Maternal and Neonatal Case Definitions for Maternal Immunisation (MI) and has co-authored guidelines for vaccines in pregnant women clinical trials. The definitions are being utilized in clinical research, observational studies and AEFI surveillance for Vaccines and Maternal and Child Health Research globally. She has helped to develop Research Ethics for Vaccines, Epidemics, and New Technologies.

She is a member of various International Steering Committees, including the Wellcome Trust's Independent Review Panel (IRP) for Clinical Study Data Requests and Vivli (Centre for Global Clinical Research Data Sharing), Vice-Chair for the Brighton Collaboration Science Board, International Alliance for Biological Standardization Human Vaccine Committee, Maternal Immunization Pharmacovigilance programs for Low and Medium Income Countries, the Gates Foundation funded Global Alignment of Immunization Safety Assessment in Pregnancy (GAIA), UK-Medical Research Council funded Immunizing Pregnant Women and Infants (IMPRINT) network etc. She is an Expert Working Group member of the PREVENT (Pregnancy Research Ethics for Vaccines, Epidemics, and New Technologies) Wellcome Trust Project, Viral Vector Vaccine Safety Working Group, European Commission Expert Evaluator for Vaccine Research, Medical Advisory Panel member of Group B Strep Support, UK, the International Society for Pharmacoepidemiology (IABS) Vaccine Special Interest and Maternal Immunization Expert Groups and the Vienna Vaccine Safety Initiative working on global vaccine research and safety. She was on the International Steering Committee of the WHO Consultation on Safety of Immunization in Pregnancy in Mothers and Newborn Children and served as the Session Chair for the WHO Pediatric Deliberative Session on neonatal events in maternal immunization. She was a member of the Gates Foundation's Global Health Clinical Consortium Leadership Group, Harvard University's Multi-Regional Clinical Trial Group, Maternal Immunization Pharmacovigilance programs for Low and Medium Income Countries, and expert groups involved in the development of a WHO globally integrated Vaccine Safety Monitoring system, Safety Standards for Malaria, Tuberculosis and AIDS Vaccine trials.

Dr. Kochhar serves on International Scientific Advisory Boards, as Guest Faculty for International Vaccinology Programs [including the LIVE (Leading International Vaccinology Education) Master's Program, Europe] and the Oxford Vaccinology course, as a reviewer for international journals, and has published numerous peer-reviewed publications and book chapters. She has been awarded the Yale World Fellowship for 2011 (Yale University's International Leadership Program); Vaccinology Fellowship Award for significant achievements in Vaccinology

from Fondation Mérieux and University of Geneva; Global Leadership Awards from Eli Lilly & Company, Indianapolis, U.S.A; Bharat Jyoti (Light of India) Award for medical achievements and the Serviers Young Investigator Award by Institut de Recherches Internationales, Servier, France.

BERNARD LO, M.D., 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. At the National Institutes of Health (NIH), he serves on Data and Safety Monitoring Committees for HIV vaccine trials and the Long-term Oxygen Treatment Trial. Dr. Lo also serves on the Board of Directors of Association for the Accreditation of Human Research Protection Programs (AAHRPP) 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 NIH Recombinant DNA Advisory Committee, and the Ethics Subcommittee, and the Advisory Committee to the Director of the Centers for Disease Control and Prevention. As 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 of Life Sciences at the National Academy of Science (NAS).

Dr. Lo and his colleagues have published around 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.

DEVEN MCGRAW, J.D., is the Chief Regulatory Officer for Ciitizen. Prior to joining Ciitizen, she directed U.S. health privacy and security policy through her roles as Deputy Director for Health Information Privacy at the HHS Office for Civil Rights (the office that oversees HIPAA policy and enforcement) and Chief Privacy Officer (Acting) of the Office of the National Coordinator for Health IT. Deven also advised PCORNet (the Patient Centered Outcomes Research Network), as well as the federal All of Us Research Initiative, on HIPAA and patient-donated data research initiatives.

DINA PALTOO, Ph.D, M.P.H, is the assistant director for policy development in the Office of the Director at the National Library of Medicine. Dr. Paltoo leads NLM's policy and legislative activities that promote stewardship and access to scientific and clinical data and information, as well as health information technology. She also works across the NIH and Federal agencies on initiatives and activities relevant to these topics, including open science and data science. Prior to joining NLM, Dr. Paltoo was the Director of the Division of Scientific Data Sharing Policy within the NIH Director's Office of Science Policy (OSP). While there, she was responsible for NIH policy efforts in scientific data sharing and management, open science, and genomics and health, including, for example, NIH policies for the sharing of genomic data and the NIH-Lacks family agreement and policy on the sharing of HeLa genome sequence data. She previously was the Director of the Genetics, Health, and Society Program in OSP. Dr. Paltoo joined OSP from NIH's National Heart, Lung, and Blood Institute, where she was a Program Director in genetics, pharmacogenetics, and personalized medicine and led activities to promote the sharing of these and other data. She also served as a scientific advisor on the Department of Health and Human Services Secretary's Personalized Healthcare Initiative, was a National Cancer Institute Cancer Prevention Fellow in Molecular Epidemiology, and taught at Howard and Morgan State Universities. Dr. Paltoo received her Ph.D.

in Physiology and Biophysics from Howard University, was a postdoctoral fellow in Cellular Biophysics and Biochemistry at the University of Medicine and Dentistry of New Jersey, and obtained her M.P.H. from the Johns Hopkins Bloomberg School of Public Health (with a concentration in Epidemiology and Biostatistics).

LIZ ROBERTS, M.Sc., is Senior Director, Global Lead, Transparency and Data Sharing at UCB. She relocated from the U.K. to the U.S. in 2009 and is currently based in Maryland. She is responsible for establishing the strategic framework, guiding principles, and corporate policies that inform transparency and responsible data sharing. In addition, she is responsible for aligning best practices that will enable current and future global disclosure requirements including Global Clinical Trial Registration & Results Disclosure, Clinical Data Sharing, Lay Summaries, and other Transparency-related activities. Ms. Roberts represents UCB on the TransCelerate workstream for Clinical Data Transparency and is involved with sharing data for secondary research via ClinicalStudyRequest.com (CSDR). Prior to this role in Transparency, she had more than 20 years' experience working as a statistician in the pharmaceutical industry and has and M.Sc.in Applied Statistics.

FRANK ROCKHOLD, PH.D., Sc.M., is a fulltime Professor of Biostatistics and Bioinformatics at Duke University Medical Center *(Scholars at Duke),* Affiliate Professor of Biostatistics at Virginia Commonwealth University, and Managing Partner of HunterRockhold, Inc. His 40+-year career includes senior research positions at Lilly, Merck, and GlaxoSmithKline, where he retired as Chief Safety Officer and Senior Vice President of Global Clinical Safety and Pharmacovigilance. He has held faculty appointments at six different universities. Dr. Rockhold served for 9 years on the board of directors of the non-profit CDISC, most recently as Chairman, and is past president of the Society for Clinical Trials and a past member of the PCORI Clinical Trials Advisory Panel. He is currently on the board of the Frontier Science and Technology Research Foundation and a technical advisor to EMA.

Dr. Rockhold has diverse research interests and consulting experience in industry and academia including clinical trials design, data monitoring, benefit/risk, safety and pharmacovigilance and has been a leader in the scientific community in promoting data disclosure and transparency in clinical research. Frank is widely published in major scientific journals across a wide variety of research topics.

Frank holds a BA in Statistics from The University of Connecticut, an ScM in Biostatistics from The Johns Hopkins University, and a PhD in Biostatistics from the Medical College of Virginia at Virginia Commonwealth University. Frank is an Elected Fellow of both the American Statistical Association and the Society for Clinical Trials, a Fellow of the Royal Statistical Society, an Accredited Professional Statistician, PStat®, and a Chartered Statistician, CStat.

SHARON TERRY, M.A., is President and CEO of Genetic Alliance, an enterprise engaging individuals, families and communities to transform health. Genetic Alliance works to provide programs, products and tools for ordinary people to take charge of their health and to further biomedical research. Terry cofounded PXE International, a research advocacy organization for the genetic condition pseudoxanthoma elasticum (PXE), in response to the diagnosis of PXE in her two children in 1994. With her husband, she co-discovered the ABCC6 gene, patented it to ensure ethical stewardship in 2000, and assigned their rights to the foundation. She subsequently developed a diagnostic test and conducts clinical trials. She is the author of 150 peer-reviewed papers, of which 30 are clinical PXE studies. Her story is the topic of her TED Talk and TED Radio Hour.

In her focus at the forefront of consumer participation in genetic research, services and policy, Terry serves in a leadership role on many of the major international and national organizations, including the Precision Medicine Initiative Cohort Advisory Panel; Accelerating Medicines Partnership; the Cures Acceleration Network Review Board, and the Advisory Council, National Center for Accelerating Translation Science, NIH; National Academy

of Medicine Roundtable on Genomics and Precision Health; the PhenX scientific advisory board; the Global Alliance for Genomics and Health; the International Rare Disease Research Consortium Executive Committee; and as Founding President of EspeRare Foundation of Geneva, Switzerland. Terry is co-founder of the Genetic Alliance Registry and Biobank. She is on the editorial boards of several journals, including Genome, Patient Engagement Editor for Genetic Alliance's official journal Genetic Testing and Molecular Biomarkers, Chief Patient Advisor for Clinical and Translational Science. She led the coalition that was instrumental in the passage of the Genetic Information Nondiscrimination Act.

Terry received an honorary doctorate from Iona College for her community engagement work in 2006; the Research!America Distinguished Organization Advocacy Award and an inaugural member of Disruptive Women in Health Care in 2009; and the Clinical Research Forum and Foundation's Annual Award for Leadership in Public Advocacy in 2011. She was named one of FDA's "30 Heroes for the Thirtieth Anniversary of the Orphan Drug Act" in 2013. She is co-inventor of the Platform for Engaging Everyone Responsibly (PEER), receiving a large grant from the Robert Wood Johnson Foundation in 2014. PEER undergirds the Community Engaged Network for All (CENA), a PCORnet member since 2013. She is Co-PI of the PCORnet Coordinating Center and Chair of the PCORnet Engagement Committee. She was a member of the Blue Ribbon Panel's Working Group on Enhanced Data Sharing for the Cancer Moonshot. She was named a National Associate of the National Research Council, National Academies of Engineering, Sciences, and Medicine for her extraordinary service. She received the Health 2.0 Health Activist award in 2016. In 2017, she co-founded the People Centered Research Foundation. Terry is an Ashoka Fellow. She is an avid student of Gestalt Awareness Practice. With her husband Patrick, she paraglidings, runs, and dreams of spending more time writing and reflecting.

DEBORAH ZARIN, M.D., is the Program Director, *Advancing the Clinical Trials Enterprise,* and Member of the Faculty, Harvard Medical School. She was the Director of ClinicalTrials.gov between 2005 and 2018. In that capacity, she oversaw the world's largest clinical trials registry, as well as the development and implementation of the first public database for summary clinical trial results. She also played a major role in the development and implementation of key legal and policy mandates for clinical trial reporting, including regulations under FDAAA (42 CFR Part 11) and the NIH trial reporting policy. Dr Zarin's recent research has been on the quality of trial reporting, as well as issues in the design and analysis of clinical trials. Previous positions held by Dr. Zarin include the Director, Technology Assessment Program, at the Agency for Healthcare Research and Quality, and the Director of the Practice Guidelines program at the American Psychiatric Association. In these positions, Dr. Zarin conducted systematic reviews and related analyses in support of evidence based clinical and policy recommendations. Dr. Zarin graduated from Stanford University and received her doctorate in medicine from Harvard Medical School. She completed a clinical decision making fellowship and a pediatric internship, and is board certified in general psychiatry as well as in child and adolescent psychiatry.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE
Forum on Drug Discovery, Development, and Translation
Forum on Neuroscience and Nervous System Disorders
National Cancer Policy Forum
Roundtable on Genomics and Precision Health

Speaker Biographies

COLIN BAIGENT, B.M., B.CH., M.SC., M.A., is Director of the MRC Population Health Research Unit and Deputy Director of the Clinical Trial Service Unit and Epidemiological Studies Unit (CTSU) at the University of Oxford, where he is Professor of Epidemiology. He was elected a Fellow of the Academy of Medical Sciences in 2019.

Professor Baigent's main scientific interests are (i) the design and conduct of large-scale streamlined randomized trials, and (ii) in the use of collaborative meta-analyses of individual patient data from randomized trials for the prevention and treatment of cardiovascular disease, and for the study of adverse drug effects. He has led some of the world's largest collaborative meta-analyses of randomized trials, typically with individual participant data, resulting in landmark papers that have helped determine, for example, the effects of statins in different types of patients (the Cholesterol Treatment Trialists' [CTT] Collaboration), aspirin (and other antiplatelet drugs), fibrinolytic therapy (in both acute myocardial infarction and ischaemic stroke), and the cardiovascular hazards of non-steroidal anti-inflammatory drugs (NSAIDs).

He has a particular interest in the epidemiology of cardiovascular disease in renal failure, where there are special difficulties owing to distortion of associations due to reverse causality. He led the Study of Heart and Renal Protection (SHARP), the largest ever randomized trial in patients with moderate-to-severe chronic kidney disease (CKD), recruiting 9438 patients in nearly 400 hospitals in 18 countries, which showed that simvastatin 20mg plus ezetimibe 10mg daily significantly reduced the risk of atherosclerotic events. His group is now coordinating the EMPA-KIDNEY trial, a trial comparing empagliflozin versus placebo in 5000 patients with CKD, which is designed to assess the effects of empagliflozin on progression to ESRD or cardiovascular death.

SEAN COADY, M.S., M.A., is a statistician and Acting Deputy Branch Chief in the Epidemiology Branch, Division of Cardiovascular Sciences, at the NIH National Heart, Lung, and Blood Institute. In addition to maintaining a portfolio of training and research project grants, Mr. Coady has managed the NHLBI Data Repository since 2002.

RAMIN DARON, M. ENG., is Vice President, Data Architecture and Technology at Takeda Pharmaceutical's Data Sciences Institute focused on innovation in R&D to drive improvements in execution and novel insights through analytics and technology. Ramin has over 20 years of experience with increasing responsibility within the pharmaceutical and information technology industries, with interest in creative and pragmatic approaches to design, implementation and growth of impactful initiatives. He holds a master's degree in Engineering Science from Penn State University and a bachelor's degree in Mechanical Engineering from Drexel University and currently based in Boston.

DAVID DEMETS, PH.D., is currently the Max Halperin Professor of Biostatistics, Emeritus, and former Chair of the Department of Biostatistics and Medical Informatics at the University of Wisconsin – Madison. He received his PhD in biostatistics in 1970 from the University of Minnesota. Following a

postdoctoral appointment in the Division of Computer Research and Technology at the National Institutes of Health (1970-72), he spent ten years (1972-1982) at the National Heart, Lung and Blood Institute at the National Institutes of Health where he was a member of and later became chief of the Biostatistics Branch. In 1982, he joined the University of Wisconsin and later founded the Department of Biostatistics and Medical Informatics which he chaired until 2009. In 2017, He became emeritus professor.

He has co-authored four texts, *Fundamentals of Clinical Trials*, *Data Monitoring in Clinical Trials: A Case Studies Approach*. *Data Monitoring Committees in Clinical Trials: A Practical Perspective, and Statistical Methods for Clinical Trials.* He has served on numerous NIH and industry-sponsored Data Safety and Monitoring Committees for clinical trials in diverse disciplines. He served on the Board of Directors of the Society for Clinical Trials (1983-1987) American Statistical Association (1987-89), as well as having been President of the Society for Clinical Trials (1989) and President of the Eastern North American Region (ENAR) of the Biometric Society (1993). In addition, he was Elected Fellow of the International Statistics Institute in 1984, the American Statistical Association in 1986, the Association for the Advancement of Science in 1998, the Society for Clinical Trials in 2006 and the American Medical Informatics Association in 2008. In 2013, he was elected as a member of the Institute of Medicine/now the National Academy of Medicine. His research interests include the design, data monitoring and analysis of clinical trials, especially large Phase III randomized clinical trials. He is well known for his work on sequential statistical methods for monitoring interim data for early evidence of intervention benefit or possible harm.

ERNIE HAWK, M.D., M.P.H., is vice president and division head for Cancer Prevention and Population Sciences at The University of Texas MD Anderson Cancer Center and holds the Boone Pickens Distinguished Chair for Early Prevention of Cancer. He also leads the Duncan Family Institute for Cancer Prevention and Risk Assessment and serves as co-director of the Moon Shot-supported Cancer Prevention and Control Platform. His personal research interests over the last two decades include preclinical and clinical drug development for cancer prevention, the conduct of a variety of clinical trials in cancer prevention, and the inclusion of diverse and underserved populations in cancer clinical trials, clinical research, and cancer control programs to improve outcomes and promote equity. He earned his medical degree from Wayne State University and his MPH degree from Johns Hopkins University. He received training in internal medicine and served as a staff physician at Emory University, completed a medical oncology fellowship at the University of California, San Francisco, and completed a cancer prevention fellowship at the National Cancer Institute.

MARK HELFAND, M.D., M.S., F.A.C.P., is a staff physician in the VA Portland Health Care System and Professor of Medicine at Oregon Health & Science University. Dr. Helfand received an AB and BS from Stanford University, an MD and MPH from University of Illinois Medical School, and an MS in health services research from Stanford University. He is board-certified in Internal Medicine. He has conducted systematic reviews for a wide variety of organizations, including the American College of Physicians, the US Preventive Services Task Force, NASEM, and the VA Healthcare System. In addition to research synthesis, his research focuses on scientific communication and peer review. He was Editor-in-chief of Medical Decision Making from 2005 to 2012 and currently leads the program for external peer review of research reports for the Patient Centered Outcomes Research Institute (PCORI). He is also a member of the PCORI Methodology Committee.

MARTIN HO, M.S., Martin is Associate Director of Science for Patient Inputs and Real-World Evidence, Office of Biostatistics & Epidemiology, U.S. Food and Drug Administration Center for Biologics Evaluation and Research (CBER). At CBER, he leads research efforts and establish review practices regarding quantitative patient inputs, real-world evidence (RWE), and digital health. He also represents CBER to coauthor multiple guidance documents, including PFDD and Digital Health technologies. He is CBER's methodological lead for guidance development and building review capacities for clinical

outcome assessments and patient preference information, as well as site-less clinical trials. Prior to CBER, he served as Associate Director for Quantitative Innovations at Office of Surveillance and Biometrics, FDA, Center for Device and Radiological Health, playing similar roles and leading the real-world performance component of the Center's Digital Health Precertification Program. He co-chairs the American Statistical Association (ASA) Real-World Evidence Scientific Working Group. He is also the past president of the FDA Statistical Association and Chair of the ASA Medical Device and Diagnostic Section.

GEORGINA HUMPHREYS, PH.D., is currently a clinical data sharing manager at Wellcome Trust, a board member at ISRTCN, and an associate editor at the Transactions of the RSTMH. Georgie is committed to maximising the benefits from research data and has experience in clinical trials, both in UK academic institutions, and in the field in East Africa. She completed an MSc at the London School of Tropical Medicine and Hygiene and a PhD at University of Glasgow. Georgie then worked as a postdoctoral researcher in Tanzania, before moving back to the UK and joining the University of Oxford where she worked for 6 years on individual patient data meta analyses. Georgie joined Wellcome in December 2018 and currently leads work on clinical trial transparency and data sharing. Georgie sits on the ISRCTN Board, the BMGF Highly Efficient Clinical Trials Advisory Board, the Health Research Authority Transparency Forum, and the IDDO Ebola Data Sharing Steering Committee.

LYRIC JORGENSON, PH.D., is the Deputy Director for the Office of Science Policy at the National Institutes of Health. In this position, she provides senior leadership in the development and oversight of cross-cutting biomedical research policies and programs considered to be of high-priority to NIH and the United States Government. Most recently, she was also the Deputy Executive Director of the White House Cancer Moonshot Task Force in the Office of the Vice President in the Obama administration, where she directed and coordinated cancer-related activities across the Federal government and worked to leverage investments across sectors to dramatically accelerate progress in cancer prevention, diagnosis, and treatment.

Prior to joining the Office of Science Policy, she was a senior science policy advisor and analyst under the Deputy Director for Science, Outreach, and Policy and assisted in the creation of new, high impact science and policy initiatives such as the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative and the National Center for Advancing Translational Sciences (NCATS). She was also an AAAS Science and Technology Fellow and has received numerous awards in recognition of her accomplishments and service.

Dr. Jorgenson earned a doctorate degree from the Graduate Program for Neuroscience at the University of Minnesota-Twin Cities where she conducted research in neurodevelopment with a focus on learning and memory systems. She earned a Bachelor's degree in Psychology from Denison University.

PANDURANG KULKARNI, PH.D., is the Chief Analytics Officer – R&D / Vice President of Biometrics and Advanced Analytics Organization at Eli Lilly and Company. He joined Eli Lilly in 2000 and has held numerous leadership positions including technical and management positions within and outside of Statistics. He has given numerous invited presentations across the globe to provide training, Continuing Medical Education, and workshops on the use of statistics in medical research. He co-led the transparency efforts of sharing placebo data with Trancelerate, and has been instrumental in enabling and ensuring the success of Transparancy effort on sharing Lilly Clinical Data through Vivli and through Data Sphere. He has published more than 50 articles in statistics and medical areas in peer reviewed journals.

REBECCA KUSH, PH.D., has a dedication to accelerating learning health cycles, particularly focused on the research aspect and its link with healthcare. Dr. Kush is currently the Chief Scientific Officer for Elligo Health Research (Austin, TX), President of Catalysis and Fellow for the Translational Research

Innovation Center of the Foundation for Biomedical Research and Innovation (Kobe, Japan). She serves on Boards for the Learning Health Community, ACRES, Litmus Health and Saama.

Dr. Kush has over 40 years of experience in medical research and related process improvement, technology and standards. She is Founder and President Emeritus of the Clinical Data Interchange Standards Consortium (CDISC), a non-profit (501c3) standards development organization (SDO) that developed and harmonized a suite of global clinical research data standards to support clinical research in therapeutic areas that affect over 1.5 billion lives; CDISC standards and controlled terminology (maintained by NCI/EVS) are now required for data submitted to FDA and Japan's PMDA for approval of new therapies.

Dr. Kush launched the CDISC Healthcare Link Initiative, which was integral in the development of the HITSP Interoperability Specification #158 (research use case) and several IHE Profiles, including as Retrieve Form for Data Capture (RFD). She co-founded the initial clinical research and clinical genomics working groups within HL7 and spearheaded efforts through which the collaborative Biomedical Research Integrated Domain Group (BRIDG) model, which bridges research and healthcare, became an ISO, CDISC and HL7 standard. Dr. Kush is currently participating (through Elligo) in FDA-led projects funded through the PCOR Trust Fund on common data model harmonization (CDMH) to facilitate the use of RWE for regulatory purposes. She was co-chair of the Bridging Collaborative and author of a white paper on the System of Accelerating Research (SOAR).

Dr. Kush served previously on the U.S. Federal Health IT Standards Committee (HITSC) for over five years, the National Cancer Institute's Center for Biomedical Informatics IT Advisory Committee, the Coalition for Accelerating Standards and Therapies (CFAST), the BRIDG Board and the Boards of DIA and HL7. She has participated in prior advisory groups for the National Academies of Medicine, including those on Data Sharing and Digital Infrastructure for a Learning Health System. She also participated in the development of a consensus document on Data Sharing through the EU CORBEL Initiative.

Prior to spending 20 years as President and CEO of CDISC, Dr. Kush held positions with academia, the U.S. National Institutes of Health, a global clinical research organization and biopharma companies in the U.S. and Japan. She earned her doctorate in Physiology and Pharmacology from the University of California San Diego School of Medicine and a B.S. in Biology and Chemistry with honors from the University of New Mexico. She has publications in Science Translational Medicine, New England Journal of Medicine, British Medical Journal and other journals, in addition to chapters in two editions of the Springer Clinical Research Informatics textbook.

HARLAN KRUMHOLZ, M.D., S.M., is a cardiologist and health care researcher at Yale University and Yale-New Haven Hospital. He received a B.S. from Yale, an M.D. from Harvard Medical School, and a masters in health policy and management from the Harvard University School of Public Health. He is the Harold H. Hines, Jr. Professor of Medicine and director of the Yale Center for Outcomes Research and Evaluation (CORE), one of the nation's first and most productive research units dedicated to producing innovations to improve patient outcomes and promote better population health. He directed the Robert Wood Johnson Foundation Clinical Scholars Program at Yale, which prepared physicians to become future leaders in health care, from 1996–2017.

Dr. Krumholz has been honored by membership in the National Academy of Medicine, the Association of American Physicians, and the American Society for Clinical Investigation. He was named a Distinguished Scientist of the American Heart Association and founded the organization's Quality of Care and Outcomes Research Council annual conference. He was a member of the Advisory Committee to the Director of the National Institutes of Health and was a founding Governor of the Patient-Centered

Outcomes Research Institute. He is the founder of HugoHealth, a patient-centric platform to engage people as partners in research and facilitate the secure movement of digital health data, co-founder of Refactor Health, an enterprise healthcare AI-augmented data management company, and a founder of medRxiv, a preprint server for the medical and health sciences. He received the Friendship Award from the People's Republic of China in recognition of his collaborative efforts to develop a national cardiovascular research network.
Dr. Krumholz was the founding editor of *Circulation: Cardiovascular Quality and Outcomes* and was editor of *CardioExchange*, a social media site of the publisher of the *New England Journal of Medicine*. He has published more than 1000 articles, has a regular blog on Forbes.com, and has contributed to the *New York Times Wellness* blog, the *New York Times* op-ed page, and *National Public Radio Shots* blog.

REBECCA LI, PH.D., is the Executive Director of Vivli and the Co-director of the Research ethics program at the Center for Bioethics at the Harvard Medical School. Previous to her current role she was the Executive Director of the MRCT Center of Brigham and Women's Hospital and Harvard for over 5 years and remains a Senior Advisor at the Center. She has over 25 years of experience spanning the entire drug development process with experience in Biotech, Pharma and CRO environments. She completed a Fellowship in 2013 in the Division of Medical Ethics at Harvard Medical School. She earned her PhD in Chemical and Biomolecular Engineering from Johns Hopkins University.

TIANJING LI, PH.D, is an Associate Professor in the Department of Ophthalmology at University of Colorado Anschutz Medical Campus. She also holds an adjunct Associate Professor position in the Department of Epidemiology at Johns Hopkins Bloomberg School of Public Health. The primary goal of Dr. Li's research is to develop, evaluate, and disseminate efficient methods for comparing healthcare interventions and to provide trust-worthy evidence for decision-making. Dr. Li has worked with Cochrane in various capacities for 15 years. She served as the Associate Director for Cochrane United States from 2012 to the Center's closure in 2018. Currently, in addition to her role as a Coordinating Editor for Cochrane Eyes and Vision, she co-convenes the Cochrane Comparing Multiple Interventions Methods Group. She is an Associate Scientific Editor for the 2nd edition of the Cochrane Handbook for Systematic Reviews of Interventions. Outside of Cochrane, Dr. Li serves as a Co-Editor-in-Chief for the journal *Trials* and the Reviews Editor for *JAMA Ophthalmology*. She is an elected member of the Society for Research Synthesis Methodology and was awarded the Society's inaugural Early Career Award in 2016. She received the Anne Anderson Award in 2019, being recognized for her cumulative accomplishment, originality and independence of thought, leadership, and inspiration for women within Cochrane.

BILL LOUV, PH.D., joined Project Data Sphere in March 2018 as President. Bill held key leadership positions in the pharmaceutical industry for nearly 30 years. He joined the pharmaceutical industry in 1986 as head of biostatistics at Merrell Dow and advanced to the position of VP of biostatistics, epidemiology and clinical data management at GlaxoWellcome in 1998. Bill made a significant career change in 1999 when he became VP of IT for GlaxoSmithKline's R&D organization. Bill was named Chief Information Officer for GSK in 2007. In 2011, Bill was promoted to Senior VP of Core Business Services which included IT, Procurement, Accounting, and Real Estate. Bill was a member of GSK's Corporate Executive Management team from 2007 until his retirement in May, 2016. Bill has consulted with many organizations on health care analytics and opportunities to leverage big data. He is a Non-Executive Director of River Logic Inc., a leader in Prescriptive Analytics and Integrated Business Planning, and he is Deputy Chairman of ClinPal, a cloud-based clinical trial platform. In his early career Bill was a member of technical staff at Bell Laboratories where he developed forecasting algorithms for signaling networks. Subsequently he was Associate Professor of Biostatistics at the University of Alabama at Birmingham. Bill published more than 25 academic papers while at these research organizations.

AMY NURNBERGER, M.S., is the Program Head for Data Management Services at MIT libraries. Amy also serves as an Adjunct Assistant Professor in the Learning Analytics program at Teachers College, Columbia University. Within the broader research data community, Amy is a co-chair of the Research Data Alliance (RDA)/World Data Services Publishing Data Workflows Working Group and the RDA Education and Training on Handling Research Data Interest Group, and is the elected co-chair of the RDA Organisational Advisory Board. She also sits on the ACRL Research and Scholarly Environment Committee and the editorial board of Patterns - The Science of Data. Prior to entering the field of research data management, Amy held positions in the hospital and pharmaceutical industries.

DEBORAH C. PEEL, M.D., is the Founder & President of Patient Privacy Rights and the world's leading advocate for patients' rights to control the use of personal health information in electronic systems. She is also a practicing physician and Freudian psychoanalyst. She became an expert and privacy warrior to stop patients from being harmed. The lack of medical privacy causes millions of US citizens to avoid early diagnosis and treatment for cancer, depression, and STIs every year. Her passion is informing the public about privacy-enhancing technologies and the major fixes needed in law and policy, so they can join the battle to restore our civil and human rights to health privacy. In 2004, she formed Patient Privacy Rights (PPR), a 501c3 non-profit organization to educate Americans about the urgent need to restore patient control over health data. In 2006, Dr. Peel founded the bipartisan Coalition for Patient Privacy. The coalition includes over 50+ national organizations, representing 10.3 million people who want to control the use of personal health information. In 2007, Microsoft Corporation joined the Coalition. Since 2007, Dr. Peel has been included on ModernHealthcare magazine's "100 Most Influential in Healthcare" list four times. In 2013, she was named one of the "Top Ten Influencers in Health InfoSec" by HealthcareInfoSecurity. Dr. Peel is the catalyst and creator of the annual International Summits on the Future of Health Privacy. The summits are the only venue where national and international experts from advocacy, academia, government, and industry come together and debate urgent threats to health privacy and realistic solutions.

JOSEPH ROSS, M.D., M.H.S., is a Professor of Medicine (General Medicine) and of Public Health (Health Policy and Management) at the Yale School of Medicine, a member of the Center for Outcomes Research and Evaluation at Yale-New Haven Health System, and Co-Director of the National Clinician Scholars Program at Yale. With expertise in performance measure development and the translation of clinical research into practice, his research examines the use and delivery of higher quality care and issues related to pharmaceutical and medical device regulation, evidence development, postmarket surveillance, and clinical adoption. Dr. Ross co-directs the Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation (CERSI), the Yale Open Data Access (YODA) Project, and the Collaboration for Research Integrity and Transparency (CRIT) at Yale Law School. He has published more than 400 articles in peer-reviewed biomedical journals and is currently an Associate Editor at JAMA Internal Medicine.

SCOTT SHAUNESSY is the Chair at ClinicalStudyDataRequest.com as well as the CEO at ideaPoint- an Anaqua Company. He has over 20 years of experience as an executive in both Business-to-Business and Consumer-facing businesses and is an expert in best practices for Business Development and Licensing, Venture Investing, R&D Collaboration and Clinical Trial Data Transparency processes in the Pharmaceutical, Medical Device, Biotech, Nutrition, Healthcare and Consumer Products industries. As Chairman of ClinicalTrialDataRequest.com (CSDR), Scott is involved in the day to day operations of CSDR and an industry thought leader on this subject. Before his management career, Scott spent eight years as a professional hockey player in the National Hockey League and American Hockey League.

IDA SIM, M.D, PH.D., is Professor of Medicine at the University of California, San Francisco and co-directs Informatics and Research Innovation at UCSF's Clinical and Translational Sciences Institute. She

is co-founder of two non-profit organizations: Vivli and Open mHealth. Dr. Sim earned her MD and PhD in Medical Informatics from Stanford University, where her dissertation was on computational methods for data sharing of clinical trial results. She Primary Care Internal Medicine residency at the Massachusetts General Hospital, and fellowships in general medicine and medical informatics at Stanford.

In 2005-6, Dr. Sim led the World Health Organization's International Clinical Trials Registry Platform which established the first global policy on clinical trial registration and defined the common 20-item Trial Registration Data Set. She was a member of the 2015 Institute of Medicine committee on "Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk." In 2017, Dr. Sim co-founded Vivli – a global neutral data sharing platform for participant-level clinical trial data – and continues to serve as Vivli's Principal Technical Advisor.

Dr. Sim is also a national leader in mobile health and co-founder of Open mHealth, a non-profit organization building open APIs and tools for integrating mobile health data. She has served on multiple national advisory committees on health information infrastructure for clinical care and research. She is a recipient of the United States Presidential Early Career Award for Scientists and Engineers (PECASE), a Fellow of the American College of Medical Informatics, and a member of the American Society for Clinical Investigation. She is a practicing primary care clinician.

MATTHEW SYDES, M.SC., is a Professor in Clinical Trials and Methodology at the Medical Research Council Clinical Trials Unit at UCL (University College London). Matthew is responsible for leading the unit's Trial Conduct Methodology activities as well as teams conducting research in prostate cancer and osteosarcoma. Matthew has a particular interest in improving clinical trial conduct, particularly around: the use of routinely-collected electronic health records (EHR) to support and run trials with Health Data Research UK; running trials with a view to regulatory use and submission; proportionate and efficient trial monitoring; clinical trial data sharing; communication of trial findings; adaptive and efficient designs for late-phase trials, including in uncommon conditions; and the functioning of Data Monitoring Committees. On this, he was involved in the DAMOCLES project which set standards for (independent) Data Monitoring Committees and led to the widespread use of charters for trial committees. He is part of the faculty for UCL's regular 1-day course on Data Monitoring Committees in practice. Matthew has served as a member of IDMCs and TSCs, often as chair, for more around 50 trials, attending around 200 meetings. Matthew teaches on these topics on the UCL Institute of Clinical Trials and Methodology's MSc in Clinical Trials. He is supervising a number of PhD students in these areas of methodological priority.

In terms of clinical trials in prostate cancer, Matt has served as lead statistician for STAMPEDE, a multi-arm multi-stage platform protocol for a clinical trial which has repeatedly reported practice defining results for adding chemotherapy, abiraterone and radiotherapy in both the metastatic and non-metastatic setting, the latter building on the MRC PR07 trial. The efficient and larger RADICALS trial for post-operative treatment will soon report results on multiple clinical questions for men with early prostate cancer. In osteosarcoma, the trans-Atlantic EURAMOS-1 trial was is the largest RCT conducted in this rare disease, thanks to an extensive international collaboration and has defined the standard-of-care. Matt commonly advises on trial design and conduct of external trials.

Externally, Matt chairs the Scientific Committee for the International Clinical Trials and Methodology Conference (ICTMC) 2019, and serves on the executive committee for the MRC Network of Hubs for Trial Methodology Research and its successor Partnership for Trials Methodology Research. He is currently a member of Cancer Research UK's Clinical Trials Monitoring Panel, having previously served on main clinical trials funding body. He spent 7 years serving on a national Research Ethics Committee.

He reviews grant applications and articles for a wide range of funding bodies and journals. He is always keen to share experiences with other researchers and has presented widely. He has served on the faculty for clinical trials and statistic courses, including co-directing the EORTC's annual “Statistics for Non-statisticians” course, and serving on the faculty of ECCO-AACR-EORTC-ESMO's (“Zeist” / "Flims") Workshop on Methods in Clinical Cancer Research, and at 3 of ASCO's International Clinical Trials Workshops joint with national oncology organisations in Turkey, Greece and Morocco.

Matt joined the MRC Cancer Trials Office (in Cambridge) back in 1995, moving to London with the formation of the MRC Clinical Trials Unit in 1998 and on into UCL in 2013. He chairs the Unit’s Protocol Review Committee, led the development of a new template protocol for clinical trials and is a member of the organising committee for the Unit’s series of Monday Lunchtime Seminars, and is a co-opted member of UCL’s Research Ethics Committee.

MOSES TAYLOR, JR., P.E., Moses Taylor is a Senior Associate at Structural Integrity Associates, Inc., an engineering consulting firm headquartered in San Jose, CA. The company offers expert consultation in the prevention and control of structural and mechanical failures. As Senior Advisor, he supports the management and technical staff with implementation of the project management systems and oversees the activities associated with a group of clients as Client Manager. Mr. Taylor also manages large projects in emerging product areas such as seismic qualification of large power plant components that require multi-discipline teams, demanding schedules, and intensive client interaction.

Mr. Taylor received his Bachelor of Engineering degree in Mechanical Engineering from Vanderbilt University and an MBA in Financial Management from American University in Washington, D.C. His undergraduate career included service in the Naval ROTC program which led to a commission in the U.S. Navy where he began his career as a nuclear engineer at the Naval Nuclear Propulsion Program headquarters in Washington, D.C. From there, his career transitioned to engineering consulting firms serving the commercial power industry. With more than 35 years of experience, his role has been primarily as a project manager for a variety of projects for domestic and international utilities and the U.S. Department of Energy. He has successfully managed applied research and development projects, turnkey design, licensing, fabrication and installation projects and field modification and testing projects. As a manager of project managers, Mr. Taylor developed project management systems and procedures and provided project management consulting services to the Safeguards Division of the International Atomic Energy Agency (IAEA) in Vienna, Austria.

Mr. Taylor is a Licensed Professional Engineer and a member of the American Society of Mechanical Engineers and the Project Management Institute. A native of Nashville, TN, he now resides in Rocklin, CA with his wife. He has two adult children and three adorable grandchildren. Mr. Taylor is active in his local church and enjoys singing locally and internationally in an award winning 100+ member choral society and orchestra. He enjoys local getaways, international travel and walking for pleasure and exercise.

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the nation • Improving health

For more information visit www.iom.edu/datasharing

Sharing Clinical Trial Data

Maximizing Benefits, Minimizing Risk

Although clinical trials generate vast amounts of data, a large portion is never published or made available to other researchers. Data sharing could advance scientific discovery and improve clinical care by maximizing the knowledge gained from data collected in trials, stimulating new ideas for research, and avoiding unnecessarily duplicative trials. But data sharing also entails significant risks, burdens, and challenges. Policies are needed to protect the privacy of participants, the investment of funders and sponsors, the academic recognition of investigators, and the validity of analyses, among other concerns.

With support from 23 public- and private-sector sponsors in the United States and abroad, the Institute of Medicine (IOM) assembled a committee to develop guiding principles and a practical framework for the responsible sharing of clinical trial data. In its report, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*, the committee concludes that sharing data is in the public interest, but a multi-stakeholder effort is needed to develop a culture, infrastructure, and policies that will foster responsible sharing—now and in the future.

Data sharing could advance scientific discovery and improve clinical care by maximizing the knowledge gained from data collected in trials, stimulating new ideas for research, and avoiding unnecessarily duplicative trials.

Guiding Principles for Sharing Clinical Trial Data

The ultimate goal of data sharing should be to increase scientific knowledge, leading to better therapies for patients. With this goal in mind, the IOM committee presents the following guiding principles for responsible sharing of clinical trial data:

- Maximize the benefits of clinical trials while minimizing the risks of data sharing.
- Respect individual participants whose data are shared.

- Increase public trust in clinical trials and the sharing of trial data.
- Conduct the sharing of trial data in a fair manner.

These principles should be balanced in the context of specific trials and stakeholder needs, including concerns about the potential harms and costs of data sharing.

Collaboration among a broad set of stakeholders is needed to create a culture in which responsible data sharing is incentivized and best practices are disseminated widely. To foster such a culture, including strategies to develop infrastructure, oversight, and sustainability, the IOM report details specific actions for funders and sponsors of clinical trials; disease advocacy organizations; regulatory and research oversight agencies; research ethics committees or institutional review boards; investigators; research institutions and universities; academic journals; and membership and professional societies.

Optimal Timing for Data Sharing

There are many different types of data generated during the course of a clinical trial, including individual participant data (including raw data or the analyzable dataset); metadata (for example, trial protocol, statistical analysis plan, analytic code); and summary-level data (for example, lay summaries and clinical study reports). Sharing each type carries different benefits, risks, and challenges. For example, making the analyzable dataset available to researchers allows reanalysis and replication of trial results, but could lead to privacy concerns and inappropriate use. Furthermore, the analyzable dataset must be accompanied by metadata to ensure that secondary analyses are rigorous and efficient. Taking into account these and other considerations, the IOM committee identifies the optimal stage in the clinical trial lifecycle at which each data type should be shared, and under what conditions.

Decisions about the timing of data sharing should balance several goals:

1. allow a fair opportunity for clinical trialists to publish results before secondary investigators gain access to the data;
2. allow secondary investigators to access unpublished trial data after a fair period has passed or reproduce the findings of a published analysis; and
3. protect the commercial interests of sponsors in gaining regulatory approval for a product so that they receive fair financial rewards for their investment.

The IOM committee acknowledges the importance of allowing ample time after the completion of a trial for original investigators to complete their analyses; however, the committee concludes that this period should extend no longer than 18 months. When that period has passed—regardless of whether the trial results have been published—the IOM committee finds that the scientific process is best served by allowing other investigators to access the data. However, if the trial is part of a submission to a regulatory agency for approval, an exception should be made, and the data should be shared no later than 30 days after regulatory approval or 18 months after product abandonment.

When trial findings are published before the 18-month period has passed, the committee recommends that the supporting analytic dataset be shared within 6 months of publication. Although many practical constraints currently prevent the release of the analytic dataset simultaneously with publication, the committee expresses its hope that, as systems for responsible data sharing evolve, simultaneous sharing will become the standard.

Due to the wide variation in clinical trial types, the IOM committee recognizes that there will be necessary exceptions to its timing recommendations. These recommendations are meant to be professional standards rather than inflexible rules. In some cases, it may be appropriate to share data later than recommended; in others—particu-

Collaboration among a broad set of stakeholders is needed to create a culture in which responsible data sharing is incentivized and best practices are disseminated widely.

larly for trials likely to have major clinical, public health, or policy implications—it may be best to share data sooner. It is important to note that the committee's data sharing recommendations do not apply to trials that are already complete, or "legacy" trials. Decisions to share legacy data should be made on a case by case basis, although the committee urges sponsors and investigators to prioritize the sharing of data from legacy trials whose findings influence decisions about clinical care.

Access to Clinical Trial Data

Many of the risks associated with sharing clinical trial data may be mitigated by controlling which parties can access data and under what conditions. Policies for granting access to data should be in the service of several goals—protecting the privacy of participants; reducing risk of invalid analyses or misuse; avoiding undue burdens on data users and harm to investigators and sponsors; and enhancing public trust in clinical trial data sharing.

The committee believes that open, public access to clinical trial data is appropriate for sharing clinical trial results and may be desirable for sharing other types of data when all stakeholders—sponsors, investigators, and participants—are comfortable and believe the benefits outweigh the risks. But in many cases, stakeholders may have concerns about granting open access, including risks to privacy and security. A number of provisions could help assuage such concerns, including de-identification and data use agreements. Case-by-case reviews of data access requests could mitigate risks but may inhibit valid secondary analyses and stifle innovation if too restrictive. Reviews should be conducted by independent panels that include representatives from community, patient, and disease advocacy groups and should ensure transparent policies and procedures. Finally, the committee urges stakeholders to share lessons and best practices for data access policies as data sharing practices evolve.

The Future of Clinical Trial Data Sharing

Although increased data sharing holds promise for scientific advancement, significant barriers remain. The IOM committee identifies several key challenge areas:

- *Infrastructure*: Currently, there are insufficient platforms to efficiently store and manage the breadth of trial data.
- *Technology*: At present, data sharing platforms are not consistently discoverable, searchable, or interoperable.
- *Workforce*: The clinical trials ecosystem lacks an adequate workforce to manage the operational and technical aspects of data sharing.
- *Sustainability*: For a system of data sharing to be sustainable, costs will need to be distributed equitably across both generators and users of data.

The committee outlines a conceptual business model for sustainable and equitable data sharing.

Because data sharing benefits multiple stakeholders—including the public, insurers, health care providers, and researchers—all of these stakeholders should also bear some of the costs of the data sharing enterprise. Additional sources of funding, such as philanthropy, should be explored. Finally, the committee notes an ongoing need for accurate measurements of the costs of data sharing.

In order for responsible data sharing to become pervasive, sustained, and rooted as a professional norm, these and other challenges will have to be addressed collaboratively by diverse institutions and stakeholders. To promote discussion and exchange of ideas among these groups and to foster agreement around best practices, standards, and incentives, the committee recommends the formation of a global, multi-stakeholder body to address current and future challenges.

Conclusion

Clinical trials are essential to determining the safety and efficacy of new health treatments, but limited data sharing prevents maximum utilization of knowledge gained. In short, the current system fails to provide an adequate return on the investments of trial participants, investigators, and sponsors. Greater data sharing could enhance public well-being by accelerating the drug discovery and development process, reducing redundant research, and facilitating scientific innovation. Before these benefits can be realized, however, stakeholders must confront significant risks and challenges. In *Sharing Clinical Trial Data*, the IOM committee provides a practical and ethical framework to help stakeholders navigate this complex terrain.

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the nation • Improving health

500 Fifth Street, NW
Washington, DC 20001
TEL 202.334.2352
FAX 202.334.1412

www.iom.edu

The Institute of Medicine serves as adviser to the nation to improve health. Established in 1970 under the charter of the National Academy of Sciences, the Institute of Medicine provides independent, objective, evidence-based advice to policy makers, health professionals, the private sector, and the public.

Sharing Clinical Trial Data

MAXIMIZING BENEFITS, MINIMIZING RISK

Recommendation at a Glance: When to Share Data

Rationale for Responsible Sharing of Clinical Trial Data

Clinical trials play a crucial role in advancing medical innovation and represent a significant investment from all involved — including trial participants, sponsors, and researchers. Data are generated throughout the clinical trial lifecycle, but results are often not published in a timely manner, and many data are not shared beyond the original investigators.

Data sharing could advance scientific discovery and improve clinical care by maximizing knowledge gained from data collected in trials, stimulating new ideas for research, and avoiding unnecessarily duplicative trials; however, to reduce potential harms, policies are needed to protect the privacy and consent of participants, the validity of analyses, the investment of funders and sponsors, and the academic recognition of investigators.

To answer this need, an Institute of Medicine consensus study recommends guiding principles and a practical framework to enhance clinical trial **data sharing**, the practice of making data from scientific research available—with or without restrictions—for **secondary uses**, which include re-analyses, new analyses and meta-analyses. This brochure focuses exclusively on the committee's **recommendation for when to share specific types of data.**

There are three types of data that should be shared:

SUMMARY DATA

Data commonly generated based on analysis of the individual participant data from a clinical trial (e.g., summary-level results posted on registries, lay summaries, publications, and clinical study reports (CSRs) used for regulatory application)

INDIVIDUAL PARTICIPANT DATA (IPD)

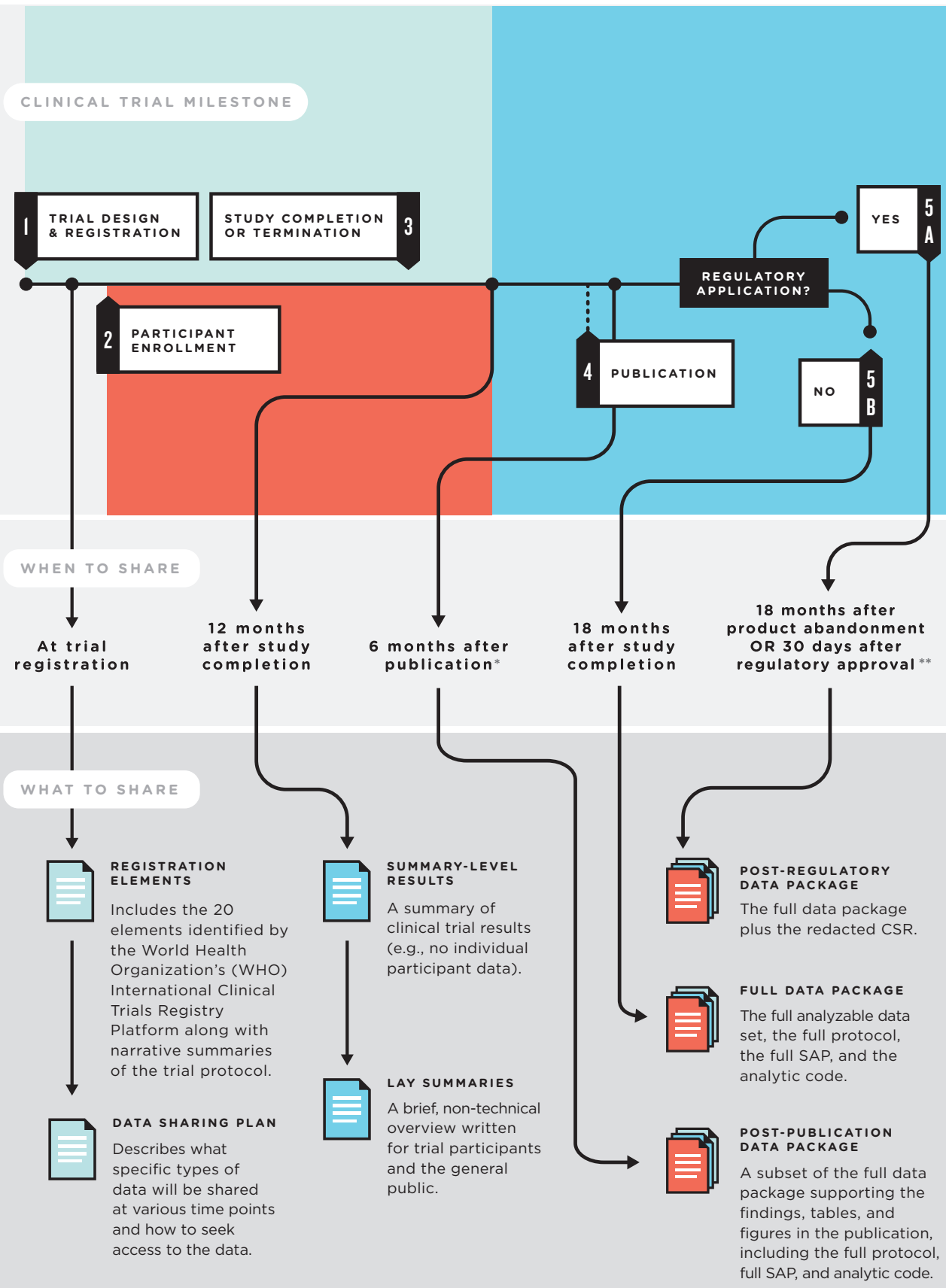
Data that are collected from participants (e.g., the raw data) and then cleaned, abstracted, coded, and transcribed to become the analyzable data

METADATA

"Data about the data" (e.g., protocol, statistical analysis plan (SAP), and analytic code)

What data should be shared and *when* during the clinical trial lifecycle in order to help amplify scientific knowledge worldwide while minimizing risk?

The following chart outlines the **major stages of the clinical trial lifecycle** and **recommends when to share specific data packages** in common scenarios.



* No later than 6 months after publication applies to all studies, whether intended and or not intended to support regulatory applications and regardless of the timing of publication relative to study completion, although publication is most likely to occur after study completion.

** Sharing of the post-regulatory data package should occur: 30 days after approval or 18 months after study completion, whichever is later; 18 months after abandonment of the product or indication. This applies to all studies intended and to support regulatory applications, even if abandonment occurs prior to actual regulatory application.

Download the full report at *www.iom.edu/datasharing.*

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES
Advising the nation • Improving health

INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Advising the nation • Improving health

For more information visit www.iom.edu/datasharing

Sharing Clinical Trial Data

Maximizing Benefits, Minimizing Risk

Recommendations

RECOMMENDATION 1: Stakeholders in clinical trials should foster a culture in which data sharing is the expected norm, and should commit to responsible strategies aimed at maximizing the benefits, minimizing the risks, and overcoming the challenges of sharing clinical trial data for all parties.

Funders and sponsors should

- promote the development of a sustainable infrastructure and mechanism by which data can be shared, in accordance with the terms and conditions of grants and contracts;
- provide funding to investigators for sharing of clinical trial data as a line item in grants and contracts
- include prior data sharing as a measure of impact when deciding about future funding;
- include and enforce requirements in the terms and conditions of grants and contracts that investigators will make clinical trial data available for sharing under the conditions recommended in this report; and
- fund and promote the development and adoption of common data elements.

Disease advocacy organizations should

- require data sharing plans as part of protocol reviews and criteria for funding grants;
- provide guidance and educational programs on data sharing for clinical trial participants;
- require data sharing plans as a condition for promoting clinical trials to their constituents; and
- contribute funding to enable data sharing.

Regulatory and research oversight bodies should

- work with industry and other stakeholders to develop and harmonize new clinical study report (CSR) templates that do not include commercially confidential information or personally identifiable data;
- work with regulatory authorities around the world to harmonize requirements and practices to support the responsible sharing of clinical trial data; and

- issue clear guidance that the sharing of clinical trial data is expected, and that the role of Research Ethics Committees or Institutional Review Boards (IRBs) is to encourage and facilitate the responsible and ethical conduct of data sharing through the adoption of protections such as those recommended by this committee and the emerging best practices of clinical trial data sharing initiatives.

Research Ethics Committees or IRBs should

- provide guidance for clinical trialists and templates for informed consent for participants that enable responsible data sharing;
- consider data sharing plans when assessing the benefits and risks of clinical trials; and
- adopt protections for participants as recommended by this committee and the emerging best practices of clinical trial data sharing initiatives.

Investigators and sponsors should

- design clinical trials and manage trial data with the expectation that data will be shared;
- adopt common data elements in new clinical trial protocols unless there is a compelling scientific reason not to do so;
- explain to participants during the informed consent process
  - what data will (and will not) be shared with the individual participants during and after the trial,
  - the potential risks to privacy associated with the collection and sharing of data during and after the trial and a summary of the types of protections employed to mitigate this risk, and
  - under what conditions the trial data may be shared (with regulators, investigators, etc.) beyond the trial team; and
- make clinical trial data available at the times and under the conditions recommended in this report.

Research institutions and universities should

- ensure that investigators from their institutions share data from clinical trials in accordance with the recommendations in this report and the terms and conditions of grants and contracts;
- promote the development of a sustainable infrastructure and mechanisms for data sharing;
- make sharing of clinical trial data a consideration in promotion of faculty members and assessment of programs; and
- provide training for data science and quantitative scientists to facilitate sharing and analysis of clinical trial data.

Journals should

- require authors of both primary and secondary analyses of clinical trial data to
  - document that they have submitted a data sharing plan at a site that shares data with and meets the data requirements of the World Health Organization's International Clinical Trials Registry Platform before enrolling participants, and
  - commit to releasing the analytic data set underlying published analyses, tables, figures, and results no later than the times specified in this report;
- require that submitted manuscripts using existing data sets from clinical trials, in whole or in part, cite these data appropriately; and
- require that any published secondary analyses provide the data and metadata at the same level as in the original publication.

Membership and professional societies should

- establish policies that members should participate in sharing clinical trial data as part of their professional responsibilities;
- require as a condition of submitting abstracts to a meeting of the society and manuscripts to the journal of the society that clinical trial data will be shared in accordance with the recommendations in this report; and
- collaborate on and promote the development and use of common data elements relevant to their members.

RECOMMENDATION 2: Sponsors and investigators should share the various types of clinical trial data no later than the times specified below. Sponsors and investigators who decide to make data available for sharing before these times are encouraged to do so.

Trial registration:

- The data sharing plan for a clinical trial (i.e., what data will be shared when and under what conditions) should be publicly available at a third-party site that shares data with and meets the data requirements of WHO's International Clinical Trials Registry Platform; this should occur before the first participant is enrolled.

Study completion:

- Summary-level results of clinical trials (including adverse event summaries) should be made publicly available no later than 12 months after study completion.
- Lay summaries of results should be made available to trial participants concurrently with the sharing of summary-level results, no later than 12 months after study completion.
- The full data package (including the full analyzable data set, the full protocol,[1] the full statistical analysis plan, and the analytic code) should be shared no later than 18 months after study completion (unless the trial is in support of a regulatory application).

Publication:

- The post-publication data package (including the subset of the analyzable data set supporting the findings, tables, and figures in the publication and the full protocol, full statistical analysis plan, and analytic code that supports the published results) should be shared no later than 6 months after publication.

Regulatory application:

- For studies of products or new indications that are approved, the post-regulatory data package (including the full analyzable data set and clinical study report redacted for commercially or personal confidential information, together with the full protocol, full statistical analysis plan, and analytic code) should be shared 30 days after regulatory approval or 18 months after study completion, whichever occurs later.

[1] Includes the protocol in place at the start of the trial, any modifications, and the final protocol.

- For studies of new products or new indications for a marketed product that are abandoned, the post-regulatory data package should be shared no later than 18 months after abandonment. However, if the product is licensed to another party for further development, these data need be shared only after publication, approval, or final abandonment.

RECOMMENDATION 3: Holders of clinical trial data should mitigate the risks and enhance the benefits of sharing sensitive clinical trial data by implementing operational strategies that include employing data use agreements, designating an independent review panel, including members of the lay public in governance, and making access to clinical trial data transparent. Specifically, they should take the following actions:

- Employ data use agreements that include provisions aimed at protecting clinical trial participants, advancing the goal of producing scientifically valid secondary analyses, giving credit to the investigators who collected the clinical trial data, protecting the intellectual property interests of sponsors, and ultimately improving patient care.
- Employ other appropriate techniques for protecting privacy, in addition to de-identification and data security.
- Designate an independent review panel—in lieu of the sponsor or investigator of a clinical trial—if requests for access to clinical trial data will be reviewed for approval.
- Include lay representatives (e.g., patients, members of the public, and/or representatives of disease advocacy groups) on the independent review panel that reviews and approves data access requests.
- Make access to clinical trial data transparent by publicly reporting
  - the organizational structure, policies, procedures (e.g., criteria for determining access and conditions of use), and membership of the independent review panel that makes decisions about access to clinical trial data; and
  - a summary of the decisions regarding requests for data access, including the number of requests and approvals and the reasons for disapprovals.
- Learn from experience by collecting data on the outcomes of data sharing policies, procedures, and technical approaches (including the benefits, risks, and costs), and share information and lessons learned with clinical trial sponsors, the public, and other organizations sharing clinical trial data.

RECOMMENDATION 4: The sponsors of this study should take the lead, together with or via a trusted impartial organization(s), to convene a multistakeholder body with global reach and broad representation to address, in an ongoing process, the key infrastructure, technological, sustainability, and workforce challenges associated with the sharing of clinical trial data.

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

JUNE 8, 2017

Bridging the Data-Sharing Divide — Seeing the Devil in the Details, Not the Other Camp

Lisa Rosenbaum, M.D.

The movement toward sharing data from clinical trials has divided the scientific community, and the battle lines were evident at a recent summit sponsored by the *Journal*. On one side stand many clinical trialists, whose lifeblood — randomized, controlled trials (RCTs) — may be threatened by data sharing. On the other side stand data scientists — many of them hailing from the genetics community, whose sharing of data markedly accelerated progress in that field.

At a time when RCT funding is shrinking, trialists know that sharing data adds substantial costs to clinical trial execution; a requirement to share data might mean that fewer trials, and smaller ones, will be conducted. Many trialists also worry that complex data will be misinterpreted by people who weren't involved in generating them, and who may therefore produce misleading results. Furthermore, journal publications are the currency of academic advancement. Researchers often invest 5 to 10 years gathering trial data, expecting to write several papers after their primary publication. An expectation that data will be shared quickly may therefore create a disincentive for conducting RCTs.

Data scientists promoting data sharing are joined by some members of the medical community, who point to abundant unpublished studies with negative results as missed learning opportunities and invitations to wasteful repetition of trials. Some proponents see resistance to data sharing as motivated purely by self-interest. As Isaac Kohane, chair of the Department of Biomedical Informatics at Harvard Medical School, recounted, when geneticists began aggregating their data, there were notable holdouts who, fearing being scooped, withheld data and slowed the community's progress. Yet as Ewan Birney, a geneticist who codirects the European BioInformatics Institute, noted at the summit, "once everyone has done it for a little bit of time, you will forget you had these arguments."

And everyone may have to do it soon. The National Institutes of Health now requires that grant applicants outline a data-sharing plan, as do the Cancer Moonshot, the Gates Foundation, and the Wellcome Trust. But many details need to be worked out, from incentive structures to sustain data generation, to standards for data exchange, to identification of the subset of clinical questions for which sharing is most cost-effec-

tive. Indeed, the focus of the data-sharing summit was less about whether to share data and more about how best to do so.

Perhaps the most incisive question, posed by Rory Collins, a University of Oxford epidemiologist and trialist, was the most obvious one: What problem are we trying to solve? The advancement of science depends on the open exchange of ideas and the opportunity to replicate or refute others' findings. But will data sharing address our current system's shortcomings in a way that advances science? For example, though it's troubling when trials provide incomplete information about adverse events, requiring the sharing of individual patient data from every trial might not be the best way to fix that problem. A more effective solution, Collins suggests, may be publishing, alongside the primary trial results, an easily accessible appendix containing adverse-event data in tabular form. Proponents of data sharing also believe it will allow other investigators to generate new insights and hypotheses. But will such insights advance health in a way that justifies the cost?

Preliminary evidence reveals less enthusiasm than anticipated for using shared RCT data. In 2007, GlaxoSmithKline created the website clinicalstudydatarequest.com (CSDR), where data from at least 3049 trials are currently available, from 13 industry sponsors. According to the independent panel that reviewed research proposals, in the first 2 years, 177 proposals were submitted, most of them for a new study and publication, but despite substantial investment by industry sponsors, only four manuscripts have been submitted for publication thus far.[1] Brian Strom, one of the panel members, noted that because industry analyzes its data so exhaustively in anticipation of intensive interrogation from the Food and Drug Administration, it's possible that nonindustry data will yield more new findings. But industry's resources also far exceed academia's, so relatively speaking, data sharing's costs for academics will be far greater.

The more substantial challenge described by investigators seeking to use others' data, however, seems to be the burdensome nature of analyzing data behind a firewall. Rather than receiving data to analyze themselves, investigators submitted to CSDR statistical inquiries that were run by the repository's managers, a time-consuming process. The consensus, according to Strom, was that "true data sharing would be preferable to data access on a dedicated website."

So what is true data sharing? Does the work required to overcome these challenges, or the outputs achieved, differ when sharing is imposed from the outside rather than motivated by prospectively identified common goals? Clinical investigators, after all, have long collaborated in efforts to address unanswered questions. More than 30 years ago, for instance, Collins led the creation of the Clinical Trial Service Unit, a global consortium of trialists who sought to share data and pool their results. Though it required a tremendous time investment and endless communication among investigators to understand each other's data sets, there was a shared sense of purpose and pride in the clinically meaningful results. For instance, though it was believed that tamoxifen reduced recurrence but did not improve survival among women with breast cancer, through a planned combined analysis with longer follow-up and more patients, the group found that there was a survival benefit — transforming the standard of care.[2]

Collins, therefore, firmly believes in data sharing's benefits, but he recommends considering all the likely hitches. Comparing the relative ease of data sharing in 1995, when a cholesterol-treatment meta-analysis was prospectively planned,[3] to current "clunky" and more time-consuming processes, Collins notes, "It is ironic that as a result of the data-sharing agenda and the formalization of the systems, it is now more difficult to get access to the data." The aggregation of vast genetics databases suggests that these technical and bureaucratic challenges are growing pains. But clinical trial data sets may be sufficiently complex that streamlining data exchange will require extensive input from the data's generators. Ideally, the trialist community will create uniform standards and data will be collected with those standards in mind.

A greater risk may be to the clinical trialist community. It's assumed that data sharing will advance the public health, but will the public benefit if there are steep declines in the number and size of clinical trials? Though more "open" science may yield as-yet-unimagined innovations, unplanned and retrospective secondary analyses can only generate, not test, hypotheses. The type of hypothesis testing that can advance treatment of disease will always depend on active and motivated clinical trialists asking questions prospectively.

And though tinkering with data-exclusivity periods and designations of academic credit may

reduce the disincentives created by data sharing, I think there is something at stake here that incentives can't solve: our capacity to rationally weigh trade-offs as we debate how best to advance science. While the recent summit was civil and collaborative, the tenor of the broader data-sharing conversation has framed the matter as one of trialist self-interest versus public good. But such a frame vastly oversimplifies the situation — and tends to entrench people in polarized positions, articulated with righteous indignation.

The indignation of data-sharing advocates arises in part from the claim that the absence of data sharing slows the development of cures. In addition, at a political moment when promises of data democratization overshadow faith in traditional expertise, reservations about data sharing are easily dismissed as elitist — as are the experts who point out misunderstandings of a topic they've spent years studying. The value placed on transparency also contributes: any resistance to greater openness is branded as secrecy and deceit. Finally, the deepest (and perhaps most valid) source of moral outrage may be the sentiment that clinical trial data aren't ours to begin with, that they should belong to the patients who put themselves at risk to participate. And in principle, patients want their data shared.

An audio interview with Dr. Drazen is available at NEJM.org

But patients also want better treatments for their diseases. And though data sharing may sometimes lead to better treatments, it may also divert limited resources to types of research that are less fruitful than RCTs, impeding the evidence generation required for improving care. The irony in the framing of this debate is that to share data in a way that advances knowledge, we must be open to one another's experience and expertise, setting aside ideology in pursuit of more objective truths. Fulfilling this obligation, as we refine the scientific process, will require not only sharing what we find but also resisting the temptation to demonize those who see different paths to our shared goal.

Disclosure forms provided by the author are available at NEJM.org.

Dr. Rosenbaum is a national correspondent for the *Journal*.

This article was published on April 26, 2017, at NEJM.org.

1. Strom BL, Buyse ME, Hughes J, Knoppers BM. Data sharing — is the juice worth the squeeze? N Engl J Med 2016;375:1608-9.
2. Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer: an overview of 61 randomized trials among 28,896 women. N Engl J Med 1988;319:1681-92.
3. Cholesterol Treatment Trialists' (CTT) Collaboration. Protocol for a prospective collaborative overview of all current and planned randomized trials of cholesterol treatment regimens. Am J Cardiol 1995;75:1130-4.

DOI: 10.1056/NEJMp1704482

Whose Data Are They Anyway? Can a Patient Perspective Advance the Data-Sharing Debate?

Charlotte J. Haug, M.D., Ph.D.

Most patients haven't thought much about data sharing, according to Sara Riggare, but those who have "find the current system unreasonable. Patients expect that health care professionals and researchers use patient data in the best possible way. That there is a fight over what the best way is is perplexing and disappointing."

Riggare is an engineer and doctoral student at the Health Informatics Center at Karolinska Institutet in Stockholm, where she researches models and methods for "digital self-care" in chronic disease — ways to use technology in monitoring and treating oneself. She is also a patient. Riggare had her first symptoms of Parkinson's disease in her early teens and calls herself a "digital patient." Actively engaged in her own care, she advocates both for patients' right to access their own medical data and for the health care system to more actively use patients' experiences as a resource. She shares her opinions on her blog, "Sara. Not patient but im-patient."[1]

Like Riggare, the patients who participated in the recent *Journal* summit on aligning incentives for data sharing want their data shared quickly, especially to ensure that other patients know about possible side effects. But they also want some control over how the data are shared. For example, they would be more hesitant to participate if commercial or other interests were involved — for instance, if health care systems wanted to use the data to decide whether to provide care to certain groups or if drug or insurance companies had a commercial interest in them.

SPECIAL ARTICLE

Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing

Michelle M. Mello, J.D., Ph.D., Van Lieou, B.S., and Steven N. Goodman, M.D., Ph.D.

ABSTRACT

BACKGROUND

Sharing of participant-level clinical trial data has potential benefits, but concerns about potential harms to research participants have led some pharmaceutical sponsors and investigators to urge caution. Little is known about clinical trial participants' perceptions of the risks of data sharing.

METHODS

We conducted a structured survey of 771 current and recent participants from a diverse sample of clinical trials at three academic medical centers in the United States. Surveys were distributed by mail (350 completed surveys) and in clinic waiting rooms (421 completed surveys) (overall response rate, 79%).

RESULTS

Less than 8% of respondents felt that the potential negative consequences of data sharing outweighed the benefits. A total of 93% were very or somewhat likely to allow their own data to be shared with university scientists, and 82% were very or somewhat likely to share with scientists in for-profit companies. Willingness to share data did not vary appreciably with the purpose for which the data would be used, with the exception that fewer participants were willing to share their data for use in litigation. The respondents' greatest concerns were that data sharing might make others less willing to enroll in clinical trials (37% very or somewhat concerned), that data would be used for marketing purposes (34%), or that data could be stolen (30%). Less concern was expressed about discrimination (22%) and exploitation of data for profit (20%).

CONCLUSIONS

In our study, few clinical trial participants had strong concerns about the risks of data sharing. Provided that adequate security safeguards were in place, most participants were willing to share their data for a wide range of uses. (Funded by the Greenwall Foundation.)

From the Department of Health Research and Policy, Stanford University School of Medicine (M.M.M., V.L., S.N.G.) and Stanford Law School (M.M.M.) — both in Stanford, CA. Address reprint requests to Dr. Mello at Stanford Law School, 559 Nathan Abbott Way, Stanford, CA 94305, or at mmello@law.stanford.edu.

N Engl J Med 2018;378:2202-11.
DOI: 10.1056/NEJMsa1713258

We are rapidly moving toward a world in which broad sharing of participant-level clinical trial data is the norm.[1-4] The European Medicines Agency has implemented a policy to expand public access to data concerning products it approves,[5,6] the Food and Drug Administration is considering how to expand access to data pooled within a product class,[7] major research sponsors[8-12] and journal editors[13] have begun promoting data sharing, and lawmakers' interest[14] has resulted in legislation authorizing the National Institutes of Health to require all of its grantees to share data.[15,16] Pharmaceutical industry associations have committed to making data more accessible,[17] and several data platforms are now available.[11,18-21]

Previous work has identified diverse potential benefits of expanding access to participant-level data.[1,4,22] These benefits include deterring inaccurate reporting of trial results,[4,23,24] accelerating scientific discovery,[25] and exploring questions that are not answerable within individual trials.[4,26] In addition, data sharing helps fulfill the ethical obligation to make the most of research participants' contributions to science.[13,27-30]

Yet some investigators and industry sponsors of clinical trials have expressed hesitancy about the swift move toward broad data sharing. These groups have shifted from opposing data sharing to supporting it[31,32]; however, several concerns have led them to urge caution, limit what they share, and resist some initiatives as going too far.[32,33] Chief among these are concerns about potential harm to research participants.[17,32,34,35] Sponsors and investigators express worries that participants' privacy cannot be adequately protected, particularly in light of the fact that experts have demonstrated that it is possible to reidentify participant-level data.[35-39] Some pharmaceutical company representatives warn that the threat to privacy posed by data sharing will chill willingness to participate in trials, thereby delaying the availability of new therapies.[36,38]

It is unclear to what extent participants in clinical trials share these concerns. There is a large body of empirical literature concerning people's preferences related to biobanking[40,41] but not about clinical trials. When patient advocacy groups have spoken about data sharing, they have sometimes been challenged as parroting the views of pharmaceutical companies that financially support them rather than conveying trial participants' views.[42] One commentator recently remarked that in debates about data sharing, "Both sides claim to have the patient's and the public's best interests at heart, but not many partisans of either camp have asked patients what those interests are."[43] To investigate this issue, we surveyed a large sample of participants in a diverse group of clinical trials.

Methods

Participants

Survey participants had been enrolled, or were the parent or guardian of someone who had been enrolled, in an interventional clinical trial within the previous 2 years. We obtained agreement from nine principal investigators (PIs) in clinical trials at three academic medical centers to facilitate access to their trial participants, including one PI who provided access to all trials in the university's Clinical and Translational Science Institute.

We aimed for a broadly representative sample of trials that would be sufficient to provide at least 1200 potential survey participants. We selected the PIs we approached on the basis of personal contacts and stressed our interest in ensuring representation of racial and ethnic minority groups and persons with major health problems.

The final sample included both community-based trials (e.g., involving smoking cessation or diabetes prevention) and hospital-based trials (e.g., involving cancer or kidney disease). Within these trials, all the participants were eligible for the survey unless the trial team judged them as having cognitive impairment or being unable to respond to questions in English. The study was approved by the institutional review boards at Stanford and at the medical centers that provided access to the trial participants.

Questionnaire Development

A 10-page structured survey questionnaire was used to elicit clinical trial participants' views on the sharing of data from clinical trials. Details of the survey development work, which included the use of focus groups, consultation with ex-

perts and community advisory boards, and pilot testing, are provided in Sections 2 and 5 in the Supplementary Appendix, available with the full text of this article at NEJM.org.

The questionnaire provided plain-English definitions of clinical trial, data sharing, and clinical trial data (Section 6 in the Supplementary Appendix). It included reminders that the survey was asking about sharing of individual-level information about trial participants, not research results, and that respondents should assume that the data were deidentified.

SURVEY ADMINISTRATION

Clinical trial PIs chose from among three methods of survey delivery: email, regular mail, or in-person distribution in study clinic waiting rooms. Four PIs chose regular mail, four chose the clinic, and one used both. All surveys were completed on paper, and the clinic staff's interaction with respondents was limited to a receptionist or research assistant handing out and collecting the questionnaires (Section 1 in the Supplementary Appendix).

The surveys were accompanied by informed consent information and a $40 gift card. The responses were identified by participant identification number only.

STATISTICAL ANALYSIS

Responses were manually entered into a database in the Stanford University REDCap Survey system[44] and analyzed with the use of Stata software, version 13 (StataCorp). In addition to univariate statistics and cross-tabulations, multivariable logistic-regression models were run to identify predictors of the expression of negative views of data sharing. The following outcomes were modeled: perceiving the potential negative consequences of data sharing to outweigh the benefits (either strongly, moderately, or a little); being somewhat or very unlikely to allow one's own trial data to be shared with scientists in not-for-profit settings; and being somewhat or very unlikely to allow data to be shared with scientists in drug companies. To account for missing data, multiple imputation was performed with the Stata "mi" platform. Details of the model construction and regression results are provided in Sections 3 and 4 in the Supplementary Appendix.

RESULTS

SAMPLE CHARACTERISTICS

Completed surveys were received from 771 of the 978 invited trial participants (79%) and included 350 mailed surveys and 421 surveys completed in the clinic (Section 1 in the Supplementary Appendix). Respondents were fairly evenly distributed across the three academic medical centers (33%, 27%, and 40%) and were drawn from 119 different trials. Percentages based on the 771 respondents have a 95% confidence interval no wider than ±3.6 percentage points.

Table 1, and Table S6 in the Supplementary Appendix, show the characteristics of the sample. Within the previous 2 years, 42% of the respondents had participated in a clinical trial as a person with the health condition being studied, 55% as a healthy volunteer or person at risk for the studied health condition, and 3% as both. The two most common topics studied in the trials were diabetes and issues related to nutrition, weight, and vitamin supplementation. A total of 90% of respondents were trial participants themselves, and 7% were parents of participants. More than 94% of the respondents reported having had positive experiences as clinical trial participants. Half were motivated to participate in the trial by the prospect of a health benefit, 33% by altruism, and 16% by other factors.

PERCEIVED RISKS OF DATA SHARING

For 9 of 11 potential consequences of data sharing, less than 10% of the respondents said they were "very concerned" and less than one third were "very" or "somewhat" concerned about the risk (Fig. 1). A total of 20% to 26% of the respondents were very or somewhat concerned about discrimination, reidentification, and exploitation of data for profit. Respondents were more concerned that data sharing could deter people from enrolling in clinical trials (37%), that companies might use the information for marketing purposes (34%), or that their data could be stolen (30%). Asked to select the most important potential risk, respondents expressed divergent views, with the most common choices being that the information might be stolen (15%) or used for marketing purposes (11%) and that others might be more reluctant to en-

roll in clinical trials if they knew their data would be shared (10%) (Table S7 in the Supplementary Appendix).

PERCEIVED BENEFITS OF DATA SHARING

Strong majorities of respondents (67% to 82%) believed that data sharing would yield "a great deal" or "a lot" of several benefits (Fig. 2). In contrast, 43% believed it would help lawyers prove their case in product liability lawsuits. When respondents were asked to choose the most important benefit of data sharing, the most popular choices were making sure people's participation in clinical trials leads to the most scientific benefit possible (18%) and helping to get answers to scientific questions faster (17%) (Table S8 in the Supplementary Appendix). More than 85% of respondents expected that scientists in universities and other not-for-profit settings would benefit "a great deal" or "a lot" from data sharing; 81% of respondents had this expectation for physicians taking care of patients, 79% for companies developing medical products, and 72% for patients (Table S9 in the Supplementary Appendix)

OVERALL SUPPORT FOR DATA SHARING

In response to a question at the end of the survey, 82% of respondents indicated that they perceived that the benefits of data sharing outweighed the negative aspects, 8% felt the negative aspects outweighed the benefits, and 10% considered them equal (Table S10 in the Supplementary Appendix).

A total of 93% of respondents said they were very (69%) or moderately (24%) likely to allow their clinical trial data to be shared with scientists in universities and other not-for-profit organizations (Table 2), and 4% were very or somewhat unlikely to share. Although respondents had less trust in drug companies (18% trusted them a great deal or a lot) and health insurance companies (15%) than in universities (63%), 82% reported that they would be very or somewhat willing to share data with for-profit companies, whereas 8% were very or somewhat unwilling to share (Table 2, and Table S11 in the Supplementary Appendix).

Willingness to share data varied little according to the purpose for which it would be used — with the exception of its use in lawsuits, al-

Table 1. Sample Characteristics as Reported in the Survey.*

Characteristic	No. of Participants/ Total No. (%) (N=771)
Female sex	380/762 (49.9)
Age	
<25 yr	63/762 (8.3)
25–44 yr	177/762 (23.2)
45–64 yr	286/762 (37.5)
≥65 yr	236/762 (31.0)
Hispanic ethnic group	101/759 (13.3)
Race	
White	518/768 (67.4)
Black or African American	113/768 (14.7)
American Indian or Alaskan Native	51/768 (6.6)
Asian	25/768 (3.3)
Other	61/768 (7.9)
Education	
Less than high school	40/752 (5.3)
High-school diploma	125/752 (16.6)
Some college	206/752 (27.4)
College degree	238/752 (31.6)
Graduate degree	143/752 (19.0)
Annual family income	
Less than $15,000 to $24,999	173/742 (23.3)
$25,000 to $54,999	206/742 (27.8)
$55,000 to $99,999	189/742 (25.5)
$100,000 or higher	174/742 (23.5)
Health status	
Excellent	168/757 (22.2)
Good	420/757 (55.5)
Fair	156/757 (20.6)
Poor	13/757 (1.7)
Trial topic	
Nutrition, weight, or vitamins	172/771 (22.3)
Diabetes	172/771 (22.3)
Cardiovascular disease	71/771 (9.2)
Aging, neurodegenerative disease, or memory	64/771 (8.3)
Tobacco use	52/771 (6.7)
Liver disease	49/771 (6.4)
Mental illness	41/771 (5.3)
Cancer	39/771 (5.1)
Kidney disease	26/771 (3.4)
Other	85/771 (11.0)
Overall experience as a trial participant	
Very positive	573/752 (76.2)
Somewhat positive	136/752 (18.1)
Neither positive nor negative	34/752 (4.5)
Somewhat negative	9/752 (1.2)
Very negative	0

* All characteristics with exception of trial topic were reported by the participant in the survey. Percentages may not total 100 because of rounding. Further details are provided in Section 6 in the Supplementary Appendix.

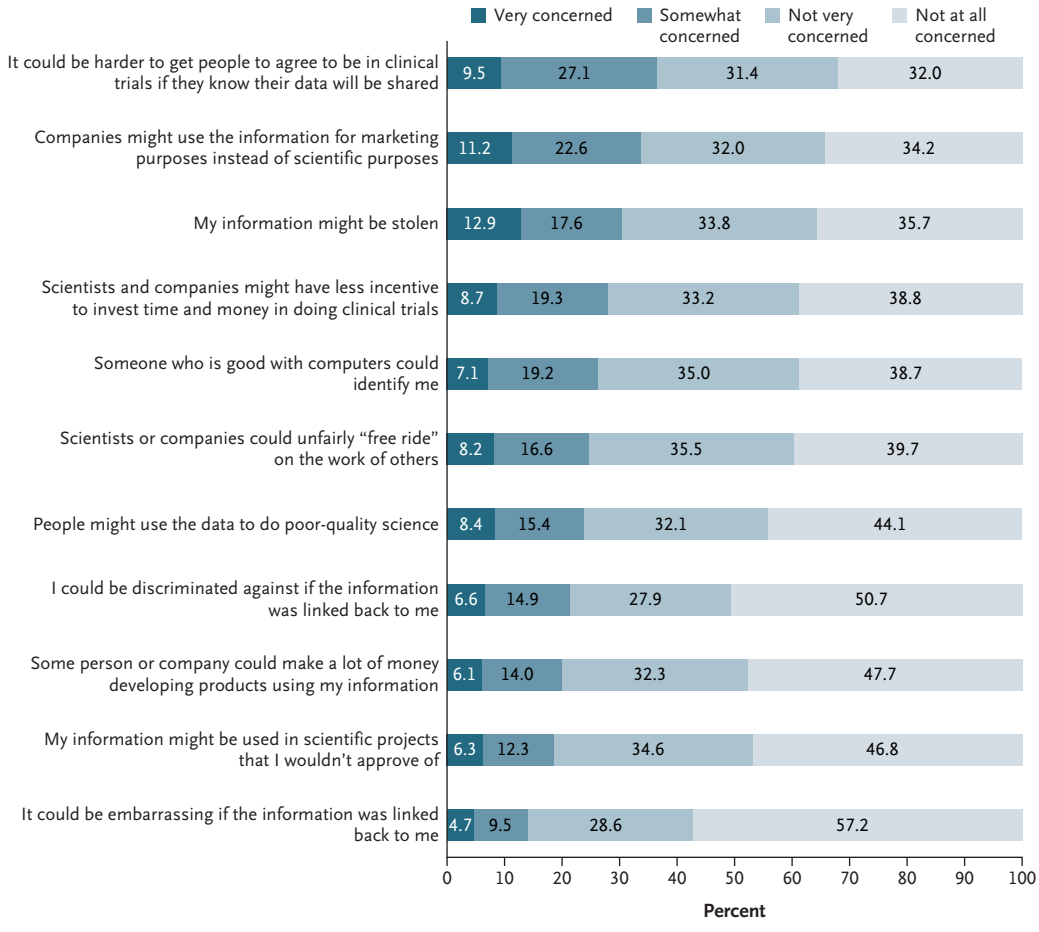


Figure 1. Level of Concern about Potential Consequences of Data Sharing.

Shown are the responses to an item worded as "How concerned are you about the following potential consequences of sharing *anonymous*, *individual* clinical trial data?" Numbers were rounded to the nearest tenth. The accuracy (95% confidence interval) of the percentages close to 50% is ±3.6 percentage points, diminishing to ±2.2 percentage points for percentages close to 10%.

though a majority of respondents were still willing to share even for that purpose (Table 2). No appreciable differences were found between uses that did and uses that did not benefit the participant directly or between uses for verifying previous research results and uses for making new discoveries.

Among the write-in comments, the most dominant theme was the need to help others as much as possible. Many commenters expressed confidence in the deidentification of data. Several urged greater cooperation and less competition among scientists.

PREDICTORS OF ATTITUDES

In multivariable modeling, the likelihood that a respondent would feel that the negative aspects of data sharing outweighed the benefits was significantly higher among those who felt that other people generally could not be trusted (odds ratio, 2.3; 95% confidence interval [CI], 1.2 to 4.6) and among those who were concerned about the risk of reidentification (odds ratio, 2.4; 95% CI, 1.2 to 4.5) or about information theft (odds ratio, 2.2; 95% CI, 1.2 to 4.1) (Section 4 in the Supplementary Appendix). The only other significant predictor was having a

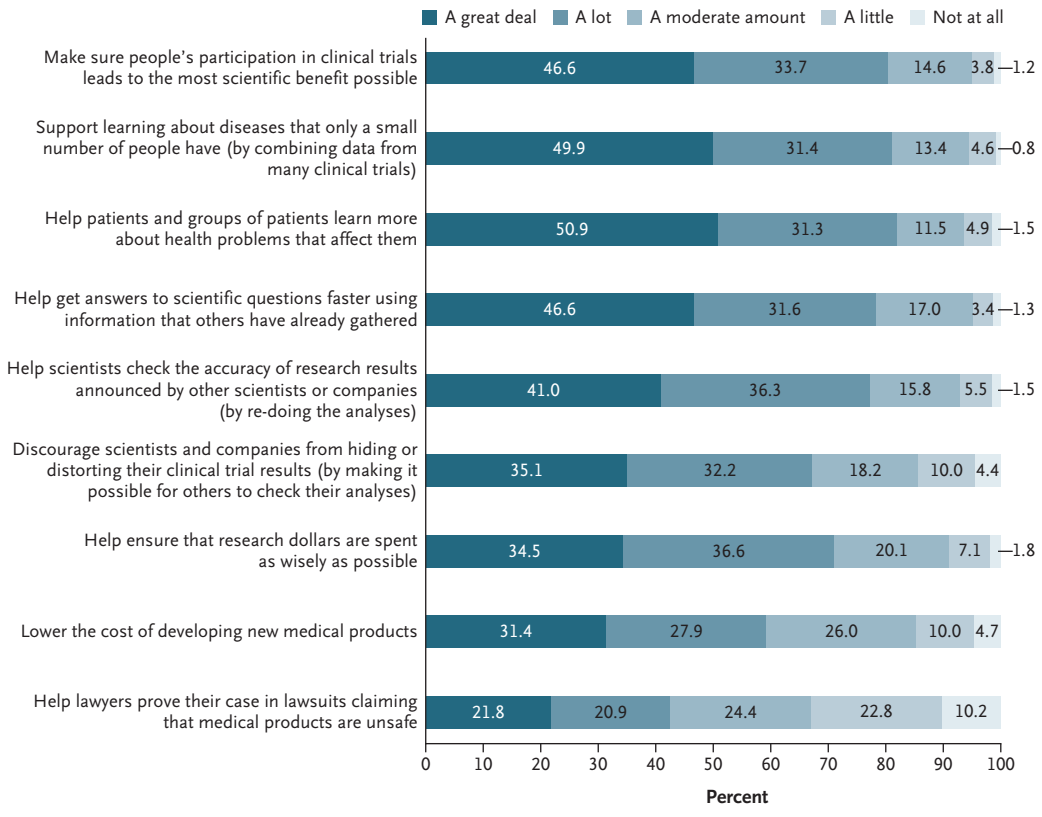


Figure 2. Perceived Benefits of Data Sharing.

Shown are the responses to an item worded as "How much do you think sharing *anonymous*, *individual* clinical trial data can . . . ." Numbers were rounded to the nearest tenth. The accuracy (95% confidence interval) of percentages close to 50% is ±3.6 percentage points, diminishing to ±2.2 percentage points for percentages close to 10%.

college degree, which was associated with a lower likelihood of feeling that the negative aspects of data sharing outweighed the benefits (odds ratio, 0.27; 95% CI, 0.2 to 0.5).

A low level of trust in people was also a significant predictor of being somewhat or very unlikely to share one's own data with scientists in not-for-profit contexts (odds ratio, 3.7; 95% CI, 1.6 to 8.3) or drug companies (odds ratio, 2.5; 95% CI, 1.3 to 4.8). Low trust in drug companies was a significant predictor of unwillingness to share data with drug-company scientists (odds ratio, 2.4; 95% CI, 1.4 to 4.2). Having a college degree was associated with a significantly lower likelihood of refusing to share data with not-for-profit scientists (odds ratio, 0.28; 95% CI, 0.10 to 0.78).

DISCUSSION

In this study assessing the views of clinical trial participants on the sharing of participant-level clinical trial data beyond genomic information, several key messages emerged. First, most of the clinical trial participants in our study believed that the benefits of data sharing outweighed the potential negative aspects and were willing to share their data. Their willingness to share was high regardless of the way in which the data would be used, with the exception of litigation, and it extended to uses that involved no prospect of direct benefit to themselves or their family members. Despite low levels of trust in pharmaceutical companies, most trial participants were willing to share their data with them.

Table 2. Willingness of Clinical Trial Participants to Share Their Data, According to Type of Use and Recipient.*

Type of Use or Recipient	Very Likely	Somewhat Likely	Neither Likely nor Unlikely	Somewhat Unlikely	Very Unlikely
			percent of respondents		
Type of use					
To help patients and groups of patients learn more about health problems that affect them	77.8	18.8	2.6	0.4	0.4
To do research on health problems that affect my family or me	78.3	17.1	3.2	1.1	0.5
To do research that will help others	79.9	17.1	2.0	0.4	0.7
To help get answers to scientific questions faster using information that others have already gathered	72.2	22.6	3.4	1.2	0.5
To help scientists check the accuracy of research results announced by other scientists or companies (by redoing the analyses)	70.9	22.6	3.8	1.5	1.2
To learn more about diseases that only a small number of people have (by combining data from many clinical trials)	69.1	22.1	5.9	1.7	1.2
To help lawyers prove their case in lawsuits claiming that medical products are unsafe	27.9	24.5	26.9	12.7	8.0
Recipient					
Scientists in universities and other not-for-profit organizations	69.2	24.0	3.3	1.7	1.8
Scientists in companies developing medical products, such as prescription drugs	53.4	28.5	10.6	5.4	2.1

* Shown are the responses to items worded as "How likely would you be to allow your *anonymous*, *individual* clinical trial data to be used in the following ways?" (for type of use) or "How likely would you be to allow your *anonymous*, *individual* clinical trial data to be shared with . . . ." (for recipient). Numbers were rounded to the nearest tenth. The accuracy (95% confidence interval) of percentages close to 50% is ±3.6 percentage points, diminishing to ±2.2 for percentages close to 10%.

The respondents' lack of differentiation among different data users and uses contrasts with previous study findings related to biobank participation. Those studies consistently showed substantially less willingness to share biospecimens with researchers in for-profit companies than with university researchers.[45-53] One study showed the same effect for sharing information from electronic health records (EHRs) for research purposes.[54]

The willingness of the respondents in our study to share clinical trial data was greater than that found in many previous studies that involved participants' attitudes toward research use of biospecimens or EHR data.[40,41,54-56] Expanding access to clinical trial data shares some ethical complexities with biobanking, such as how to obtain meaningful informed consent, but genetic information raises special concerns.[45,57] On the other hand, clinical trial data include information from medical records and questionnaires that reveals much more about participants than biospecimens. Some such information — for example, sexual orientation or substance use — may carry serious social risks.[38] A further consideration is that with rare exceptions,[58] biobanking studies presume that an institutional review board will approve future uses of the data — a safeguard that may not be present for sharing of clinical trial data. Finally, biobanking and EHR studies have generally presumed that the data would be used by qualified researchers, but some proposals for "open access" data sharing are not so limited.[1,4]

The values and concerns of clinical trial participants may differ from those of the general

public, patients in general, or other populations surveyed in biobanking and EHR studies. Clinical trial participants typically constitute a small proportion of the people who are eligible for participation and may represent those who are least bothered by data sharing and most enthusiastic about contributing to science. Their familiarity with physician-researchers may impart especially high trust in research and researchers.[59] Indeed, nearly all of our respondents reported very positive experiences as trial participants.

Our findings are broadly consistent with other literature on engagement in clinical trials in underscoring the idea that altruism as well as self-regarding motivations influence participation decisions.[60,61] In write-in comments, many respondents expressed the view that agreeing to broad use of their data was inherent in agreeing to participate in research.

A second finding of our study is that even when presented with a list of negative potential consequences, most trial participants do not express substantial concern about the risks of data sharing. On average, across the negative consequences they considered, approximately 8% of respondents were very concerned and 17% somewhat concerned. However, a substantial minority of respondents did express some concern, especially about discouraging others from volunteering for trials (37% somewhat or very concerned), having information used for marketing (34%), and having information stolen (31%). Many potential harms that trial sponsors and investigators worry a great deal about, such as reidentification and discrimination, were not of great concern to a sizable majority of participants, a finding that differs from surveys about biobanking that highlight these issues as leading concerns.[62]

Third, multivariable analysis revealed few differences in views across participant subgroups. Despite concern that distrust in research among African Americans may extend to data sharing,[1,46,58,63] we found no significant differences according to race. Because few of our respondents expressed negative views of data sharing, only large subgroup differences were detectable.

Our study had limitations. The respondents were relatively healthy: approximately a quarter characterized their health status as fair or poor. Although health status was not a significant predictor of attitudes in our models, a less healthy group of respondents might have reported different views. Our response rate was high, but we cannot exclude the possibility of nonresponse bias. Some people may decline to enroll in clinical trials out of concern that their data might be shared, and they are not represented in our sample. The survey concepts were complex, and although we conducted pilot work to clarify questions, some respondents may have had comprehension difficulties or lacked sufficient understanding of data sharing to meaningfully assess the potential consequences. Finally, respondents' actual willingness to share their data might be lower than their hypothetical willingness. Previous research on genomic data, however, has shown the reverse.[59,62]

Our findings suggest that concerns about trial participants' attitudes toward data sharing invoked by companies and investigators who caution against it may be exaggerated. Participants perceive data sharing to have many benefits, and most are willing to share their data. Finally, participants' concern about the use of their data for marketing is worth addressing. Data repositories could require data requesters to attest that no marketing use will occur, and consent documents could offer assurances about this requirement.

Reaching a world in which the sharing of clinical trial data is routine requires surmounting several challenges — financial, technical, and operational. But in this survey, participants' objections to data sharing did not appear to be a sizable barrier.

Supported by a grant from the Greenwall Foundation. Research information technology used in the study was funded by a grant from the National Center for Research Resources, National Institutes of Health (Clinical and Translational Sciences Award UL1 TR001085, to Stanford University).

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank the clinical trial principal investigators and teams for their generosity in facilitating access to the trial participants and administering the survey; Harry Selker, Edward Kuczynski, Anantha Shekhar, Laurie Trevino, and Brenda Hudson for connecting us with the trial teams; Sayeh Fattahi, Cynthia Rinaldo, and Quinn Walker for research assistance; and Jon Krosnick, Eric Campbell, Joe Ross, Jennifer Miller, Randy Stafford, and members of Dr. Stafford's American Indian Community Action Board for feedback on earlier drafts of the survey questionnaire.

REFERENCES

1. Institute of Medicine. Sharing clinical trial data: maximizing benefits, minimizing risk. Washington, DC: National Academies Press, 2015.
2. Zarin DA. Participant-level data and the new frontier in trial transparency. N Engl J Med 2013;369:468-9.
3. Loder E. Sharing data from clinical trials: where we are and what lies ahead. BMJ 2013;347:f4794.
4. Mello MM, Francer JK, Wilenzick M, Teden P, Bierer BE, Barnes M. Preparing for responsible sharing of clinical trial data. N Engl J Med 2013;369:1651-8.
5. European Medicines Agency. Clinical data publication (http://www.ema.europa.eu/ema/?curl=pages/special_topics/general/general_content_000555.jsp).
6. Davis AL, Miller JD. The European Medicines Agency and publication of clinical study reports: a challenge for the US FDA. JAMA 2017;317:905-6.
7. Food and Drug Administration. Availability of masked and de-identified non-summary safety and efficacy data: request for comments. Fed Regist 2013;78(107):33421-3 (https://www.federalregister.gov/articles/2013/06/04/2013-13083/availability-of-masked-and-de-identified-non-summary-safety-and-efficacy-data-request-for-comments).
8. Coady SA, Wagner E. Sharing individual level data from observational studies and clinical trials: a perspective from NHLBI. Trials 2013;14:201.
9. Walport M, Brest P. Sharing research data to improve public health. Lancet 2011;377:537-9.
10. Jumbe NL, Murray JC, Kern S. Data sharing and inductive learning — toward healthy birth, growth, and development. N Engl J Med 2016;374:2415-7.
11. Krumholz HM, Waldstreicher J. The Yale Open Data Access (YODA) Project — a mechanism for data sharing. N Engl J Med 2016;375:403-5.
12. Green AK, Reeder-Hayes KE, Corty RW, et al. The Project Data Sphere initiative: accelerating cancer research by sharing data. Oncologist 2015;20(5):464-e20.
13. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials — a requirement of the International Committee of Medical Journal Editors. N Engl J Med 2017;376:2277-9.
14. Warren E. Strengthening research through data sharing. N Engl J Med 2016;375:401-3.
15. 21st Century Cures Act, Pub. L. No. 114-255 § 2014 (2016).
16. Majumder MA, Guerrini CJ, Bollinger JM, Cook-Deegan R, McGuire AL. Sharing data under the 21st Century Cures Act. Genet Med 2017;19:1289-94.
17. PhRMA, EfPIA. Principles for responsible clinical trial data sharing. 2014 (http://www.phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf).
18. Rockhold F, Nisen P, Freeman A. Data sharing at a crossroads. N Engl J Med 2016;375:1115-7.
19. Krumholz HM, Gross CP, Blount KL, et al. Sea change in open science and data sharing: leadership by industry. Circ Cardiovasc Qual Outcomes 2014;7:499-504.
20. Bierer BE, Li R, Barnes M, Sim I. A global, neutral platform for sharing trial data. N Engl J Med 2016;374:2411-3.
21. Strom BL, Buyse ME, Hughes J, Knoppers BM. Data sharing — is the juice worth the squeeze? N Engl J Med 2016;375:1608-9.
22. Eichler HG, Pétavy F, Pignatti F, Rasi G. Access to patient-level trial data — a boon to drug developers. N Engl J Med 2013;369:1577-9.
23. Shining a light on trial data. Nat Biotechnol 2012;30:371.
24. Krumholz HM, Ross JS. A model for dissemination and independent analysis of industry data. JAMA 2011;306:1593-4.
25. Whitty CJ, Mundel T, Farrar J, Heymann DL, Davies SC, Walport MJ. Providing incentives to share data early in health emergencies: the role of journal editors. Lancet 2015;386:1797-8.
26. Eichler HG, Abadie E, Breckenridge A, Leufkens H, Rasi G. Open clinical trial data for all? A view from regulators. PLoS Med 2012;9(4):e1001202.
27. Gøtzsche PC. Why we need easy access to all data from all clinical trials and how to accomplish it. Trials 2011;12:249.
28. Lemmens T. Pharmaceutical knowledge governance: a human rights perspective. J Law Med Ethics 2013;41:163-84.
29. Bauchner H, Golub RM, Fontanarosa PB. Data sharing: an ethical and scientific imperative. JAMA 2016;315:1237-9.
30. Bertagnolli MM, Sartor O, Chabner BA, et al. Advantages of a truly open-access data-sharing model. N Engl J Med 2017;376:1178-81.
31. PhRMA statement on clinical trials and *Bad Pharma*. Press release of PhRMA, February 5, 2013 (http://www.phrma.org/press-release/phrma-statement-on-clinical-trials-and-bad-pharma).
32. Rathi V, Dzara K, Gross CP, et al. Sharing of clinical trial data among trialists: a cross sectional survey. BMJ 2012;345:e7570.
33. Statement on European General Court decision on EMA. Press release of PhRMA, April 30, 2013 (http://www.phrma.org/press-release/phrma-statement-on-decision-of-the-european-general-court-on-ema-disclosure-of-company-data).
34. Berlin JA, Morris S, Rockhold F, Askie L, Ghersi D, Waldstreicher J. Bumps and bridges on the road to responsible sharing of clinical trial data. Clin Trials 2014;11:7-12.
35. Kaye J. The tension between data sharing and the protection of privacy in genomics research. Annu Rev Genomics Hum Genet 2012;13:415-31.
36. Castellani J. Are clinical trial data shared sufficiently today? Yes. BMJ 2013;347:f1881.
37. Homer N, Szelinger S, Redman M, et al. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. PLoS Genet 2008;4(8):e1000167.
38. Rosenblatt M, Jain SH, Cahill M. Sharing of clinical trial data: benefits, risks, and uniform principles. Ann Intern Med 2015;162:306-7.
39. Alemayehu D, Anziano RJ, Levenstein M. Perspectives on clinical trial data transparency and disclosure. Contemp Clin Trials 2014;39:28-33.
40. Garrison NA, Sathe NA, Antommaria AHM, et al. A systematic literature review of individuals' perspectives on broad consent and data sharing in the United States. Genet Med 2016;18:663-71.
41. Shabani M, Bezuidenhout L, Borry P. Attitudes of research participants and the general public towards genomic data sharing: a systematic literature review. Expert Rev Mol Diagn 2014;14:1053-65.
42. Sample I. Big pharma mobilising patients in battle over drugs trials data. The Guardian. July 21, 2013 (https://www.theguardian.com/business/2013/jul/21/big-pharma-secret-drugs-trials).
43. Haug CJ. Whose data are they anyway? Can a patient perspective advance the data-sharing debate? N Engl J Med 2017;376:2203-5.
44. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) — a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377-81.
45. Pentz RD, Billot L, Wendler D. Research on stored biological samples: views of African American and White American cancer patients. Am J Med Genet A 2006;140:733-9.
46. Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. Public opinion about the importance of privacy in biobank research. Am J Hum Genet 2009;85:643-54.
47. Helft PR, Champion VL, Eckles R, Johnson CS, Meslin EM. Cancer patients' attitudes toward future research uses of stored human biological materials. J Empir Res Hum Res Ethics 2007;2:15-22.

48. Beskow LM, Dean E. Informed consent for biorepositories: assessing prospective participants' understanding and opinions. Cancer Epidemiol Biomarkers Prev 2008;17:1440-51.
49. Gornick MC, Ryan KA, Kim SY. Impact of non-welfare interests on willingness to donate to biobanks: an experimental survey. J Empir Res Hum Res Ethics 2014;9:22-33.
50. Trinidad SB, Fullerton SM, Bares JM, Jarvik GP, Larson EB, Burke W. Genomic research and wide data sharing: views of prospective participants. Genet Med 2010;12:486-95.
51. Rogith D, Yusuf RA, Hovick SR, et al. Attitudes regarding privacy of genomic information in personalized cancer therapy. J Am Med Inform Assoc 2014;21(e2):e320-e325.
52. Long MD, Cadigan RJ, Cook SF, et al. Perceptions of patients with inflammatory bowel diseases on biobanking. Inflamm Bowel Dis 2015;21:132-8.
53. Hapgood R, McCabe C, Shickle D. Public preferences for participation in a large DNA cohort study: a discrete choice experiment. HEDS discussion paper 04/05. 2004 (http://eprints.whiterose.ac.uk/10939/).
54. Grande D, Mitra N, Shah A, Wan F, Asch DA. Public preferences about secondary uses of electronic health information. JAMA Intern Med 2013;173:1798-806.
55. Truven Health Analytics, National Public Radio. Health poll: data privacy. November 2014 (https://truvenhealth.com/Portals/0/NPR-Truven-Health-Poll/NPRPulseDataPrivacy_Nov2014.pdf).
56. Luchenski SA, Reed JE, Marston C, Papoutsi C, Majeed A, Bell D. Patient and public views on electronic health records and their uses in the United kingdom: cross-sectional survey. J Med Internet Res 2013;15(8):e160.
57. Melas PA, Sjöholm LK, Forsner T, et al. Examining the public refusal to consent to DNA biobanking: empirical data from a Swedish population-based study. J Med Ethics 2010;36:93-8.
58. Storr CL, Or F, Eaton WW, Ialongo N. Genetic research participation in a young adult community sample. J Community Genet 2014;5:363-75.
59. Johnsson L, Helgesson G, Rafnar T, et al. Hypothetical and factual willingness to participate in biobank research. Eur J Hum Genet 2010;18:1261-4.
60. Stunkel L, Grady C. More than the money: a review of the literature examining healthy volunteer motivations. Contemp Clin Trials 2011;32:342-52.
61. Truong TH, Weeks JC, Cook EF, Joffe S. Altruism among participants in cancer clinical trials. Clin Trials 2011;8:616-23.
62. Oliver JM, Slashinski MJ, Wang T, Kelly PA, Hilsenbeck SG, McGuire AL. Balancing the risks and benefits of genomic data sharing: genome research participants' perspectives. Public Health Genomics 2012;15:106-14.
63. Sanderson SC, Brothers KB, Mercaldo ND, et al. Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US. Am J Hum Genet 2017;100:414-27.

SOUNDING BOARD

Data Sharing from Clinical Trials — A Research Funder's Perspective

Robert Kiley, Tony Peatfield, Jennifer Hansen, and Fiona Reddington

The Wellcome Trust, the Medical Research Council, Cancer Research UK, and the Bill and Melinda Gates Foundation share a common vision for maximizing the value of data that are generated through the trials we fund. We are committed to ensuring that the data from published clinical trials can be accessed by researchers so they can validate key findings, stimulate further inquiry, and ultimately deliver lifesaving results.

The sharing of data during the outbreak of Ebola virus disease in West Africa that began in 2014 helped researchers to trace the origins of the final few cases and bring the epidemic under control.[1] And the challenge organized by the *Journal* to encourage researchers to use data from the Systolic Blood Pressure Intervention Trial (SPRINT) demonstrated the vast potential for those data to be reused to develop new applications and uncover new knowledge.[2]

The recent announcement by the International Committee of Medical Journal Editors (ICMJE) on data-sharing statements for clinical trials[3] is a step in the right direction but falls short of realizing our vision. The ICMJE has not mandated data sharing as a requirement for publication, and we find the example statements it provides to be vague and open to interpretation. Crucially, the requirements do not recognize that some research funders already have mandates for data sharing.

POLICY

As funders of medical research, we recognize the importance of the appropriate sharing of clinical-trial data for reasons of transparency, good practice, and accelerated dissemination of results to the broader community. There is now a clear consensus that the results of all clinical trials must be reported in a timely manner, as set out in a joint statement by the World Health Organization regarding public disclosure of results from clinical trials.[4] In addition, all our organizations have implemented data-sharing policies requiring that the data from studies we have funded will be made available to other researchers at the time of publication. This requirement applies equally to clinical trials.

These policies, however, do not mean that such data have to be openly available for anyone to access on the Web. We fully recognize that some data — and especially clinical-trial data — may contain sensitive, personal information about research participants, and these data need to be shared in a manner that protects participants' privacy and confidentiality and respects the terms under which they consented to take part in the study. Such an approach might include the use of managed-access procedures, whereby requests to access data are reviewed by an independent committee, and of data-access agreements that place appropriate restrictions on how the data may be used.

As funders, we also recognize the many challenges to data sharing[5] — most notably, those related to resources, equity, and incentives.

RESOURCES

Sharing data is not a cost-free activity. Data need to be collected, preserved, curated, and stored in standardized formats in order to be useful to the scientific community. We need to support technical solutions that enable researchers to easily discover, access, and reuse the data in order to reap the benefits of accelerating discovery, enabling research reproducibility, and preventing redundancy. In addition, funding bodies are increasingly requiring that researchers develop data-management plans as part of research proposals, and we support the justified costs of delivering these plans as an integral part of funding the research. We anticipate that the data-sharing

statements required by the ICMJE can, in part, be derived from researchers' data-management-and-sharing plans.

Funders are actively working in partnership to support the development of community resources that facilitate access to clinical-trial data and reduce the burden on trialists. In particular, our organizations are planning to participate in the ClinicalStudyDataRequest.com platform,[6] which currently includes trial data from 13 pharmaceutical companies, as a mechanism for listing and providing managed access to data from clinical trials that we have funded.

EQUITY

Particular concerns have been raised over the effect of more stringent requirements for sharing data from clinical trials that are conducted in low-income and middle-income countries — specifically, that requiring researchers in such countries to share data with better-resourced groups elsewhere may put them at an unfair disadvantage and that benefits will not necessarily be shared with the communities that participated in the research.

Our organizations are strongly committed to establishing trusted and equitable systems for data-access governance in these settings, which may include terms that require users to contribute to training and capacity development or to share the resulting outcomes. However, the fundamental requirement to ensure that data are accessible at the time of publication still holds firm.

INCENTIVES

Arguably, the biggest challenge to data sharing is the sense that researchers are not given incentives to share data — and worse, many researchers believe they are disadvantaging themselves by doing so. A recent survey of Wellcome Trust–funded researchers showed that the potential loss of publication opportunities — along with the belief that publishing is the only currency for successful grant funding and academic advancement — was a key factor in the inhibition of data sharing.[7]

As funders, we need to tackle this issue head-on and demonstrate that we value the sharing of data — as well as other outputs, such as software and materials (e.g., antibodies, cell lines, and reagents) — and will take these outputs into account when reviewing grant and job applications. In parallel, we will make it clear that we focus on the scientific content of an article, rather than its publication metrics or the name of the journal in which it was published. We commit to clearly communicating these values to the members of our grant-reviewing panels.

But we need to do more. The Wellcome Trust is reexamining its grant-application process to see how it can shift the emphasis from publications to a wider set of outputs. The Wellcome Innovator Awards program invites applicants to describe their key achievements and the significance in their field. These statements can be supported with reference to peer-reviewed articles, but also with other research outputs, such as patents, data sets, software, and materials.[8] Such a model could be applied more broadly. Asking applicants to explain how they support the values of open research — transparency, reproducibility, and early access to results — is also worthy of consideration.[9]

More broadly, there is a need to support and foster community-wide efforts in this realm. Such efforts include accelerating the uptake of consistent approaches for data citation that allow the use of data to be acknowledged and tracked. The recently announced initiative exploring the value of awarding "data authorship" to researchers whose data are used or reused is also one we are following with interest.[10]

CONCLUSIONS

Medical research saves lives, and as the challenges in our world continue to outweigh the resources, collaboration and cooperation among members of the global research community will be essential in maximizing the effect of funded research. It is simply unacceptable that the data from published clinical trials are not made available to researchers and used to their fullest potential to improve health.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Wellcome Trust (R.K.), the Medical Research Council (T.P.), and Cancer Research UK (F.R.) — all in London; and the Bill and Melinda Gates Foundation, Seattle (J.H.).

1. Arias A, Watson SJ, Asogun D, et al. Rapid outbreak sequencing of Ebola virus in Sierra Leone identifies transmission chains linked to sporadic cases. Virus Evol 2016;2(1):vew016.
2. The SPRINT Challenge winning submissions (https://challenge.nejm.org/pages/winners).
3. Taichman DB, Sahni P, Pinborg A, et al. Data sharing state-

ments for clinical trials — a requirement of the International Committee of Medical Journal Editors. N Engl J Med 2017;376:2277-9.
4. Joint statement on public disclosure of results from clinical trials. Geneva: World Health Organization, May 18, 2017 (http://www.who.int/ictrp/results/jointstatement/en/).
5. van Panhuis WG, Paul P, Emerson C, et al. A systematic review of barriers to data sharing in public health. BMC Public Health 2014;14:1144.
6. ClinicalStudyDataRequest.com home page (https://clinicalstudydatarequest.com).
7. Van den Eynden V, Knight G, Vlad A, et al. Survey of Wellcome researchers and their attitudes to open research. figshare. October 31, 2016 (https://doi.org/10.6084/m9.figshare.4055448.v1).
8. Wellcome Trust. Innovator awards (https://wellcome.ac.uk/funding/innovator-awards).
9. Schönbrodt F. Changing hiring practices towards research transparency: the first open science statement in a professorship advertisement. January 6, 2016 (http://www.nicebread.de/open-science-hiring-practices/).
10. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. N Engl J Med 2017;376:1684-7.

DOI: 10.1056/NEJMsb1708278

EDITORIALS

Data Sharing Statements for Clinical Trials — A Requirement of the International Committee of Medical Journal Editors

The International Committee of Medical Journal Editors (ICMJE) believes there is an ethical obligation to responsibly share data generated by interventional clinical trials because trial participants have put themselves at risk. In January 2016 we published a proposal aimed at helping to create an environment in which the sharing of deidentified individual participant data becomes the norm. In response to our request for feedback we received many comments from individuals and groups.[1] Some applauded the proposals while others expressed disappointment they did not more quickly create a commitment to data sharing. Many raised valid concerns regarding the feasibility of the proposed requirements, the necessary resources, the real or perceived risks to trial participants, and the need to protect the interests of patients and researchers.

It is encouraging that data sharing is already occurring in some settings. Over the past year, however, we have learned that the challenges are substantial and the requisite mechanisms are not in place to mandate universal data sharing at this time. Although many issues must be addressed for data sharing to become the norm, we remain committed to this goal.

Therefore, ICMJE will require the following as conditions of consideration for publication of a clinical trial report in our member journals:

1. As of July 1, 2018, manuscripts submitted to ICMJE journals that report the results of clinical trials must contain a data sharing statement as described below.

2. Clinical trials that begin enrolling participants on or after January 1, 2019, must include a data sharing plan in the trial's registration. The ICMJE's policy regarding trial registration is explained at www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html. If the data sharing plan changes after registration this should be reflected in the statement submitted and published with the manuscript, and updated in the registry record.

Data sharing statements must indicate the following: whether individual deidentified participant data (including data dictionaries) will be shared; what data in particular will be shared; whether additional, related documents will be available (e.g., study protocol, statistical analysis plan, etc.); when the data will become available and for how long; by what access criteria data will be shared (including with whom, for what types of analyses, and by what mechanism). Illustrative examples of data sharing statements that would meet these requirements are in the Table.

These initial requirements do not yet mandate data sharing, but investigators should be aware that editors may take into consideration data sharing statements when making editorial decisions. These minimum requirements are intended to move the research enterprise closer to fulfilling our ethical obligation to participants. Some ICMJE member journals already maintain, or may choose to adopt, more stringent requirements for data sharing.

Sharing clinical trial data is one step in the process articulated by the World Health Organization (WHO) and other professional organizations as best practice for clinical trials: universal prospective registration; public disclosure of results from all clinical trials (including through journal publication); and data sharing. Although universal compliance with the requirement to

Table 1. Examples of Data Sharing Statements That Fulfill These ICMJE Requirements*

	Example 1	Example 2	Example 3	Example 4
Will individual participant data be available (including data dictionaries)?	Yes	Yes	Yes	No
What data in particular will be shared?	All of the individual participant data collected during the trial, after deidentification.	Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures, and appendices).	Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures and appendices).	Not available
What other documents will be available?	Study Protocol, Statistical Analysis Plan, Informed Consent Form, Clinical Study Report, Analytic Code	Study Protocol, Statistical Analysis Plan, Analytic Code	Study Protocol	Not available
When will data be available (start and end dates)?	Immediately following publication. No end date.	Beginning 3 months and ending 5 years following article publication.	Beginning 9 months and ending 36 months following article publication.	Not applicable
With whom?	Anyone who wishes to access the data.	Researchers who provide a methodologically sound proposal.	Investigators whose proposed use of the data has been approved by an independent review committee ("learned intermediary") identified for this purpose.	Not applicable
For what types of analyses?	Any purpose.	To achieve aims in the approved proposal.	For individual participant data meta-analysis.	Not applicable
By what mechanism will data be made available?	Data are available indefinitely at (*link to be included*).	Proposals should be directed to xxx@yyy. To gain access, data requestors will need to sign a data access agreement. Data are available for 5 years at a third party website (*link to be included*).	Proposals may be submitted up to 36 months following article publication. After 36 months the data will be available in our University's data warehouse but without investigator support other than deposited metadata. Information regarding submitting proposals and accessing data may be found at (*link to be included*).	Not applicable

* These examples are meant to illustrate a range of, but not all, data sharing options.

prospectively register clinical trials has not yet been achieved and requires continued emphasis, we must work toward fulfilling the other steps of best practice as well — including data sharing.

As we move forward into this new norm where data are shared, greater understanding and collaboration among funders, ethics committees, journals, trialists, data analysts, participants, and others will be required. We are currently working with members of the research community to facilitate practical solutions to enable data sharing. The United States Office for Human Research Protections has indicated that provided the appropriate conditions are met by those receiving them, the sharing of deidentified individual participant data from clinical trials does not require separate consent from trial participants.[2] Specific elements to enable data sharing statements that meet these requirements have been adopted at ClinicalTrials.gov (https://prsinfo.clinicaltrials.gov/definitions.html#shareData). The WHO also supports the addition of such elements at the primary registries of the International Clinical Trials Registry Platform. Unresolved issues remain, including appropriate scholarly credit to those who share data, and the resources needed for data access, the transparent processing of data requests, and data archiving. We welcome creative solutions to these problems at www.icmje.org.

We envision a global research community in which sharing deidentified data becomes the norm. Working toward this vision will help maximize the knowledge gained from the efforts and sacrifices of clinical trial participants.

Darren B. Taichman, M.D., Ph.D.
Secretary, ICMJE, Executive Deputy Editor, *Annals of Internal Medicine*

Peush Sahni, M.B., B.S., M.S., Ph.D.
Representative and Past President, World Association of Medical Editors

Anja Pinborg, M.D.
Scientific Editor-in-Chief, *Ugeskrift for Laeger (Danish Medical Journal)*

Larry Peiperl, M.D.
Chief Editor, *PLOS Medicine*

Christine Laine, M.D., M.P.H.
Editor-in-Chief, *Annals of Internal Medicine*

Astrid James, M.B., B.S.
Deputy Editor, *The Lancet*

Sung-Tae Hong, M.D., Ph.D.
Editor-in-Chief, *Journal of Korean Medical Science*

Abraham Haileamlak, M.D.
Editor-in-Chief, *Ethiopian Journal of Health Sciences*

Laragh Gollogly, M.D., M.P.H.
Editor, *Bulletin of the World Health Organization*, Coordinator, WHO Press

Fiona Godlee, F.R.C.P.
Editor-in-Chief, *BMJ (British Medical Journal)*

Frank A. Frizelle, M.B., Ch.B., F.R.A.C.S.
Editor-in-Chief, *New Zealand Medical Journal*

Fernando Florenzano, M.D.
Editor, *Revista Médica de Chile (Medical Journal of Chile)*

Jeffrey M. Drazen, M.D.
Editor-in-Chief, *New England Journal of Medicine*

Howard Bauchner, M.D.
Editor-in-Chief, *JAMA (Journal of the American Medical Association)* and the JAMA Network

Christopher Baethge, M.D.
Chief Scientific Editor, *Deutsches Ärzteblatt (German Medical Journal) & Deutsches Ärzteblatt International*

Joyce Backus, M.S.L.S.
Representative and Associate Director for Library Operations, National Library of Medicine

Editor's note: This editorial is being published simultaneously in *Annals of Internal Medicine*, *BMJ (British Medical Journal)*, *Bulletin of the World Health Organization*, *Deutsches Ärzteblatt (German Medical Journal)*, *Ethiopian Journal of Health Sciences*, *JAMA (Journal of the American Medical Association)*, *Journal of Korean Medical Science*, *New England Journal of Medicine*, *New Zealand Medical Journal*, *PLOS Medicine*, *The Lancet*, *Revista Médica de Chile (Medical Journal of Chile)*, and *Ugeskrift for Laeger (Danish Medical Journal)*.

Disclaimer: Dr. Sahni's affiliation as representative and past president of the World Association of Medical Editors (WAME) does not imply endorsement by WAME member journals that are not part of the ICMJE.

This editorial was published on June 5, 2017, at NEJM.org.

1. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors. Ann Intern Med 2016;164:505-6.
2. Menikoff J. Letter from Jerry Menikoff, M.D., J.D., Director, Office for Human Research Protections, to ICMJE Secretariat. March 7, 2017 (http://www.icmje.org/).

DOI: 10.1056/NEJMe1705439

DISCUSSION PAPER

Individual Patient-Level Data Sharing for Continuous Learning: A Strategy for Trial Data Sharing

Richard E. Kuntz, MD, Medtronic, Inc.; **Elliott M. Antman, MD, FAHA, MACC,** Harvard Medical School; **Robert M. Califf, MD, MACC,** Duke University; **Julie R. Ingelfinger, MD,** Harvard Medical School; **Harlan M. Krumholz, MD, SM,** Yale School of Medicine; **Alexander Ommaya, DSc,** Association of American Medical Colleges; **Eric D. Peterson, MD, MPH, FAHA, FACC,** Duke Clinical Research Institute; **Joseph S. Ross, MD, MHS,** Yale University School of Medicine; **Joanne Waldstreicher, MD,** Johnson & Johnson; **Shirley V. Wang, PhD, MSc,** Harvard Medical School; **Deborah A. Zarin, MD,** Harvard Medical School; **Danielle M. Whicher, PhD, MHS,** National Academy of Medicine; **Sameer M. Siddiqi, PhD,** National Academy of Medicine; and **Marianne Hamilton Lopez, PhD, MPA,** Duke-Margolis Center for Health Policy

July 1, 2019

Introduction

The National Academy of Medicine (NAM) has prioritized the use of clinical and administrative health care data as a core utility for a continuously learning health system[1] and for advancing the health and health care of Americans. There is increasing acceptance that sharing data constitutes a key strategy for continuous and real-time improvement in the effectiveness and efficiency of patient care and for the enhancement of research transparency and reproducibility [1]. "Individual patient-level data (IPD)[2] sharing" refers to "widespread, third-party access to the IPD and associated documentation from clinical trials" to achieve broad societal and scientific benefits [2].

1 The "Learning Health System" is a system in which science, informatics, incentives, and culture are aligned for continuous improvement and innovation, with best practices seamlessly embedded in the care process, patients and families as active participants in all elements, and new knowledge captured as an integral by-product of the care experience. SOURCE: Institute of Medicine. 2013. Best care at lower cost: The path to continuously learning health care in America. Washington, DC: The National Academies Press.

2 This includes individual patient-level data (e.g., raw data or an analyzable dataset); metadata, or "data about the data" (e.g., protocol, statistical analysis plan, and analytic code); and summary-level data (e.g., summary-level results posted on registries, lay summaries, publications, and clinical study reports).

Analyses of existing IPD may lead to a better understanding of current evidence, the generation of new information to support informed health care decision making, and improved transparency of original research findings, which, in turn, may enhance data integrity and public confidence in the overall clinical trial enterprise [3,4,5]. Public registration of key study details at study inception and the reporting of summary results through platforms such as ClinicalTrials.gov and other registries in the World Health Organization International Clinical Trials Registry Platform have already improved clinical research transparency. IPD sharing represents the next step in facilitating the transformation of raw study data to the aggregated data that form the basis of statistical analyses and reported results [2]. Making IPD and associated metadata available after study completion for clinical trials and observational studies can benefit the research community by enhancing transparency and enabling careful examination of the data and methods used by the primary research team (e.g., as demonstrated in *Box 3*), which is important, given ongoing concerns about the reproducibility of research studies [6]. Although industry has different incentives and concerns from academia, for academic investigators, the benefits of data shar-

ing include preservation and accessibility of their data, increased citation of the work, and increased visibility and opportunity for new collaborations [7].

Although there are potential benefits to IPD sharing, there are also many barriers that have yet to be addressed [6,8,9,10,11]. Contentious issues include consent for data sharing and the sharing of anonymized data, sustainable infrastructure and resources to support the preparation of IPD and metadata, and the heterogeneity of data repositories and related tools [9,12,13]. Also, limited guidance exists on the role of the primary research team in preparing IPD, the responsibilities of secondary research teams to ensure valid analyses, and the process by which conflicting findings should be reconciled [14,15]. For example, Natale, Stagg, and Zhang, and Gay, Baldridge, and Huffman demonstrate the potential difficulty of IPD reanalysis, given differences in population and endpoint definitions, and reliance on primary investigators to explain datasets and facilitate data use [16,17].

While not focused exclusively on sharing IPD, the Future of Research Communications and e-Scholarship group (FORCE11), as a step to address some of these barriers, published the first iteration of the FAIR (findable, accessible, interoperable, and reusable) principles in 2016, which aim to improve management and stewardship [18]. More recently, the International Committee of Medical Journal Editors (ICMJE) issued a statement on data sharing for clinical trials. As of July 2018, all ICMJE member journals began requiring that articles reporting results from clinical trials include a data-sharing statement.[3] Furthermore, for any clinical trials that began enrolling participants after January 1, 2019, the data-sharing plan needs to be included as part of the trial's registration [19].

Additionally, articles by Ohmann and colleagues offer consensus-based principles and recommendations for addressing common barriers, such as incentivizing, resourcing, and planning for IPD sharing during the design of an original study; structuring data and metadata using widely recognized standards; managing repository data and access; and monitoring data sharing [9,12]. Finally, the responsible sharing of clinical trial data was also the focus of a 2015 Institute of Medicine (now NAM) report, *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk,* that offers guiding principles for responsible data sharing and describes the benefits, risks, and challenges for a variety of stakeholders, including participants, sponsors, regulators, investigators, research institutions, journals, and professional societies [8].

The present paper aims to describe strategies for addressing outstanding challenges to IPD sharing that were identified through a collaborative effort facilitated by the NAM and through a review of relevant literature and selected IPD repositories. It builds on previous efforts by providing specific case examples of IPD sharing efforts (several of which are being led by members of the author group), focusing specifically on issues of greatest relevance to the US context and considering data from both commercial and noncommercial sources. While the authors present multiple viewpoints on how best to share IPD, all believe that important scientific contributions may be derived from leveraging previously acquired data for additional research and analysis.

The NAM Collaboration

To discuss the outstanding questions related to IPD sharing, the NAM hosted a meeting of the Clinical Effectiveness Research Innovation Collaborative in November 2016. During the discussion, meeting participants called for a more substantial and strategic focus on how to facilitate IPD sharing effectively, efficiently, and ethically. To address this charge, the authors of this paper have collected examples and drawn from their personal experiences to develop a set of actionable steps that may help promote responsible and widespread sharing of IPD from clinical trials that involve participants from the United States. We also identify the stakeholders responsible for each of the identified action steps.

Our paper does not describe all possible benefits and harms that may be associated with IPD sharing initiatives, nor does it include all possible financial considerations. Instead, the purpose is to create a policy and practice agenda that could lead to more robust and evidence-based IPD sharing efforts within the United States. Enhancing continuous learning from

3 According to the ICMJE website, data-sharing statements must include the following information: "whether individual deidentified participant data (including data dictionaries) will be shared ('undecided' is not an acceptable answer); what data in particular will be shared; whether additional, related documents will be available (e.g., study protocol, statistical analysis plan, etc.); when the data will become available and for how long; by what access criteria data will be shared (including with whom, for what types of analyses, and by what mechanism)." SOURCE: International Committee of Medical Journal Editors. 2019. Recommendations: Clinical trials. http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html#two (accessed May 15, 2019).

further analyses and the study of original data, without additional risk to patients and with maximum benefit for society, requires work, resources, culture change, and collaboration. To promote the necessary changes, we focus on operationalizing Recommendations 1, 3, and 4 of the NAM Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk publication: developing a data-sharing culture; implementing operational strategies to maximize benefits and minimize risks; and addressing infrastructure, technological, sustainability, and workforce challenges associated with IPD sharing [8].

Examples of IPD Sharing Initiatives

Over the past decade, there has been meaningful progress in activities, regulation, and practices associated with IPD sharing. Several governmental and nongovernmental agencies have either developed or are in the process of developing guidance related to IPD sharing, clinical study reports, summary results, and trial registration (e.g., the European Medicines Agency [EMA], ICMJE, the National Institutes of Health [NIH], the US Department of Veterans Affairs, the US Food and Drug Administration [FDA], and the World Health Organization) [19,20,21,22,23,24,25,26,27,28]. Broader policy changes have also emerged. For example, the European Union's General Data Protection Regulation regarding individual data privacy and accountability may have consequences for clinical research and patient care [29]. Many pharmaceutical companies have also established their own policies for IPD sharing [30], and data sharing is encouraged and incentivized through

Box 1 | Case Example 1: The Yale University Open Data Access (YODA) Project

Description: Initiated in 2013, the YODA Project is a voluntary, industry-supported effort to promote open science and data sharing. This initiative has made individual patient-level data and reports of clinical research available from three industry sponsors: Johnson & Johnson, Medtronic Inc., and SI-BONE Inc.

Governance structure: This initiative is overseen by an independent steering committee, which includes researchers, editors, ethicists, and members of the public. The names of the steering committee members, all decisions made, and all submitted research proposals which pre-specify the project plan are publicly posted. All requests are reviewed by the YODA Project for completeness and scientific merit; external review is used to assess scientific merit on a case-by-case basis. If approved, all data users must sign an institutional data use agreement (DUA) that explicitly precludes re-identification and data distribution.

Factors facilitating sharing: Factors include: (1) transparency and accessibility, including metadata and documentation, such as trial enrollment and study demographics for subgroup analysis; (2) YODA Project independence, including maintenance of full authority over data requests; (3) sponsor entitlement to exclusive data use after trial completion for up to 18 months; (4) the absence of data access fees; and (5) employment of DUA and data security measures to protect patient privacy.

Factors affecting costs and timeliness: Resource-intensive aspects of this initiative include: (1) establishing a transparent platform to support data sharing, including a trial request system and associated metadata; (2) responding to data queries; (3) sponsoring data de-identification and metadata preparation for external sharing; and (4) having the external user time, resources, and expertise needed to perform data analysis and prepare the findings for publication. Other factors affecting the timeliness of responses to data queries include institutional review and negotiation of DUAs.

Impact: Over 115 data requests have been received across all sponsors, all of which have been approved (provided the requested data were available) as of May 2019 and a majority of which were for multiple trials. Eighteen publications and 25 conference presentations have resulted from shared data thus far.

SOURCE: Developed by authors. Description of case example sourced from author experience and from: Krumholz, H. M., and J. Waldstreicher. 2016. The Yale Open Data Access (YODA) Project—a mechanism for data sharing. *The New England Journal of Medicine* 375(5):403-405. doi:10.1056/NEJMp1607342.
Ross, J. S., J. Waldstreicher, S. Bamford, J. A. Berlin, K. Childers, N. R. Desai, G. Gamble, C. P. Gross, R. Kuntz, R. Lehman, P. Lins, S. A. Morris, J. D. Ritchie, and H. M. Krumholz. 2018. Overview and experience of the YODA project with clinical trial data sharing after 5 years. *Scientific Data* 5:180268. doi:10.1038/sdata.2018.268.

Box 2 | Case Example 2: Supporting Open Access for Researchers (SOAR) Program

Description: The Duke Clinical Research Institute's SOAR initiative was created in 2013 to promote the sharing of de-identified individual patient-level data from the Duke Cardiac Catheterization Research Dataset (DukeCath), which includes information on adult patients undergoing cardiac catheterization procedures at Duke between 1985 and 2013, and clinical trials sponsored by Bristol-Myers Squibb.

Governance structure: Submitted data requests must be approved by Duke and reviewed by the institutional review board (IRB) established and compensated through a collaboration with Bristol-Myers Squibb. Proposals are evaluated on their scientific rationale and analysis plans. If approved, all users must sign an institutional data use agreement.

Factors facilitating sharing: Factors include (1) strong data security protections, including de-identification; (2) IRB oversight; and (3) contracting procedures.

Factors affecting costs and timeliness: Resource-intensive aspects of this initiative include (1) preparing and documenting DukeCath data extraction from the Duke Databank for Cardiovascular Disease, (2) establishing a data enclave, and (3) creating a "clean" and de-identified copy of the datasets.

Impact: This initiative has resulted in expanded investigator networks and collaboration, as well as enhanced awareness. To date, 57 data requests have been received, of which 22 have been approved and one has resulted in a publication.

SOURCE: Developed by authors. Description of case example sourced from author experience and from: Duke Clinical Research Institute. 2018. *SOAR data: Available datasets: Duke cardiac catheterization datasets.* https://dcri.org/our-approach/data-sharing/soar-data (accessed June 13, 2019).

various federal regulatory and funding agencies and publication bodies. While the requirements of these policies do not always directly align, they are important steps in IPD data sharing.

Additionally, several public funders and private companies—particularly some pharmaceutical and medical device companies that sponsor and conduct clinical trials—have established data repositories for secondary use and analysis [31,32,33,34,35,36,37]. Since effective clinical trial IPD sharing requires maintenance of and adequate support for data repositories, we examined six case studies of established repositories that have navigated obstacles to creating an infrastructure, developed operational strategies to maximize benefit and minimize risk, and contributed to growing a data-sharing culture.[4] We described each system's governance structure, factors facilitating sharing, factors affecting cost and timeliness, and impact to date (see *Boxes 1-6* and *Table 1*). The specific cases included are the Yale University Open Data Access (YODA) Project (Case 1); Duke Clinical Research Institute's Supporting Open Access for Researchers (SOAR) (Case 2); the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) of the National Heart, Lung, and Blood Institute (NHLBI) (Case 3); ClinicalStudyDataRequest.com (CSDR) (Case 4); Project Data Sphere (PDS) (Case 5); and the Project Genomics Evidence Neoplasia Information Exchange (GENIE) of the American Association for Cancer Research (AACR) (Case 6).

Of these six cases, a few were included because of the authors' detailed knowledge regarding these efforts, and others were added because of the availability of information on their development and procedures. The six cases were also selected because they differ in several important ways related to governance, data access models, data availability, and data type. For example, some are administered by public sector organizations and include data from publicly funded studies, whereas others are administered by groups of private sector organizations or public-private partnerships and include data from industry studies. Some use open-access data-sharing models, in which there is minimal review of data requests, while others rely on controlled access models, which use in-house or third-party expert review of data requests to provide greater protection for patients and data sponsors

4 These examples are not meant to represent an exhaustive list of available data repositories. Other repositories not discussed, but that may be of interest, include Vivli (https://vivli.org/), OpenTrials (https://opentrials.net/), and Dryad (https://datadryad.org).

Box 3 | Case Example 3: The National Heart, Lung, and Blood Institute (NHLBI) Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC)

Description: The NHLBI Data Repository—which, together with the NHLBI Biological Specimen Repository, is overseen by BioLINCC—aims to facilitate access to, and maximize the scientific value of, individual patient-level data from grants of high programmatic interest, or those with 500 or more participants and direct costs exceeding $500,000. Funding for BioLINCC is $1 million per year.

Governance structure: Submitted data requests are reviewed by the NHLBI institutional review board. NHLBI program officers oversee contractor activities, approve data requests, facilitate studies and contractor interactions, review and approve new study collections, and provide support for the resolution of issues and future directions.

Factors facilitating sharing: Factors include (1) the NHLBI's acceptance of data in any format in which it was collected, (2) study investigators' entitlement to 24 months of exclusive data use after trial completion, and (3) minimal burdens on investigators depositing data.

Factors affecting costs and timeliness: Resource-intensive aspects of this initiative include (1) maintaining the website and enhancing the portal, (2) reviewing and cleaning submitted data, (3) de-identifying data and preparing documents, (4) managing the process of requesting data, and (5) assisting investigators with data questions and responding to other queries. Other factors include contacting researchers for appropriate biospecimens to link them with associated clinical data.

Impact: About 200 data requests were processed in 2016, and about 1,000 investigators requested data from the repository. Over 800 publications have resulted from the repository data, and data requests have doubled every five years since its initiation. The data repository has resulted in new scientists being trained, expanded investigator collaborations, and enhanced transparency.

SOURCE: Developed by authors. Description of case example sourced from author experience and from: Ross, J. S., J. D. Ritchie, E. Finn, N. R. Desai, R. L. Lehman, H. M. Krumholz, and C. P. Gross. 2016. Data sharing through an NIH central database repository: A cross-sectional survey of BioLINCC users. *BMJ Open* 6(9):e012769. doi:10.1136/bmjopen-2016-012769.

[38,39]. Some offer data contributors the opportunity to review data requests for potential conflicts in terms of their publication plans, whereas others offer data users broad access to all available data.

The approaches these repositories take to collecting and accessing IPD vary. AACR's GENIE consists of voluntarily contributed data that are required to meet criteria related to quality and comprehensiveness, and must include at least 500 genomic records. The YODA Project and SOAR largely rely on robust partnerships among a small group of academic and industry data holders. Across repositories, IPD are required to be de-identified in a manner that is consistent with the Health Insurance Portability and Accountability Act and shared in a manner that is consistent with participants' informed consent in cases in which the data were not de-identified.

For example, NHLBI's BioLINCC repository, which includes data from NHLBI-funded studies, requires those contributing IPD to specify whether their data were collected with broad, unrestricted consent or tiered consent, and limits secondary uses in a manner that is consistent with consent restrictions (e.g., data can only be used for research on certain topics). BioLINCC also requires that data from funded contracts and large grants be made available within two years after the publication of primary outcome data, with additional rules for observational studies [40]. BioLINCC's historical development, described by Giffen et al. and Coady and Wagner [40,41], demonstrates the complex decisions underlying IPD repositories. BioLINCC's development entailed organizing existing data; assessing its quality; developing documentation for preparing, submitting, and requesting datasets; and developing workflows for data requests and review processes [40,41].

Coady et al. describe the use and publication record of BioLINCC. From January 2000 to May 2016, the repository received 1,116 data requests for 100 clinical studies [32]. Five years after the data request, 35

percent of studies that reused clinical trial data and 48 percent of studies that reused observational data were published [32]. A survey of investigators who had received data from BioLINCC indicated that due to time or financial resources, it was not practical to collect data of similar size and scope as those originally collected via NIH-supported work and made available through BioLINCC [42].

An analysis of data reuse requests to the YODA Project, SOAR, and CSDR indicated that between 2013 and 2015, 234 proposals were submitted [43]. For the YODA project, the vast majority of investigators requesting data (91.5 percent) are from academic institutions [44]. Although data request and data use statistics are useful indicators of IPD sharing, additional data are needed to better understand how instances of IPD sharing directly enhance patient care and quality improvement. Studies of data requests from BioLINCC suggest that investigators use data for novel research questions (72 percent), meta-analyses (7 percent), or pilot studies (9 percent); relatively few requested data for reanalysis [30,40]. These studies have also demonstrated that BioLINCC is most used by early stage investigators or trainees, which suggests "a potential role for repositories in the development of new trialists and epidemiologists" [32].

While growth in the use of clinical data repositories is promising, an important challenge for these repositories is the curation of the data, including data provenance, data formatting, and metadata quality [45]. The FDA has issued several guidance documents on data formats, metadata requirements, and related in-

Box 4 | Case Example 4: ClinicalStudyDataRequest.com (CSDR)

Description: CSDR, launched in 2013, aims to provide access to anonymized patient-level data from clinical studies sponsored or funded by a consortium of 17 research funders and industry organizations. The database of studies is publicly viewable and lists 3,623 studies spanning multiple phases and medical conditions.

Governance structure: CSDR is operated by IdeaPoint Inc. Current sponsors and funders include Astellas Pharma, Bayer, the Bill and Melinda Gates Foundation, Boehringer Ingelheim, Cancer Research UK, Daiichi Sankyo, Eisai, Eli Lilly, GlaxoSmithKline, Medical Research Council, Novartis, Roche, Sanofi, Takeda, UCB, ViiV Healthcare, and the Wellcome Trust. Proposal review is overseen by the Wellcome Trust, which serves as the secretariat of the independent review panel. Proposals are initially reviewed for completeness by the Wellcome Trust and referred to the corresponding study sponsor and/or funder for additional review. Sponsors check for feasibility and potential conflicts with the sponsor and/or funder's publication plan. After these preliminary reviews, the proposal is sent to the independent review panel for a full review. The panel assesses each proposal's scientific rationale, research plan, qualifications, potential conflicts, and publication plan. Upon approval, investigators must agree to a data-sharing agreement.

Factors facilitating sharing: Factors include (1) researchers' ability to select studies from multiple sponsors and across diseases in a single proposal, (2) researchers' ability to access data via an online SAS analytics portal, (3) data sponsors' and/or funders' ability to rapidly review proposals to identify potential conflicts or concerns, and (4) review of proposals by an independent panel administered by the Wellcome Trust.

Factors affecting costs and timeliness: Resource-intensive aspects of this initiative include (1) platform development, web maintenance, and portal enhancements; (2) review of research proposals by three parties, which can take up to 90 days; (3) and payments to independent review panel members and other experts on a per review basis.

Impact: Approximately 375 research proposals were submitted between May 2013 and January 2018. Of these, 177 have been provided access to requested data. Approximately 20 research proposals have led to publications in peer-reviewed journals. Most proposals focused on new analyses, not reanalysis of original results.

SOURCE: Developed by authors. Description of case example sourced from: ClinicalStudyDataRequest. 2018. *ClinicalStudyDataRequest.com.* https://clinicalstudydatarequest.com/Default.aspx (accessed June 13, 2019).

Box 5 | Case Example 5: Project Data Sphere (PDS)

Description: An online data platform launched in 2014, PDS uses an open access system to help researchers share, integrate, and analyze de-identified patient-level comparator arm data from 148 industry and academic Phase 3 cancer clinical trials.

Governance structure: PDS is operated and funded by the CEO Roundtable on Cancer's Life Sciences Consortium. The project is administered by a group of five officers, with additional ethical and scientific input from the executive committee, which includes nonprofit and industry members. Applicants must complete a user application form, agree to terms, and submit a brief summary of their background and initial research goals. All data in the platform are made available on an individual basis upon acceptance. All registered users must enter into an online services user agreement with PDS. Each data provider must also enter into a data-sharing agreement with PDS for each dataset provided.

Factors facilitating sharing: Factors include (1) use of a single application that provides access to all datasets; (2) use of an open access model with no applicant review panel; (3) data access typically within seven days of registration; and (4) researchers' ability to access and use data via a user-friendly SAS portal and, in some cases, download data onto their machines.

Factors affecting costs and timeliness: Resource-intensive aspects of this initiative include (1) platform development, web maintenance, and portal enhancements; (2) potential intellectual property and competitive risks for data providers as a result of the open access model; and (3) the requirement that data providers de-identify and upload data.

Impact: The platform has over 1,700 users and has facilitated more than 9,200 data downloads. It includes 148 research studies, representing over 100,000 patients. Since May 2015, there have been 11 peer-reviewed publications based on PDS data.

SOURCE: Developed by authors. Description of case example sourced from: Green, A. K., K. E. Reeder-Hayes, R. W. Corty, E. Basch, M. I. Milowsky, S. B. Dusetzina, A. V. Bennett, and W. A. Wood. 2015. The project data sphere initiative: Accelerating cancer research by sharing data. *Oncologist* 20(5):e464-e420. doi:10.1634/theoncologist.2014-0431.

formation, and since 2017, it has required that data be submitted in a format that adheres to the Clinical Data Interchange Standards Consortium (CDISC) standards [46,47]. Through Policy 0070, uniform data preparation and documentation will also soon be required for any medical product trial submitted to the EMA. The policy currently requires publication of clinical study reports but will include IPD at a later date [21]. Similarly, as noted in the case examples, data must be uniformly indexed or cataloged so that they can be located when requests are received. For BioLINCC, it took between 85 and 350 hours to prepare IPD and supporting materials for each individual study, depending on data complexity and documentation quality [32]. Given the costs and effort associated with data sharing, there is a need to deploy educational efforts that encourage researchers to use robust clinical trial designs and develop clear and usable metadata and documentation as part of trial conduct, and provide guidance on developing data-sharing efforts in collaboration with stakeholders. There is also a need to improve the efficiency of IPD preparation, incentivize publication, and support the costs of data curation.

IPD Sharing Opportunities and Obstacles

Based on our review of data repositories, as well as findings from NAM's Clinical Effectiveness Research Innovation Collaborative and our individual experiences, we have identified five opportunities for addressing the critical obstacles to sharing IPD from clinical trials (see *Box 7*). In this section, we describe specific tasks for key stakeholders—including research teams, secondary data users, journal editors, research funders, data repository owners, institutional review boards, and others—to consider in order to take full advantage of the opportunities afforded by sharing clinical trial data. Similar discussions are occurring around data generated through longitudinal studies; routine health care delivery and health delivery system data warehouses; and patient-reported outcomes [48,49,50,51].

Box 6 | Case Example 6: Project Genomics Evidence Neoplasia Information Exchange (GENIE)

Description: Project GENIE of the American Association for Cancer Research (AACR) is a registry that aims to accelerate precision oncology by combining clinical cancer genomic data with clinical outcomes from cancer patients from eight academic institutions. The project intends to inform standards for aggregating, harmonizing, and sharing clinical sequencing data collected in routine medical practice.

Governance structure: AACR provides the funding, infrastructure, and governance to administer GENIE. GENIE is supported by AACR, Genentech, and Boehringer Ingelheim. GENIE uses a federated model in which all data reside at the participating institution and are made available as needed. Each participating institution signs a master participation agreement and a data use agreement. To access the data, users must create an account and agree to the terms of access. GENIE is governed by a steering committee, which includes representatives from each participating institution and members of AACR's leadership. The steering committee reports to an external advisory board and the AACR board.

Factors facilitating sharing: Factors include (1) no requirements for research proposals from public investigators; (2) the ability of the online platform to harmonize clinical genomic and patient-level data; (3) users' ability to access the data via an online analytics platform, cBioPortal, or download it directly via Sage Bionetworks; and (4) the use of a federated model that allows data to be stored locally and made available to others only after a defined period of institutional exclusivity.

Factors affecting costs and timeliness: Resource-intensive aspects of this initiative include (1) the requirement that participating institutions agree to provide a minimum of 500 records with specific requested clinical data elements and participate in ongoing meetings, and (2) the development and maintenance of the data synthesis and analysis platform.

Impact: GENIE's first set of cancer genomic data was made available in January 2017 and updated in January 2018. The registry includes data for over 60 major cancer types. The combined data includes 39,000 de-identified records. At least one article has been published demonstrating GENIE's utility.

SOURCE: Developed by authors. Description of case example sourced from: AACR Project GENIE Consortium. 2017. AACR project GENIE: Powering precision medicine through an international consortium. *Cancer Discovery* 7(8):818-831. doi:10.1158/2159-8290.CD-17-0151.

While there are unique considerations associated with all of these data resources, there are also several commonalities, many of which are reflected in the following discussion.

1. Improve Incentives for Data Sharing for Primary Researchers and Research Institutions, Including Academic Credit for the Generation of Rich Data Sources That Are Shared and Used

In most research fields, researchers are incentivized to maintain IPD ownership and not share with the wider scientific community, to maximize publication and other traditionally valued academic opportunities. To incentivize sharing within and among academic institutions, it may be necessary to develop a system of academic credit for trialists who generate and share data that acknowledges the effort required to conduct clinical trials and organize, clean, and store IPD; and that provides meaningful credit and other incentives for making that data available to others. Academic institutions should reward, celebrate, and highlight investigators who share, particularly those whose work leads to downstream contributions.

One of several recent suggestions is that publications identify the source and location of the datasets used as the basis of the manuscript—linking the researchers who make substantial contributions to data acquisition, quality control, creation and authoring of metadata, and curation—to assist in providing academic credit for these efforts [52,53]. The appropriate mechanism for making such attributions is a topic requiring continued discussion given the potentially large number of individuals involved in producing and curating health data [53]. Additionally, research journals should consider requesting specification of the data source as a citable reference to ensure that the researchers who collected the data and the data stewards receive credit in resulting publications and repositories should pro-

vide citations for each data set included, following the example set by Dryad.

Research funders should also consider strategies to promote IPD sharing. For example, funders could require IPD sharing prospectively in appropriate requests for proposals, providing clear instructions regarding data-sharing expectations; providing repositories and an integrated process, such as the one used by BioLINCC (see *Box 3*); and, as previously mentioned, including additional infrastructure support to help cover the costs of sharing, which include organizing and managing data for reuse, governance, and oversight.

An additional concern is that even after fulfilling sponsor agreements and other regulatory requirements for data sharing, without appropriate incentives for all of the parties involved, data could be shared without appropriate metadata or other needed context and resources. Further discussion is needed on developing a clearer understanding of jurisdiction, ownership, and responsibility for shared data and metadata (see topic five below). This exploration of the various forces that could help facilitate the process of IPD sharing and consideration of potential action items may promote necessary change in culture and practices in clinical research.

2. Create General Rules to Address Patient Consent and Privacy Issues, Anticipating Future Secondary Analyses and Sharing of Primary Clinical Trial Data.

To minimize the risk and maximize the benefit of data sharing, primary research teams, institutional review boards, the federal offices of human research protection (the US Department of Health and Human Services Office for Human Research Protections, the US Department of Veterans Affairs Office of Research Oversight, and the US Department of Defense Human Research Protection Office), and associated institutions should work to address patient consent and privacy issues to anticipate future secondary analyses and sharing of primary clinical trial data. To achieve this, a general framework should be developed to determine whether consent for secondary data use is needed from participants at the beginning of a new study, how to communicate that data collected will be shared, how to exclude or contact individuals who do not want their data shared without explicit consent, how to ensure data are sufficiently de-identified for new analyses, and how to streamline the development and use of appropriate data-use agreements. Ohmann et al. suggest several practices for attaining consent for secondary use of data, including offering a lay explanation of the potential benefits and harms of data sharing; how the data will be prepared, stored, and accessed; and the practical difficulty of trial participants withdrawing their consent [9].

In addition, where possible, data intended to be shared should be prepared with de-identification in mind. Both PLoS and ICMJE data policy indicate that investigators should share de-identified data underlying their published clinical trials results [19,54]. Additional guidance regarding novel and standardized strategies to de-identify data should be broadly disseminated. In particular, there is a need for a conceptual framework and terminology describing potential causes and consequences of re-identification and potential types of identifiers and their risk of re-identification.[6] Members of the public and study participants should be engaged in the development of such a framework. The potential risks of re-identification, which may increase if de-identified IPD are combined with existing public information, include violation of patient privacy and medical identity theft, and may disproportionately affect minorities [15]. Repositories such as the YODA Project (see *Box 1*) include language in their data use agreements that explicitly forbids activities that may cause re-identification.

3. Consider the Operational Expenses Associated with Data Repositories and Develop a Framework to Identify the Stakeholders and Resources Necessary to Cover Those Operational Costs

According to Wilhelm, Oster, and Shoulson: "The investigators who lead the Alzheimer's Disease Neuroimaging Initiative have estimated that across the lifetime of the nearly $130 million project, 10% to 15% of the total costs will have been dedicated to data-sharing activities and that investigators will have spent about 15% of their time on data-sharing tasks, such as uploading data or responding to queries from outside researchers" [55]. While this example is illustrative, it is important to recognize that costs can vary substantially based on several factors, including the model used for data sharing and access, availability of technical assistance, the extensiveness of procedures for reviewing data requests, and the need and intensity for legal review of DUAs. If IPD sharing is to be more widely adopted, the associated costs should be better understood by investigators, their institutions, repositories, and

6 A preliminary overview of these concepts is provided in the Institute of Medicine's 2015 report *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk* (doi: 10.17226/18998).

Table 1 | Key Characteristics of IPD Sharing Case Examples

Case	Data-Sharing Model[5]	Data Access Criteria	Decision-Making Entity	Transparency	Time Limit
1: YODA Project	Controlled (gatekeeper-federated)	Completeness, scientific merit	Project steering committee, external reviewers	Sponsors, review process, metrics	Data-use agreement expires after one year (renewable)
2: SOAR	Controlled (gatekeeper-federated)	Scientific rationale, dissemination plan, qualifications, and analysis plans	Project staff, institutional review board	Sponsors, review process	N/A
3: BioLINCC	Controlled (gatekeeper-federated)	Significance, approach, feasibility	Project staff, funding organization	Sponsors, review process, metrics	Research materials distribution agreement expires after three years
4: CSDR	Controlled (gatekeeper-federated)	Feasibility, conflicts of interest	External organization, independent review panel	Sponsors, review process, metrics	Data-use agreement expires after one year
5: PDS	Open access	Exclusion on the basis of FDA's debarment list	Project steering committee	Sponsors, review process, metrics	Access granted for one year
6: GENIE	Open access	N/A	Project steering committee	Sponsors, review process	N/A

5 Data-sharing models are strategies for granting access to patient data and materials in order to address current clinical research challenges. Open access models are "characterized by the absence of any review panel or decision maker." Controlled access models are characterized by "some form of control by either the donor (i.e., patient), the data provider (i.e., initial organization), or an independent party." Gatekeeper models are a type of controlled model where "access to data is not at the data providers' discretion but may be granted by a distinct entity," often an institutional review board. Gatekeeper models can either be centralized or federated. With a centralized approach, data are collected and housed as part of a repository, whereas with a federated approach, the data are stored by the data providers but information about those data is made available through a web-based search system. SOURCE: Broes, S., D. Lacombe, M. Verlinden, and I. Huys. 2018. Toward a tiered model to share clinical trial data and samples in precision oncology. Frontiers in Medicine 5:6. doi:10.3389/fmed.2018.00006.

SOURCE: Developed by authors. Data sharing models adapted from: Broes, S., D. Lacombe, M. Verlinden, and I. Huys. 2018. Toward a tiered model to share clinical trial data and samples in precision oncology. *Frontiers in Medicine* 5:6. doi:10.3389/fmed.2018.00006.

Box 7 | Five Opportunities for Addressing Major Obstacles to Individual Patient-Level Data Sharing from Clinical Trials

1. Improve incentives for data sharing for primary researchers and research institutions, including academic credit for the generation of rich data sources that are shared and used.
2. Create general rules to address patient consent and privacy issues, anticipating future secondary analyses and sharing of primary clinical trial data.
3. Consider the operational expenses associated with data repositories and develop a framework to identify the stakeholders and resources necessary to cover those operational costs.
4. Develop a conceptual framework to specify what clinical trial data and associated metadata should be shared, organized, and stored, including whether stored data should be raw or derived.
5. Develop guidelines for how data repositories can promote meaningful data sharing, and for how to select an appropriate repository platform.

SOURCE: Developed by authors.

research funding organizations, and such costs should be potentially included as part of the research process. There are also recurring costs associated with data curation and data repositories that need a reliable funding stream. For example, to ensure that appropriate personnel are available to review reanalysis requests, it is necessary to have sufficient funds to support staff time after the normal conclusion of a study. These costs may also be borne by repositories such as CSDR (see *Box 4*), which relies on independent external reviewers, and PDS (see *Box 5*), which allows data sponsors to conduct an additional review of risks related to intellectual property and competitive advantage. Additionally, data repositories require considerable resources to provide governance, such as the development of common data request forms and processes.

A list of potential data-sharing tasks that may require funding, based on costs highlighted by case examples, is provided in *Box 8.* While it may be possible for primary researchers and repository owners to write some of these additional costs into funding requests, another possibility might be for funding agencies to consider separate funding streams for IPD sharing, since many funders may have an interest in extending the impact of their grantmaking by making data from funded studies easier to share. Funders could also consider developing mechanisms that support secondary research projects conducted either by the original research team or by other investigators. Ohmann et al. suggest avoiding access fees to data where possible, but they note that in some cases, "the costs of preparing data for sharing may need to be met by the secondary users" [9]. While promoting access by avoiding user fees may be possible in some situations, data repositories should be able to develop their own business models based on their needs and existing support. Existing secondary research using administrative claims data, such as Medicare and Truven Health MarketScan data, may offer useful contracting models and data-use agreements.

4. Develop a Conceptual Framework to Specify What Clinical Trial Data and Associated Metadata Should Be Shared, Organized, and Stored, Including Whether Stored Data Should Be Raw or Derived

To achieve widespread sharing of IPD, it is necessary to establish a framework and standards for data documentation, organization, and storage, for processes related to coding and analysis, and for ensuring appropriate personnel are available to review reanalysis requests and expedite decision making. Collected data should use standardized formats that facilitate use by secondary research teams and merge data from multiple trials and sponsors where possible. In the absence of such standards, primary investigators and others may be burdened by having to develop post hoc informatics tools to transform data in order to facilitate use [53]. For example, although the CDISC is widely used by industry because the data it contains follow formats required by the FDA and EMA, it is not used by the NIH and its funded researchers, which do not have such format requirements.

Box 8 | Potential Data-Sharing Tasks That May Require Funding

1. Coordinating and reviewing vendor activities (if outsourced), which could include removing personal identifiers from individual patient-level data, reviewing the protocol and statistical analysis plan, removing commercial confidential information from case report forms, and checking consent forms for data-sharing restrictions
2. Redacting historical clinical documents, which is relevant for sponsors sharing clinical trials prior to committing to data sharing
3. De-identifying and/or anonymizing data and documents, removing or recoding identifying variables or excluded cases, and investigating low-frequency cases
4. Translating data into the standardized format of the repository
5. Information hosting for clinical trial data and associated metadata
6. Managing and tracking data-sharing requests
7. Assisting investigators with data questions and related issues
8. Convening an independent review panel and relevant administration and infrastructure (e.g., request intake portal, request processing, metrics, etc.; will vary depending on data-sharing model)
9. Implementing and updating a secure data platform for hosting participant-level data and supporting documents

SOURCE: Developed by authors.

In addition to data standardization, it is necessary for trialists to share key metadata, including protocols, data dictionaries, statistical analysis plans, and template case report forms. To facilitate more consistent metadata across repositories, Canham and Ohmann propose a schema for metadata that captures, "(a) study identification data, including links to clinical trial registries; (b) data object characteristics and identifiers; and (c) data covering location, ownership and access to the data object." [568 In addition, repositories such as the YODA Project (see *Box 1*) require sponsors to prepare metadata and other documentation for external sharing to allow for subgroup analyses and other uses. With respect to clinical trial protocols, several of the repositories described in the case examples, such as PDS (see *Box 4*), include study protocols in their data query system, while others—such as the YODA Project (see *Box 1*) and CSDR (see *Box 3*)—provide access to protocols upon approval of data requests. Ohmann et al. provide suggestions on how to develop consistent citable identifiers for repositories, protocols, and datasets [9]. One potential action item for improving metadata is building on existing literature to develop standards or guidelines for investigators to facilitate defining derived variables, provide the rationales for defining such derived variables, and describe the expected responsibilities of secondary research teams aiming to perform replication or subgroup studies [57. Similar guidance has been developed for secondary analyses of administrative and health care data and may be used as a model [58,59].

5. Develop Guidelines For How Data Repositories Can Promote Meaningful Data Sharing, and for How to Select an Appropriate Repository Platform

Repository platforms provide data holders with the ability to share their data in a systematic and accessible way. As described in Table 1, data repositories rely on a range of data access models, review criteria and bodies, data-use agreements, and data-sharing and analysis platforms. Decisions regarding these factors should be made based on their alignment with the type and scale of data being stored. For example, GENIE relies on cBioPortal, an open source platform that is uniquely engineered to support analyses and visualizations of cancer genomics data. Additionally, repository owners should provide appropriate data security for data transfer and analysis systems and overall governance, such as the development of common data request forms and processes. Instructions on how to use repositories' analysis environments should be made available to researchers. Repositories should engage in discussion and planning to determine how data repositories should interoperate to reduce the potential problems associated with having different datasets available in

varied data-sharing platforms and repositories with different requirements.

Conclusion

Sharing of IPD from clinical trials and, eventually, widespread sharing of routinely generated electronic health information is necessary for realizing the vision of a continuously learning health system. The obstacles and opportunities described in this paper are meant to contextualize actionable steps that should be taken by key stakeholders to engage in IPD sharing. We realize that these obstacles and opportunities will continuously evolve with the increase of IPD sharing initiatives and new lessons learned. In that vein, we would encourage pilot testing, future research, and collaboration on other topics that ultimately effect a culture of data sharing, as there remains a pressing need to generate high-quality evidence to support all that is done in clinical medicine. Although major work has been completed to actualize the vision of IPD sharing, there are still important action steps, identified in this paper, that must be addressed.

Ultimately, driven by a desire for high-quality science that enables new discoveries and dedicated individuals and institutions—and through continued engagement with the National Academy of Medicine, among other entities—we want to encourage individual investigators, regulators, scientists, and industry to continue to work to improve IPD sharing capabilities, with the belief that partnership and collaboration offers opportunity to advance science. We would also like to encourage federal and nonfederal funders to consider approaches for making funds available to support key data-sharing tasks, as outlined in *Box 8*. We hope that the action steps presented here will add to this effort and will reinforce the importance of robust clinical trial design and conduct.

References

1. Institute of Medicine. 2010. *Clinical data as the basic staple of health learning: Creating and protecting a public good: workshop summary.* Washington, DC: The National Academies Press. doi:10.17226/12212.
2. Zarin, D. A., and T. Tse. 2016. Sharing individual participant data (IPD) within the context of the trial reporting system (TRS). *PLOS Medicine* 13(1):e1001946. doi:10.1371/journal.pmed.1001946.
3. Rathore, S. S., J. P. Curtis, Y. Wang, M. R. Bristow, and H. M. Krumholz. 2003. Association of serum digoxin concentration and outcomes in patients with heart failure. *JAMA* 289(7):871-878.
4. Rathore, S. S., Y. Wang, and H. M. Krumholz. 2002. Sex-based differences in the effect of digoxin for the treatment of heart failure. *The New England Journal of Medicine* 347(18):1403-1411. doi:10.1056/NEJMoa021266.
5. Ross, J. S., D. Madigan, M. A. Konstam, D. S. Egilman, and H. M. Krumholz. 2010. Persistence of cardiovascular risk after rofecoxib discontinuation. *Archives of Internal Medicine* 170(22):2035-2036. doi:10.1001/archinternmed.2010.461.
6. Naudet, F., C. Sakarovitch, P. Janiaud, I. Cristea, D. Fanelli, D. Moher, and J. P. A. Ioannidis. 2018. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: Survey of studies published in The BMJ and PLOS Medicine. *BMJ* 360:k400. doi:10.1136/bmj.k400.
7. McKiernan, E. C., P. E. Bourne, C. T. Brown, S. Buck, A. Kenall, J. Lin, D. McDougall, B. A. Nosek, L. Ram, C. K. Soderberg, J. R. Spies, K. Thaney, A. Updegrove, K. H. Woo, and T. Yarkoni. 2016. How open science helps researchers succeed. *Elife* 5:e16800.
8. Institute of Medicine. 2015. *Sharing clinical trial data: Maximizing benefits, minimizing risk.* Washington, DC: The National Academies Press. doi:10.17226/18998.
9. Ohmann, C., R. Banzi, S. Canham, S. Battaglia, M. Matei, C. Ariyo, L. Becnel, B. Bierer, S. Bowers, L. Clivio, M. Dias, C. Druml, H. Faure, M. Fenner, J. Galvez, D. Ghersi, C. Gluud, T. Groves, P. Houston, G. Karam, D. Kalra, R. L. Knowles, K. Krleža-Jerić, C. Kubiak, W. Kuchinke, R. Kush, A. Lukkarinen, P. S. Marques, A. Newbigging, J. O'Callaghan, P. Ravaud, I. Schlünder, D. Shanahan, H. Sitter, D. Spalding, C. Tudur-Smith, P. van Reusel, E. B. van Veen, G. R. Visser, J. Wilson, and J. Demotes-Mainard. 2017. Sharing and reuse of individual participant data from clinical trials: Principles and recommendations. *BMJ Open* 7(12):e018647. doi:10.1136/bmjopen-2017-018647.
10. PLoS Medicine editors. 2016. Can data sharing become the path of least resistance? *PLOS Medicine* 13(1):e1001949. doi:10.1371/journal.pmed.1001949.
11. Tudur Smith, C., C. Hopkins, M. R. Sydes, K. Woolfall, M. Clarke, G. Murray, and P. Williamson. 2015. How should individual participant data (IPD) from

publicly funded clinical trials be shared? *BMC Medicine* 13(1):298. doi:10.1186/s12916-015-0532-z.

12. Ohmann, C., S. Canham, R. Banzi, W. Kuchinke, and S. Battaglia. 2018. Classification of processes involved in sharing individual participant data from clinical trials. *F1000Research* 7:138. doi:10.12688/f1000research.13789.2.
13. Rowhani-Farid, A., and A. G. Barnett. 2016. Has open data arrived at the British Medical Journal (BMJ)? An observational study. *BMJ Open* 6(10). doi:10.1136/bmjopen-2016-011784.
14. Gamble, C., A. Krishan, D. Stocken, S. Lewis, E. Juszczak, C. Dore, P. R. Williamson, D. G. Altman, A. Montgomery, P. Lim, J. Berlin, S. Senn, S. Day, Y. Barbachano, and E. Loder. 2017. Guidelines for the content of statistical analysis plans in clinical trials. *JAMA* 318(23):2337-2343. doi:10.1001/jama.2017.18556.
15. Gibson, C. 2018. Moving from hope to hard work in data sharing. *JAMA Cardiology* 3(9):795-796. doi:10.1001/jamacardio.2018.0130.
16. Gay, H. C., A. S. Baldridge, and M. D. Huffman. 2018. Different population and end point definitions in reproduction analysis based on shared data—reply. *JAMA Cardiology* 3(9):894. doi:10.1001/jamacardio.2018.1791.
17. Natale, A., R. Stagg, and B. Zhang. 2018. Different population and end point definitions in reproduction analysis based on shared data. *JAMA Cardiology* 3(9):893-894. doi:10.1001/jamacardio.2018.1788.
18. Wilkinson, M. D., M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J. W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S. A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3:160018. doi:10.1038/sdata.2016.18.
19. International Committee of Medical Journal Editors. 2018. *Clinical trials: Registration and data sharing.* http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html#two (accessed June 13, 2019).
20. Bonini, S., H. G. Eichler, N. Wathion, and G. Rasi. 2014. Transparency and the European Medicines Agency—sharing of clinical trial data. *The New England Journal of Medicine* 371(26):2452-2455. doi:10.1056/NEJMp1409464.
21. *External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use.* 2017. The Netherlands: European Medicines Agency.
22. *Food and Drug Administration Amendments Act of 2007,* Pub. L. No. 110-85.
23. Moorthy, V. S., G. Karam, K. S. Vannice, and M. P. Kieny. 2015. Rationale for WHO's new position calling for prompt reporting and public disclosure of interventional clinical trial results. *PLOS Medicine* 12(4):e1001819. doi:10.1371/journal.pmed.1001819.
24. National Institutes of Health. 2016. *NIH policy on the dissemination of NIH-funded clinical trial information.* https://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-149.html (accessed June 13, 2019).
25. US Department of Health and Human Services. 2016. *42 CFR Part 11: Clinical trials registration and results information submission; final rule.* Federal Register 81(183):64982-65157.
26. US Department of Veterans Affairs. 2018. *ORD sponsored clinical trials: Registration and submission of summary results.* https://www.research.va.gov/resources/ORD_Admin/clinical_trials (accessed June 13, 2019).
27. World Medical Association. 2013. World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA* 310(20):2191-2194. doi:10.1001/jama.2013.281053.
28. US Food and Drug Administration. 2018. *Study data for submission to CDER and CBER.* https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/ucm587508.htm (accessed June 13, 2019).
29. McCall, B. 2018. What does the GDPR mean for the medical community? *Lancet* 391(10127):1249-1250. doi:10.1016/S0140-6736(18)30739-6.
30. Krumholz, H. M., C. P. Gross, K. L. Blount, J. D. Ritchie, B. Hodshon, R. Lehman, and J. S. Ross. 2014. Sea change in open science and data sharing: Leadership by industry. *Circulation: Cardiovascular Quality and Outcomes* 7(4):499-504. doi:10.1161/cir-

coutcomes.114.001166.

31. AACR Project GENIE Consortium. 2017. AACR project GENIE: Powering precision medicine through an international consortium. *Cancer Discovery* 7(8):818-831. doi:10.1158/2159-8290.CD-17-0151.
32. Coady, S. A., G. A. Mensah, E. L. Wagner, M. E. Goldfarb, D. M. Hitchcock, and C. A. Giffen. 2017. Use of the National Heart, Lung, and Blood Institute data repository. *The New England Journal of Medicine* 376(19):1849-1858. doi:10.1056/NEJMsa1603542.
33. Duke Clinical Research Institute. 2018. *SOAR data: Available datasets: Duke cardiac catheterization datasets.* https://dcri.org/our-approach/data-sharing/soar-data (accessed June 13, 2019).
34. Krumholz, H. M., and J. Waldstreicher. 2016. The Yale Open Data Access (YODA) Project—a mechanism for data sharing. *The New England Journal of Medicine* 375(5):403-405. doi:10.1056/NEJMp1607342.
35. National Institutes of Health. 2017. *NIH data sharing repositories.* https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html (accessed June 13, 2019).
36. Springer Nature Limited. 2018. *Recommended data repositories.* https://nature.com/sdata/policies/repositories (accessed June 13, 2019).
37. Strom, B. L., M. Buyse, J. Hughes, and B. M. Knoppers. 2014. Data sharing, year 1—access to data from industry-sponsored clinical trials. *The New England Journal of Medicine* 371(22):2052-2054. doi:10.1056/NEJMp1411794.
38. Broes, S., D. Lacombe, M. Verlinden, and I. Huys. 2018. Toward a tiered model to share clinical trial data and samples in precision oncology. *Frontiers in Medicine* 5:6. doi:10.3389/fmed.2018.00006.
39. Sydes, M. R., A. L. Johnson, S. K. Meredith, M. Rauchenberger, A. South, and M. K. Parmar. 2015. Sharing data from clinical trials: The rationale for a controlled access approach. *Trials* 16(1):104. doi:10.1186/s13063-015-0604-6.
40. Coady, S. A., and E. Wagner. 2013. Sharing individual level data from observational studies and clinical trials: A perspective from NHLBI. *Trials* 14:201. doi:10.1186/1745-6215-14-201.
41. Giffen, C. A., L. E. Carroll, J. T. Adams, S. P. Brennan, S. A. Coady, and E. L. Wagner. 2015. Providing contemporary access to historical biospecimen collections: Development of the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC). *Biopreservation Biobank* 13(4):271-279. doi:10.1089/bio.2014.0050.
42. Ross, J. S., J. D. Ritchie, E. Finn, N. R. Desai, R. L. Lehman, H. M. Krumholz, and C. P. Gross. 2016. Data sharing through an NIH central database repository: A cross-sectional survey of BioLINCC users. *BMJ Open* 6(9):e012769. doi:10.1136/bmjopen-2016-012769.
43. Navar, A., M. J. Pencina, J. A. Rymer, D. M. Louzao, and E. D. Peterson. 2016. Use of open access platforms for clinical trial data. *JAMA* 315(12):1283-1284. doi:10.1001/jama.2016.2374.
44. The YODA Project. *Submitted requests to use Johnson & Johnson data.* https://yoda.yale.edu/metrics/submitted-requests-use-johnson-johnson-data (accessed June 13,, 2019).
45. Zhu, C. S., P. F. Pinsky, J. E. Moler, A. Kukwa, J. Mabie, J. M. Rathmell, T. Riley, P. C. Prorok, and C. D. Berg. 2017. Data sharing in clinical trials: An experience with two large cancer screening trials. *PLOS Medicine* 14(5):e1002304. doi:10.1371/journal.pmed.1002304.
46. US Food and Drug Administration. 2019. *Study data standards resources.* https://www.fda.gov/ForIndustry/DataStandards/StudyDataStandards/default.htm#Catalog (accessed June 13, 2019).
47. US Food and Drug Administration. 2014. *Providing regulatory submissions in electronic format—standardized study data.* https://www.fda.gov/downloads/drugs/guidances/ucm292334.pdf (accessed June 13, 2019).
48. Budin-Ljosne, I., J. Isaeva, B. M. Knoppers, A. M. Tasse, H. Y. Shen, M. I. McCarthy, and J. R. Harris. 2014. Data sharing in large research consortia: Experiences and recommendations from ENGAGE. *European Journal of Human Genetics* 22(3):317-321. doi:10.1038/ejhg.2013.131.
49. Hickner, J., and L. A. Green. 2015. Practice-based research networks (PBRNs) in the United States: Growing and still going after all these years. *The Journal of the American Board of Family Medicine* 28(5):541-545. doi:10.3122/jabfm.2015.05.150227.
50. Poldrack, R. A., and K. J. Gorgolewski. 2014. Making big data open: Data sharing in neuroimaging. *Nature Neuroscience* 17(11):1510-1517. doi:10.1038/nn.3818.
51. van Panhuis, W. G., P. Paul, C. Emerson, J. Grefenstette, R. Wilder, A. J. Herbst, D. Heymann, and D. S. Burke. 2014. A systematic review of barriers to data sharing in public health. *BMC Public Health* 14:1144. doi:10.1186/1471-2458-14-1144.
52. Bierer, B. E., M. Crosas, and H. H. Pierce. 2017. Data authorship as an incentive to data sharing. *The New*

England Journal of Medicine 376(17):1684-1687. doi:10.1056/NEJMsb1616595.

53. Peterson, E. D., and F. W. Rockhold. 2018. Finding means to fulfill the societal and academic imperative for open data access and sharing. *JAMA Cardiology* 3(9):793-794. doi:10.1001/jamacardio.2018.0129.
54. Silva, L. 2014. PLOS' new data policy: Public access to data. *EveryONE: PLOS ONE Community Blog, February 24.* http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2 (accessed June 13, 2019).
55. Wilhelm, E. E., E. Oster, and I. Shoulson. 2014. Approaches and costs for sharing clinical research data. *JAMA* 311(12):1201-1202. doi:10.1001/jama.2014.850.
56. Canham, S., and C. Ohmann. 2016. A metadata schema for data objects in clinical research. *Trials* 17(1):557. doi:10.1186/s13063-016-1686-5.
57. Lo, B., and D. L. DeMets. 2016. Incentives for clinical trialists to share data. *The New England Journal of Medicine* 375(12):1112-1115. doi:10.1056/NEJMp1608351.
58. Berger, M. L., H. Sox, R. J. Willke, D. L. Brixner, H. G. Eichler, W. Goettsch, D. Madigan, A. Makady, S. Schneeweiss, R. Tarricone, S. V. Wang, J. Watkins, and C. D. Mullins. 2017. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiology Drug Safety* 26(9):1033-1039. doi:10.1002/pds.4297.
59. Wang, S. V., S. Schneeweiss, M. L. Berger, J. Brown, F. de Vries, I. Douglas, J. J. Gagne, R. Gini, O. Klungel, C. D. Mullins, M. D. Nguyen, J. A. Rassen, L. Smeeth, and M. Sturkenboom. 2017. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies v1.0. *Pharmacoepidemiology Drug Safety* 26(9):1018-1032. doi:10.1002/pds.4295.

DOI

https://doi.org/10.31478/201906b

Suggested Citation

Kuntz, R. E., E. M. Antman, R. M. Califf, J. R. Ingelfinger, H. M. Krumholz, A. Ommaya, E. D. Peterson, J. S. Ross, J. Waldstreicher, S. V. Wang, D. A. Zarin, D. M. Whicher, S.M. Siddiqi, and M. Hamilton Lopez,. 2019. Individual patient-level data sharing for continuous learning: A strategy for trial data sharing. NAM Perspectives. Discussion Paper, National Academy of Medicine, Washington, DC. https://doi.org/10.31478/201906b

Author Information

Richard E. Kuntz, MD is Senior Vice President, Chief Medical and Scientific Officer of Medtronic. **Elliott M. Antman, MD, FAHA, MACC,** is professor of medicine and the Associate Dean for Clinical and Translational Research at Harvard Medical School. **Robert M. Califf, MD, MACC,** is Donald F. Fortin, M.D. Professor of Cardiology at Duke University. **Julie R. Ingelfinger, MD,** is Pediatrician and Senior Consultant in Pediatric Nephrology at Harvard Medical School and the MassGeneral Hospital for Children. **Harlan M. Krumholz, MD, SM,** is Director, Center for Outcomes Research and Evaluation at the Yale University School of Medicine. **Alexander Ommaya, DSc,** is Senior Director, Clinical and Translational Research and Policy at the Association of American Medical Colleges. **Eric D. Peterson, MD, MPH, FAHA, FACC,** is Fred Cobb, M.D. Professor of Medicine at the Duke Clinical Research Institute. **Joseph S. Ross, MD, MHS,** is Associate Professor of Medicine and Associate Professor of Public Health at the Yale University School of Medicine. **Joanne Waldstreicher, MD** is Chief Medical Officer at Johnson & Johnson. **Shirley V. Wang, PhD, MSc,** is Assistant Professor of Medicine at Harvard Medical School. **Deborah A. Zarin, MD** is Program Director, Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard. **Danielle M. Whicher, PhD, MHS,** is Senior Program Officer at the National Academy of Medicine. **Sameer M. Siddiqi, PhD,** was Technical Specialist at the National Academy of Medicine at the time of this paper's authoring. **Marianne Hamilton Lopez, PhD, MPA,** is Research Director, Value-Based Payment Reform at the Duke-Margolis Center for Health Policy.

Acknowledgments

The authors would like to thank **Robert Harrington** of Stanford University, **Lynn Hudson** of the Critical Path Institute, and **Rebecca Kush** of CDISC for their important contributions to this paper.

Conflict-of-Interest Disclosures

Dr. Antman is conducting a clinical trial with Daiichi Sankyo. Dr. Califf serves on the corporate board for Cytokinetics and is board chair of the People-Centered Research Foundation. He receives consulting fees from

Merck, Biogen, Genentech, Eli Lilly, and Boehringer Ingelheim, and is employed as an advisor by Verily Life Sciences (Alphabet). Dr. Ingelfinger is salaried deputy editor of the New England Journal of Medicine. Dr. Krumholz is chair of the Cardiac Scientific Advisory Board for UnitedHealth; participant on the Life Sciences Board for IBM Watson Health; member of the Advisory Board for Element Science; member of the Physician Advisory board for Aetna; member of Advisory Board for Facebook; and owns Hugo, a personal health information platform. Dr. Peterson is a consultant for Abiomed; Amgen, Inc.; AstraZeneca; Bayer AG; Janssen Pharmaceuticals; Livongo; and Sanofi-Aventis. Dr. Peterson receives research funding in grants partially funded by Amarin; American College of Cardiology; American Heart Association; Amgen, Inc.; AstraZeneca; Baseline Study, LLC; Bayer AG; Eli Lilly & Company; Genentech; Janssen Pharmaceuticals; Merck & Co.; Novartis; Reflexion Health; Regeneron; Sanofi-Aventis; and the Society of Thoracic Surgeons. Dr. Ross recieved support through a research grant from Medtronic, Johnson & Johnson, Blue Cross-Blue Shield Association, and the Laura and John Arnold Foundation. Dr. Waldstreicher is an employee of Johnson & Johnson and a former employee of Merck & Co. Dr. Wang receives research funding in grants partially funded by Novartis Pharmaceuticals, Boehringer Ingelheim, and Johnson & Johnson. Dr. Zarin serves as a consultant for the National Library of Medicine.

Disclaimer

The views expressed in this paper are those of the authors and not necessarily of the authors' organizations, the National Academy of Medicine (NAM), or the National Academies of Sciences, Engineering, and Medicine (the National Academies). The paper is intended to help inform and stimulate discussion. It is not a report of the NAM or the National Academies.

PLOS | MEDICINE

ESSAY

Sharing Individual Participant Data (IPD) within the Context of the Trial Reporting System (TRS)

Deborah A. Zarin*, Tony Tse*

National Library of Medicine, National Institutes of Health, Bethesda, Maryland, United States of America

* dzarin@mail.nih.gov (DAZ); atse@mail.nih.gov (TT)

OPEN ACCESS

Citation: Zarin DA, Tse T (2016) Sharing Individual Participant Data (IPD) within the Context of the Trial Reporting System (TRS). PLoS Med 13(1): e1001946. doi:10.1371/journal.pmed.1001946

Published: January 19, 2016

Funding: This article was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health. The views expressed in this article are those of the authors and do not necessarily reflect the positions of the NIH. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: We have read the journal's policy and have the following conflicts: DAZ is Director and TT is an Analyst for ClinicalTrials.gov.

Abbreviations: CONSORT, Consolidated Standards of Reporting Trials; CRE, clinical research enterprise; CRF, case report form; DICOM, Digital Imaging and Communications in Medicine; FDG, fluorodeoxyglucose; GSK, GlaxoSmithKline; IOM, Institute of Medicine; IPD, individual participant data; NLM, National Library of Medicine; PET, positron emission tomography; RIAT, Restoring Invisible and

Summary Points

- The role of individual participant data (IPD) sharing can best be understood as part of an overall three-level trial reporting system (TRS) framework.
- Different "types" of IPD, which reflect varying degrees of information granularity, have different potential benefits and harms.
- Study 329 of Paxil (paroxetine) in children with depression is used as a case study to highlight the potential value of different components of the TRS.

The Institute of Medicine (IOM) [1], journal editors [2,3], and many others [4–6] have called for more widespread, third-party access to the individual participant data (IPD) and associated documentation from clinical trials (i.e., "IPD sharing"). Advocates assert that access to trial IPD will help to address well-established flaws in the current system of communicating trial results, including nonpublication, selective reporting, and lack of reproducibility [7]. Additional proposed benefits include the ability to reanalyze study data (e.g., validation and/or correction of previously published findings [8]) and to combine data from multiple studies (e.g., IPD-level meta-analyses [9]). Others note the burdens and costs associated with preparing IPD and associated documentation for sharing, the need to ensure participant privacy, and the risk of invalid analyses [10].

We do not attempt to replicate the more comprehensive analysis of IPD sharing that was conducted by the recent IOM panel [1]. However, we believe that it would be helpful at this pivotal time to consider the implications of IPD sharing within the context of the "trial reporting system" (TRS), which encompasses existing efforts to enhance access to information about trials and their findings and to improve the transparency of the clinical research enterprise (CRE) [11]. In this essay, we attempt to add precision to the ongoing discussion by examining the range of information granularity associated with different types of IPD. We then consider IPD sharing within a three-level TRS framework and illustrate the roles of these levels with a case study.

Abandoned Trials; SSRI, selective serotonin reuptake inhibitor; TRS, trial reporting system.

Provenance: Commissioned; externally peer reviewed

What Is the Nature of IPD?

As attention shifts to IPD sharing, it is instructive to consider the mechanism by which initial "raw" data collected from each trial participant are analyzed, transformed, and aggregated into the summary data reported in the results sections of journal articles, conference abstracts, press releases, and package inserts and as entries in results databases (Fig 1).

Each arrow in Fig 1 indicates a transformation of trial data. While some transformations are based on procedures prespecified in study documents (e.g., detailed criteria or algorithms in the protocol or statistical analysis plan), others likely rely on ad hoc expert judgments. For example, analyzing IPD collected for the primary outcome measure of "change in tumor size from baseline at 3 months" might involve the following decisions:

- choosing a specific imaging approach (e.g., fluorodeoxyglucose (FDG)-positron emission tomography (PET) using a specific device);
- determining a particular method for transforming 2- or 3-D images into tumor size measurements (e.g., Digital Imaging and Communications in Medicine [DICOM] standard using autocontouring to calculate the volume for the region of interest);
- applying these methods to measure tumor size for each individual at baseline and at 3 months; and
- calculating and recording the changes in size per participant.

Additional decisions must be made by the researchers about the handling of missing data, unreadable images, and other data deficiencies; determining the analysis population (e.g., all who started the study [including those who discontinued] or only those who received the full course of treatment); and aggregating the IPD for purposes of reporting and analysis (e.g., mean change in size versus proportion with a change over a certain size). The most granular data (far left in Fig 1) would provide insight into these decisions and allow independent

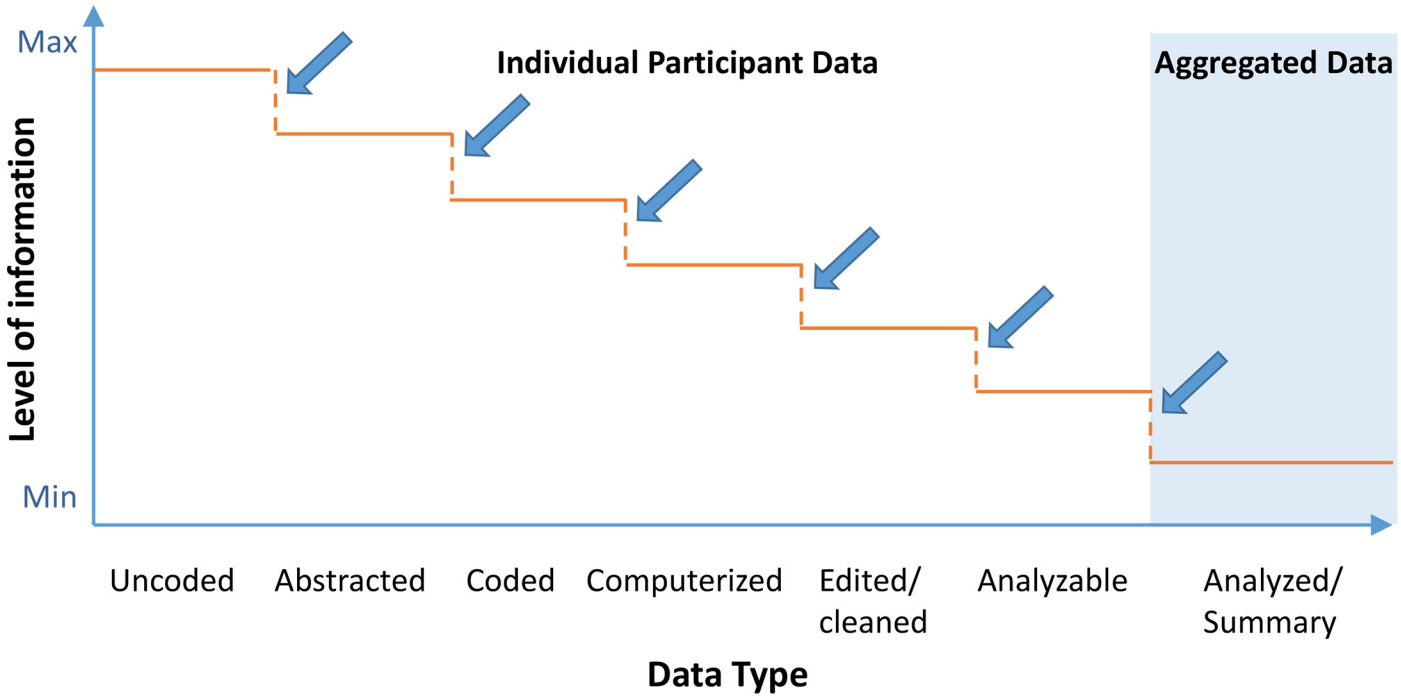


Fig 1. Schematic depicting information granularity for different types of data [12].

doi:10.1371/journal.pmed.1001946.g001

researchers to examine the implications of alternative analytic decisions. On the other hand, the least granular IPD (far right) would obscure some of these decisions and would not allow for testing the impact of different analytic methods.

Most discussions of IPD sharing policies sidestep the issue of matching IPD types with anticipated benefits and burdens. For example, third-party researchers interested in independently recoding the IPD would need access to uncoded data (i.e., data types to the left of "Coded" on the *x*-axis in Fig 1). In contrast, users who intend to replicate and confirm the reproducibility of aggregate data published in a journal article may only require access to the analyzable IPD (i.e., final type of IPD before undergoing transformation into aggregated data in Fig 1). While not an insurmountable barrier for IPD sharing policies, we believe that consideration of various data types and their uses is a timely issue for discussion within the research community, including questions such as the following:

- What standard terminology or classification should be used to describe the different data types?
- Which types of IPD should be made available systematically?
- When more than one type is available for sharing, how should they be uniquely identified and tracked (e.g., cited) within the research community?

Where Does IPD Fit in the TRS?

The TRS framework encompasses key existing and proposed efforts and is designed to increase trial transparency systematically. Fig 2 depicts the TRS as a pyramid with prospective registration at its base, summary or aggregate trial results reporting in the middle, and the sharing of trial IPD and relevant documents at its apex.

At its base, prospective registration provides a public listing of all ongoing and completed trials, along with key protocol and administrative details to allow people to identify the full set of trials conducted within a research area (e.g., antidepressant trials in children). Trial registration, if done and used appropriately, also allows for the assessment of fidelity to key protocol details, such as definition of the prespecified primary outcome measure [13]. Summary results reporting in trial registries, currently implemented at ClinicalTrials.gov and the European Union Clinical Trials Registry [14], is the next level of the TRS. Results databases—designed to ensure that aggregate trial results are reported systematically in a timely, structured, and complete manner based in part on expert trial-reporting guidelines such as the Consolidated Standards of Reporting Trials (CONSORT) statement [15] and its extensions—call attention to unacknowledged deviations from the registered protocol details [13]. Current policies are generally intended to address these two foundational levels of the TRS.

Registration information and summary results displayed as a single trial record provide the minimal, essential information needed to understand a trial and its findings. Each record also uses a format that is highly structured and searchable by a range of criteria. Ideally, users could easily retrieve information about all completed or ongoing trials for a particular clinical or policy question (e.g., to identify a need for additional research or conduct a systematic review), avoiding the biases imposed by incomplete and selective publication. Trial registration and results records are also linked, via unique registry identifiers, to relevant peer-reviewed journal publications [16]. As the use of unique registry identifiers expands (e.g., systematic reviews and press releases), an extensive network of automated, explicit linkages can provide an even more useful way to identify publicly available information about a trial from the trial record itself (Fig 3).

IPD and related documents reside at the apex of this pyramid because they are most useful within the context of the two lower levels, which serve as the foundation. Without careful use

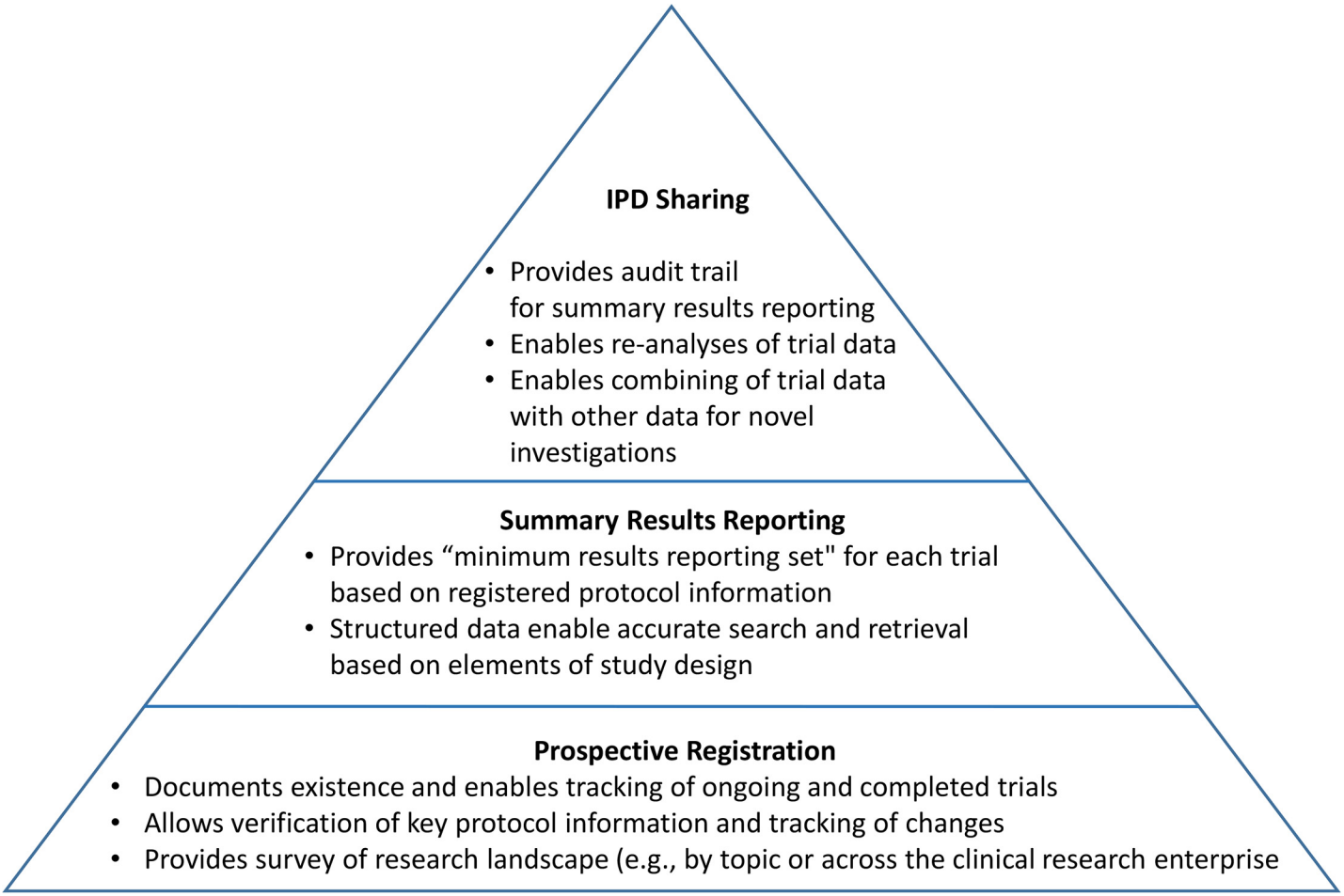


Fig 2. Schematic depicting the functions of the three key components of the TRS.

doi:10.1371/journal.pmed.1001946.g002

of trial registries and summary results databases, access to IPD might simply recreate or amplify existing reporting biases [17]. For example, analysis of trial IPD cannot mitigate biases that stem from selective release of data from only one trial among a "family" of trials for the studied population, intervention, and condition (e.g., a likely result of proposals to require the release of IPD only upon journal publication).

How Would the Three Key Components of TRS Work Together?

Case Study: Recent Reanalysis of Study 329

Study 329, sponsored by SmithKline Beecham (now GlaxoSmithKline [GSK]), was one of several studies conducted to examine the use of Paxil (paroxetine) in children with depression and the first with results to be published. The original publication of Study 329 in 2001 implied that the study results showed the safety and efficacy of Paxil in children [18]. In 2004, the New York State attorney general filed a consumer fraud lawsuit against GSK, alleging that the suppression and misreporting of trial data created the false impression that Paxil was safe and effective in depressed children [19].

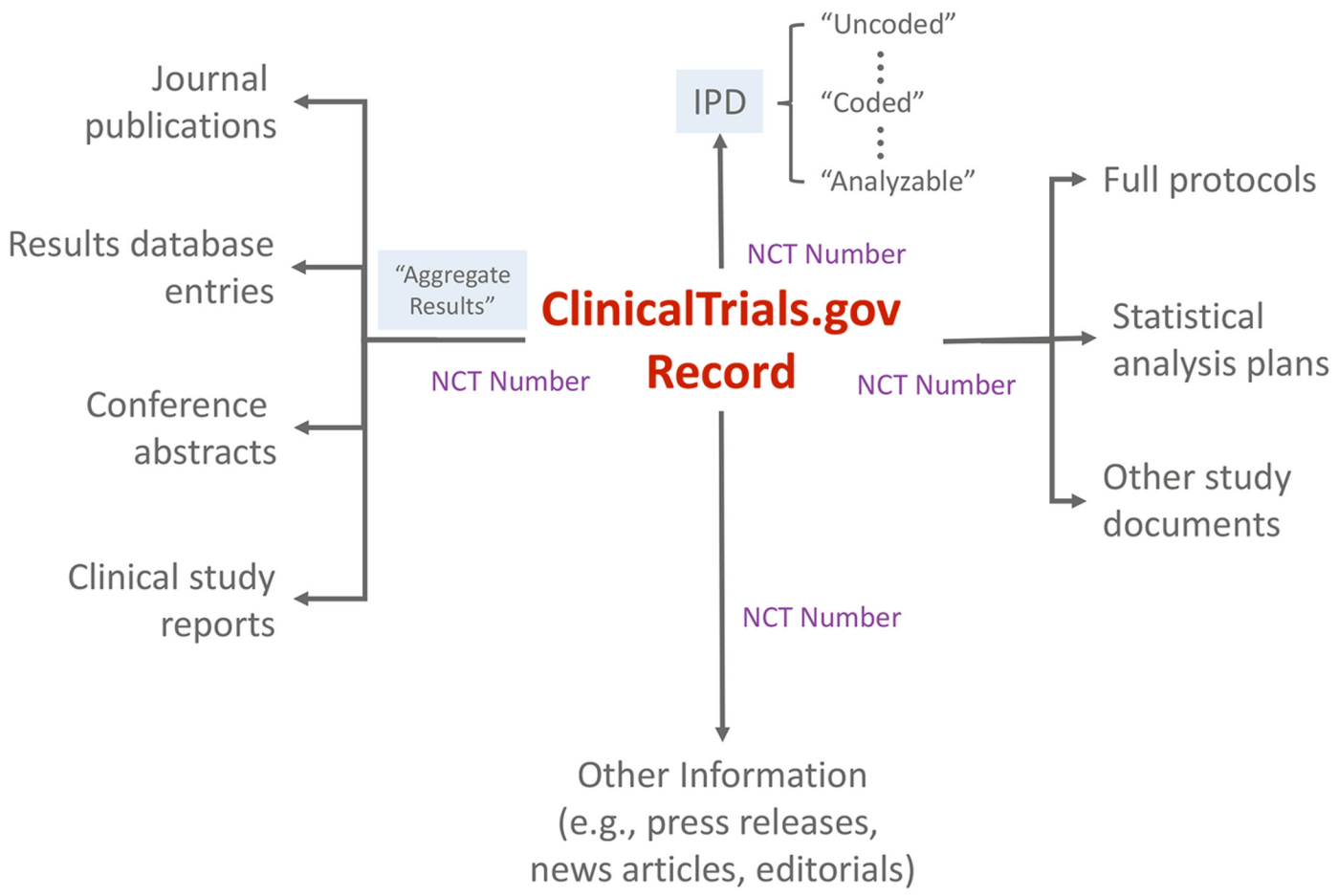


Fig 3. Schematic depicting ClinicalTrials.gov as an "information scaffold" using the record unique identifier (NCT number) to link to various online resources.

doi:10.1371/journal.pmed.1001946.g003

A newly published reanalysis, part of the Restoring Invisible and Abandoned Trials (RIAT) initiative [20], was based on access to original case report forms (CRFs) for 34% of the 275 participants [21]. These highly granular IPD datasets enabled the researchers to recategorize certain adverse events that they determined had been miscategorized originally (e.g., "mood lability" rather than the more serious "suicidality"). The reanalysis concluded that Study 329 did not show either efficacy or safety.

How Would the Problems of Study 329 Be Addressed by the Current TRS?

It would be an oversimplification to conclude that this reanalysis demonstrates the need to make IPD for all trials available. A more nuanced look at the specific problems is useful. Many of the concerns about Study 329 and the other Paxil studies might have been addressed if current policies regarding registration and results reporting had been in existence (Table 1, [22–24]). The key issue that specifically required access to IPD was the detection of miscategorization of some adverse events in the original report.

Table 1. Key issues with trials of antidepressant use in children for depression and the role of the TRS.

Key Issue	Relevant TRS Component	Comment
Lack of prospective public information about all trials of Paxil and other selective serotonin reuptake inhibitors (SSRIs) in depressed children	Prospective Registration	Registration would have provided a public list of all ongoing and completed trials of Paxil/SSRIs in depressed children
Alleged suppression of "negative" results from certain Paxil trials in depressed children [22]	Prospective Registration	Registration would have allowed the detection of trials without disclosed results
	Summary Results Reporting	Results database entries would have provided access to "minimum reporting set" including all prespecified outcome measures and all serious adverse events
Detection of selective reporting bias of efficacy and safety findings in the published results of Study 329, unacknowledged changes in outcome measures, and other issues [23]	Prospective Registration	Archival registration information would have allowed for the detection of unacknowledged changes in prespecified outcome measures and detection of nonprespecified outcome measures reported as statistically significant
	Summary Results Reporting	Structured reporting devoid of interpretation or conclusions would have made summary data publicly available while avoiding the possibility of spinning the results
Invalid and unacknowledged categorization of certain adverse events, resulting in the underreporting of suicidality [24]	Sharing Highly Granular IPD and Documents (e.g., CRFs)	Access to high-granularity IPD enabled the elucidation of data analytic decisions that had not been publicly disclosed; reanalysis was possible with different methods of categorizing adverse events

doi:10.1371/journal.pmed.1001946.t001

It is important to note that this illuminating reanalysis required access to the highly detailed IPD available in the original CRFs, represented by the far-left side of the *x*-axis in Fig 1. However, recent high-profile proposals for the sharing of IPD might not have added any clarity in the case of the Paxil studies in children beyond what could have been achieved with the optimal use of a registry and results database (i.e., two foundational levels of the pyramid in Fig 2). The reason is that journal publication serves as the "trigger" for IPD release in many of these proposals [1]), which could not possibly mitigate biases resulting from selective publication in the first place (i.e., IPD from unpublished trials would be exempt from sharing requirements). In addition, such proposed IPD policies call for the release of only the "coded" or "analyzable" dataset, which would not have allowed for the detection of miscategorization or the recategorization of the adverse events. Finally, such proposals would only require the sharing of a subset of IPD and documents for those aggregate data reported in the publication and not the full dataset, precluding secondary analyses intended to go beyond validation and reproducibility of the original publication.

Conclusion

The evolving TRS can be thought of as a pyramid, with each successive layer being dependent on the layer(s) below it. We should not allow the prospects for providing access to IPD and relevant documents to divert attention from the continuing need to ensure complete, accurate, and timely trial registration and summary results reporting—as well as attentive and consistent use of these tools by key stakeholders. In addition, IPD sharing policies and systems must consider the different benefits and burdens that would be expected from third-party access to data types of varying levels of granularity.

Acknowledgments

We thank Kristina V. Elliott, MLS, contractor to ClinicalTrials.gov, for her research assistance as a National Library of Medicine (NLM) associate fellow.

Author Contributions

Wrote the first draft of the manuscript: DAZ TT. Contributed to the writing of the manuscript: DAZ TT. Agree with the manuscript's results and conclusions: DAZ TT. All authors have read, and confirm that they meet, ICMJE criteria for authorship.

References

1. Institute of Medicine (2015) Sharing clinical trial data: maximizing benefits, minimizing risks. Washington DC: The National Academies Press.
2. Drazen JM (2015) Sharing individual patient data from clinical trials. N Engl J Med 372: 201–202. doi: 10.1056/NEJMp1415160 PMID: 25587944
3. Data Availability Policy. PLoS Med; 2016 [cited 2016 Jan 6]. http://journals.plos.org/plosmedicine/s/data-availability#loc-acceptable-data-sharing-methods.
4. Eichler HG, Abadie E, Breckenridge A, Leufkens H, Rasi G (2012) Open clinical trial data for all? A view from regulators. PLoS Med 9: e1001202. doi: 10.1371/journal.pmed.1001202 PMID: 22505851
5. Krumholz HM, Peterson ED (2014) Open access to clinical trials data. JAMA 312: 1002–1003. doi: 10.1001/jama.2014.9647 PMID: 25203080
6. Mello MM, Francer JK, Wilenzick M, Teden P, Bierer BE, et al. (2013) Preparing for responsible sharing of clinical trial data. N Engl J Med 369: 1651–1658. doi: 10.1056/NEJMhle1309073 PMID: 24144394
7. Song F, Parekh S, Hooper L, Loke YK, Ryder J, et al. (2010) Dissemination and publication of research findings: an updated review of related biases. Health Technol Assess 14: iii, ix-xi, 1–193. doi: 10.3310/hta14080 PMID: 20181324
8. Ebrahim S, Sohani ZN, Montoya L, Agarwal A, Thorlund K, et al. (2014) Reanalyses of randomized clinical trial data. JAMA 312: 1024–1032. doi: 10.1001/jama.2014.9646 PMID: 25203082
9. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, et al. (2015) Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. PLoS Med 12: e1001855. doi: 10.1371/journal.pmed.1001855 PMID: 26196287
10. Tudur Smith C, Dwan K, Altman DG, Clarke M, Riley R, et al. (2014) Sharing individual participant data from clinical trials: an opinion survey regarding the establishment of a central repository. PLoS ONE 9: e97886. doi: 10.1371/journal.pone.0097886 PMID: 24874700
11. Zarin DA, Tse T (2008) Medicine. Moving toward transparency of clinical trials. Science 319: 1340–1342. doi: 10.1126/science.1153632 PMID: 18323436
12. Zarin DA (2012) Presentation at IOM Workshop on Sharing Clinical Research Data. Washington DC.
13. Tse T, Zarin DA, Williams RJ, Ide NC (2012) The role and importance of clinical trial registries and results databases. In: Gallin JI, Ognibene FP, editors. Principles and Practice of Clinical Research 3rd ed Amsterdam: Academic Press. pp. 171–181.
14. EU Clinical Trials Register. European Medicines Agency. [cited 2016 Jan 6] https://www.clinicaltrialsregister.eu/ctr-search/search.
15. Schulz KF, Altman DG, Moher D, Group C (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med 7: e1000251. doi: 10.1371/journal.pmed.1000251 PMID: 20352064
16. Clinical Trial Registry Numbers in MEDLINE®/PubMed® Records. U.S. National Library of Medicine. [cited 2016 Jan 6] https://www.nlm.nih.gov/bsd/policy/clin_trials.html.
17. Zarin DA (2013) Participant-level data and the new frontier in trial transparency. N Engl J Med 369: 468–469. doi: 10.1056/NEJMe1307268 PMID: 23902488
18. Keller MB, Ryan ND, Strober M, Klein RG, Kutcher SP, et al. (2001) Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial. J Am Acad Child Adolesc Psychiatry 40: 762–772. PMID: 11437014
19. The People of the State of New York v. GlaxoSmithKline, Complaint, filed 2 June 2004. [cited 2016 Jan 6] http://news.findlaw.com/wsj/docs/glaxo/nyagglaxo60204cmp.pdf.
20. Doshi P, Dickersin K, Healy D, Vedula SS, Jefferson T (2013) Restoring invisible and abandoned trials: a call for people to publish the findings. BMJ 346: f2865. doi: 10.1136/bmj.f2865 PMID: 23766480
21. Le Noury J, Nardo JM, Healy D, Jureidini J, Raven M, et al. (2015) Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. BMJ 351: h4320. doi: 10.1136/bmj.h4320 PMID: 26376805
22. GSK Assurance of Discontinuance, 2004. [cited 2016 Jan 6] http://www.ag.ny.gov/sites/default/files/press-releases/archived/aug26a_04_attach2.pdf.

23. Jureidini JN, McHenry LB, Mansfield PR (2008) Clinical trials and drug promotion: Selective reporting of study 329. International Journal of Risk & Safety in Medicine 20: 76–81.

24. Leslie LK, Newman TB, Chesney PJ, Perrin JM (2005) The Food and Drug Administration's deliberations on antidepressant use in pediatric patients. Pediatrics 116: 195–204. PMID: 15995053

COMMENTARY

Moving Data Sharing Forward:
The Launch of the Vivli Platform

Rebecca Li, PhD, Vivli; **Jessica Scott, MD, JD,** GlaxoSmithKline; **Frank Rockhold, ScM, PhD,** Duke University; **Nina Hill, PhD,** Pfizer; **Julie Wood,** Vivli; and **Ida Sim, MD, PhD,** University of California at San Francisco and Vivli

November 19, 2018

Introduction

On January 14, 2015, the Institute of Medicine (IOM) released a consensus study report titled *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk* [1], which detailed guiding principles and a practical framework for responsibly managing clinical trial data to making data sharing easier. The foundational principles from the IOM report laid the groundwork for the current ecosystem of clinical trial data sharing and are guiding the implementation of a range of data sharing platforms.

The current culture and expectation for sharing of individual participant-level data (IPD) is rapidly changing. For example, the International Committee of Medical Journal Editors (ICMJE) [2] began requiring authors to provide IPD data sharing statements as of July 2018 as a condition for publication. This policy was enacted as investigators demonstrate a growing commitment to sharing trial data [3]. With journal editor policies, industry commitments, funder mandates, and regulatory requirements designed to broaden public access to this data, we appear to be at an inflection point. An increasing number of established and effective data sharing platforms and initiatives are now in existence to facilitate and coordinate access to these data. However, these platforms remain siloed from each other and from existing communities creating a fragmented ecosystem due to complex governance, policy, and security controls. As a result, cross-platform integration of data sets has been difficult [8].

In 2016 [4], the authors of this paper envisioned a new, independent, and neutral platform called Vivli that would link existing data sharing platforms and communities to break down these silos and amplify the value of data sharing, as well as provide data archiving capabilities to those who lack the ability to host data long-term. Figure 1 shows the timeline of the emergence of major data sharing platforms and initiatives and how the Vivli platform fits into the ecosystem. Vivli has established mechanisms and governance to enable sharing across several of these current platform initiatives despite the wide variety of data access mechanisms, ranging from those with more open access to those with highly controlled access. For users, providing access to a growing and diverse source of trial data has the potential to drive forward new scientific discoveries and insights.

Benefits of Participant-Level Data Sharing

Ultimately, how should the scientific value from these data sharing efforts be measured? Scientific advancement is a cumulative process, where findings from one study are replicated, refuted, verified, or extended by others. Data sharing has the possibility to incrementally advance our knowledge of mechanisms of action, predictors of efficacy, and effect sizes in different populations; can help shape the design of new trials in the planning stages; and myriad other critical pieces of knowledge that may eventually lead to influential landmark studies. Indeed, the majority of research projects

published on existing data sharing platforms do not replicate primary original findings but, rather, test new hypotheses [5]. Data sharing has the potential to increase the rate of scientific discovery, enhancing our ability to truly build on prior knowledge to maximize scientific output, while decreasing unnecessary risks to future participants.

Historically, data have been shared at the summary or aggregate level. Sharing participant-level data allows more powerful analyses, including new subgroup analyses, permitting the use of different analytic methods than was used in the original analysis, and analysis of time-to-event and continuous variables. We are now beginning to see demand for IPD from researchers as key findings, and examples of new science generated from IPD sharing is starting to be published. In 2016, the European Federation of Pharmaceutical Industries and Associations and the Pharmaceutical Research

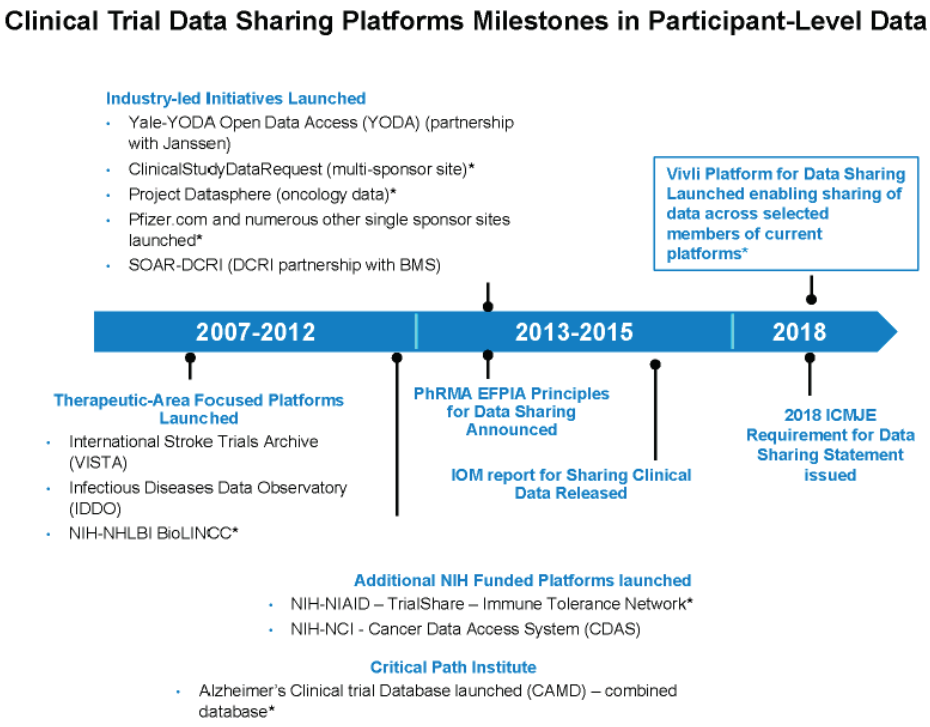


Figure 1 | Clinical Trial Data Sharing Platform Milestones in Participant-Level Data
SOURCE: Developed by authors

and Manufacturers of America jointly surveyed their members and found that 64% of their member companies received requests for data and the majority of these data requests were for participant-level data (679 out of 1,062 requests) [6]. Estimates are that over 330 peer-reviewed publications have resulted from the Project Data Sphere, NIH BioLINCC, Yale Yoda, CSDR and Pfizer data-sharing efforts.

The Vivli Platform

Vivli was designed to bridge the fragmentation in the current data sharing ecosystem and provide data archiving and hosting capacity. The authors of this paper understand the need that so much participant-level data was available for sharing, but could not be directly integrated. The platform was designed to bring together diverse sources of data for cross-platform sharing to enable the scientific community to utilize the contributions of trial participants beyond that from their original use. For example, the Vivli data use agreement ensures that the privacy rights of clinical trial participants are respected while advancing the goal of scientifically valid secondary analyses and balancing the interests of data contributors. Moreover, Vivli is implementing

advanced informatics technologies to meet FAIR (Findable, Accessible, Interoperable, Reusable) principles [7]. This cross-platform sharing will hopefully encourage and enable the incremental scientific advances described above.

Vivli is also neutral broker between data contributors and data requesters. The platform governance follows a number of central tenets, including transparently displaying the criteria for access, any exceptions to those criteria and how studies may be accessed. For requesters, Vivli can be used to search thousands of clinical studies that previously could not be easily searchable or integrated via a single portal. This robust search engine can be used by requesters which may help to target their search effectively and efficiently across multiple data contributors. The ability to find the specific data that can help scientists advance their work will enable them to work faster and more efficiently.

Vivli was created both as a response to the promise of IPD data sharing and to the frustration of not being able to use the data to its fullest extent.

In Closing

Key challenges ahead for IPD data sharing include more widespread adoption by data contributors of global data standards and the development of prospective plans for data sharing. Given the duration of a typical clinical trial, the data shared today are likely from studies planned when data sharing platforms were either non-existent or in their infancy. As culture shifts and data sharing becomes standard practice, users and contributors will need to work towards more unified processes and data standards. Additionally, there remain unresolved tensions between the potential risk of re-identification, evolving consensus standards for anonymization and maintaining data utility. Data sharing platforms have an important role to play in both shifting the culture and addressing these challenges with key partners. Ultimately, if we are successful, we will have respected the contributions of trial participants to science and maximized the opportunity to benefit human health.

References

1. Institute of Medicine (IOM). 2015. Sharing clinical trial data: Maximizing benefits, minimizing risk. Washington, DC: The National Academies Press. http://nationalacademies.org/HMD/Reports/2015/Sharing-Clinical-Trial-Data.aspx (accessed July 25, 2018).
2. Taichman, D. B., Sahni, P., Pinborg, A., Peiperl, L., Laine, C., James, A., ... & Frizelle, F. A. (2017). Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. Annals of Internal medicine, 167(1), 63-65.
3. Tannenbaum, S., Ross, J. S., Krumholz, H. M., Desai, N. R., Ritchie, J. D., Lehman, R., ... & Gross, C. P. (2017). Early experiences with journal data sharing policies: a survey of published clinical trial investigators. Annals of Internal Medicine. [Epub ahead of print 17 July 2018] doi: 10.7326/M18-0723 (accessed August 1, 2018).
4. Bierer, B. E., Li, R., Barnes, M., & Sim, I. (2016). A global, neutral platform for sharing trial data. New England Journal of Medicine, 374(25), 2411-2413.
5. Navar, A. M., Pencina, M. J., Rymer, J. A., Louzao, D. M., & Peterson, E. D. (2016). Use of open access platforms for clinical trial data. JAMA, 315(12), 1283-1284.
6. EFPIA-PHRMA. Principles for Responsible Clinical Trial Data Sharing: report on the 2106 Member Company Survey. 2016. https://www.efpia.eu/media/288603/efpia-phrma-report-on-the-2016-member-company-survey-on-the-joint-principles-for-responsible-clinical-trial-data-sharing.pdf (accessed November 11, 2018).
7. Wilkinson, M. D. et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. https://www.nature.com/articles/sdata201618 (accessed July 25, 2018).
8. Rockhold, F., P. Nisen, and A. Freeman. 2016. Data Sharing at a Crossroads. The New England Journal of Medicine 375:1115-1117. https://doi.org/10.1056/NEJMp1608086 (accessed November 11, 2018).

DOI

https://doi.org/10.31478/201811b

Suggested Citation

Li, R., J. Scott, F. Rockhold, N. Hill, J. Wood, and I. Sim. 2018. Moving Data Sharing Forward: The Launch of the Vivli Platform. *NAM Perspectives.* Commentary, National Academy of Medicine, Washington, DC. https://doi.org/10.31478/201811b.

Author Information

Rebecca Li, PhD, is Executive Director at Vivli. **Jessica Scott, MD, JD** authored this work in her role as the Director of North America Medical Policy and Advocacy

for GlaxoSmithKline. Dr. Scott is currently Head of R&D Patient Engagement at Takeda. **Frank Rockhold, ScM, PhD,** is a Professor of Biostatistics and Bioinformatics at Duke Clinical Research Institute. **Nina Hill, PhD,** is Vice President, Science Policy and Science Advocacy for Pfizer. **Julie Wood, BA,** is Director of Strategy and Operations at Vivli. **Ida Sim, MD, PhD,** is a Professor of Medicine at the University of California, San Francisco and co-directs Informatics and Research Innovation at UCSF's Clinical and Translational Sciences Institute; she is also Technical Lead at Vivli.

Acknowledgments

The authors are grateful to Jessica Baker for her assistance with manuscript preparation.

Conflict-of-Interest Disclosures

Rebecca Li and Julie Wood are currently employed by Vivli. Ida Sim is paid as a consultant by Vivli. Nina Hill is employed by Pfizer, Inc; Dr. Hill is not formally affiliated with Vivli. Frank Rockhold reports that he receives grant support from Janssen, AstraZeneca, BMS, Luitpold, Regeneron, PCORI and the NIH; that he receives personal fees from Merck Serono, Janssen, GlaxoSmithKline, Adverum, Eidos, Abbvie, Novo Nordisk, Amgen, and UCB; that he holds an equity interest in GlaxoSmithKline; and that he serves as Senior Advisor to Vivli. Jessica Scott is employed by Takeda; Dr. Scott is not formally affiliated with Vivli.

Correspondence

Questions or comments should be directed to Rebecca Li at rli@vivli.org

Disclaimer

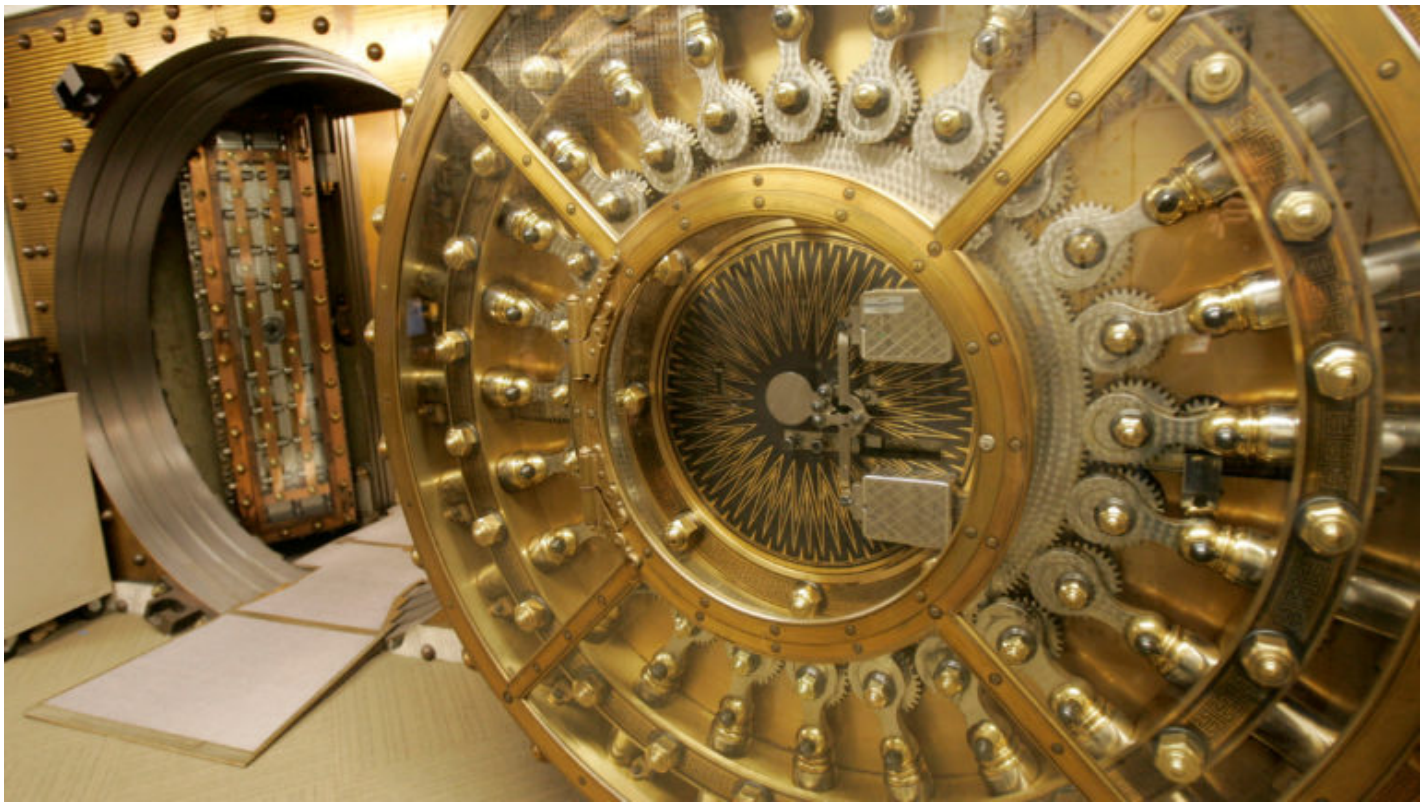
The views expressed in this paper are those of the authors and not necessarily of the authors' organizations, the National Academy of Medicine (NAM), or the National Academies of Sciences, Engineering, and Medicine (the National Academies). The paper is intended to help inform and stimulate discussion. It is not a report of the NAM or the National Academies.

STAT
1

Move clinical trial data sharing from an option to an imperative

By Rebecca Li

February 19, 2019



Jeff Chiu/AP

Data from clinical trials have long been locked away, some in this principal investigator's computer bank, some in that pharmaceutical company's cloud. For years we have been talking about opening up those vaults and freeing these data. The key has finally turned: Data sharing is becoming the new reality.

From Jan. 1, 2019, onward, the world's leading medical journals, including the New England Journal of Medicine[2], the Lancet[3], Annals of Internal Medicine[4], BMJ, and thousands more[5] require authors to disclose whether and how they plan to share deidentified raw data[6] from individual participants in their clinical trials.

What's more, researchers wishing to publish in these journals must declare their data-sharing plans in a public registry[7], such as ClinicalTrials.gov[8].

It's a radical departure from where we've been. In my former life conducting trials as a scientist in industry and for the National Institutes of Health, when I'd log onto ClinicalTrials.gov to register a new trial, I didn't have to give a second thought to if or how I'd be sharing data from the trial. Now all researchers need to think about that from the very beginning, even before the first trial participant is enrolled.

To be clear, these new journal requirements do not *require* authors to share their data, so even with this new policy researchers might be tempted not to do so. That's understandable. Clinical trials are becoming increasingly complex, costly, and time-consuming. Competition among researchers, whether they are in the biopharma industry or in academia, is fiercer than ever.

Related: [9]

Journal editor calls for 'culture change' around clinical trial data [9]

But there are rewards for sharing data. By building upon each other's work, we can move faster toward insights for diseases such as asthma, cancer, dementia, diabetes, heart disease, and more. That's good for science and patients.

There are also rewards for the researchers who share data. They can be more widely recognized for their hard work. Those who share data receive more citations[10] from other researchers. It's a reward system that builds on itself: As the number of papers that reuse a data set grows, the more citations a researcher receives. This takes on even more significance with the International Committee of Medical Journal Editors requirements, as editors have indicated[11] they "may take into consideration data sharing statements when making editorial decisions." What's more, a growing number of funders — from government to philanthropy — require grantees to share data. With more grants and more publications can come the golden prize of tenure for academics.

Even more important, sharing clinical trial[94] data honors the people who volunteer for them. They put themselves at risk, give up precious time — sometimes years — and must endure multiple medical exams, blood draws, scans, and more.

All too often I am met with surprise when I tell patient groups that clinical trials are designed to answer just a single fundamental question. They rightfully expect that their data will live on and be used to help solve future problems. Now this expectation could become reality.

Over the last few years, the culture in the pharmaceutical industry has begun to shift toward more transparent sharing[12] of data. Academic investigators[13] have also started to step forward to share their clinical trial data. Now, with the new journal requirements, these decisions are out there for the world to see, shared on a public register. I believe this public declaration will accelerate data sharing.

Related: [14]

Sharing is a cardinal virtue, but scientists still struggle with it [14]

As it becomes easier and more fruitful than ever before to share data, researchers are starting to awaken to this new reality. Several platforms are now available to help researchers share patient-level data. These platforms are tailored to address researchers' concerns — whether that's fear of losing credit for their work or losing control of their data — and resolve practical worries about how to actually share it. (Full disclosure: I work for a nonprofit institution that manages such a platform.)

I hope that 2019 is a landmark year for data sharing — a year when investigators and researchers who run clinical trials create data-sharing plans and make them public. This move to more collaborative science won't just accelerate medicine. It is an ethical imperative and our responsibility to patients worldwide.

Rebecca Li is executive director of Vivli[15], an independent, nonprofit organization that has developed a global data-sharing and analytics platform. She is also a faculty member at the Center for Bioethics[16] at the Harvard

Medical School and a senior advisor at the MRCT Center[17] of Brigham and Women's Hospital and Harvard.

About the Author

Rebecca Li

rli@vivli.org[18]
@VivliCenter[19]

Tags

Links

1. https://www.statnews.com/category/first-opinion/
2. https://www.nejm.org/
3. https://www.thelancet.com/
4. http://annals.org/aim
5. http://www.icmje.org/journals-following-the-icmje-recommendations/
6. http://www.icmje.org/news-and-editorials/data_sharing_june_2017.pdf
7. http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html
8. https://clinicaltrials.gov/
9. https://www.statnews.com/2017/04/04/clinical-trial-data-sharing/
10. https://peerj.com/articles/175/
11. https://annals.org/aim/fullarticle/2630766
12. http://phrma-docs.phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf
13. https://www.jacionline.org/article/S0091-6749(16)30601-7/fulltext
14. https://www.statnews.com/2017/03/10/data-sharing-scientists/
15. https://vivli.org/
16. http://bioethics.hms.harvard.edu/
17. https://mrctcenter.org/
18. https://www.statnews.com/2019/02/19/data-sharing-imperative-clinical-trials/mailto:rli@vivli.org
19. https://twitter.com/VivliCenter
20. https://www.statnews.com/tag/clinical-trials/
21. https://www.statnews.com/tag/research/

SCIENTIFIC DATA

OPEN

Overview and experience of the YODA Project with clinical trial data sharing after 5 years

Joseph S. Ross[1,2,3,4], Joanne Waldstreicher[5], Stephen Bamford[6], Jesse A. Berlin[5], Karla Childers[5], Nihar R. Desai[4,7], Ginger Gamble[4], Cary P. Gross[1,2,4,8], Richard Kuntz[9], Richard Lehman[10], Peter Lins[5], Sandra A. Morris[5], Jessica D. Ritchie[4] & Harlan M. Krumholz[2,3,4,7]

Received: 28 March 2018
Accepted: 24 October 2018
Published: 27 November 2018

The Yale University Open Data Access (YODA) Project has facilitated access to clinical trial data since 2013. The purpose of this article is to provide an overview of the Project, describe key decisions that were made when establishing data sharing policies, and suggest how our experience and the experiences of our first two data generator partners, Medtronic, Inc. and Johnson & Johnson, can be used to enhance other ongoing or future initiatives.

[1]Section of General Medicine, Department of Medicine, Yale School of Medicine, New Haven, CT, USA. [2]National Clinician Scholars Program, Yale School of Medicine, New Haven, CT, USA. [3]Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA. [4]Center for Outcomes Research and Evaluation, Yale–New Haven Hospital, New Haven, CT, USA. [5]Johnson & Johnson, New Brunswick, NJ, USA. [6]Janssen Pharmaceutical Companies of Johnson & Johnson, High Wycombe, UK. [7]Section of Cardiovascular Medicine, Department of Medicine, Yale School of Medicine, New Haven, CT, USA. [8]Cancer Outcomes, Public Policy, and Effectiveness Research (COPPER) Center and Yale Cancer Center, Yale School of Medicine, New Haven, CT, USA. [9]Medtronic, Inc., Minneapolis, MN, USA. [10]Stonebank Medical Ltd., Banbury, UK. Correspondence and requests for materials should be addressed to J.S.R. (email: joseph.ross@yale.edu)

Introduction

Remarkable quantities of participant-level data and aggregate-level results are generated through clinical research, much of which is never published or disseminated, limiting its contribution to current knowledge and practice. Over the past five years, many leaders within the clinical research enterprise have made bold statements and adopted policies that promote clinical research data sharing, defined by the National Academy of Medicine (formerly known as the Institute of Medicine of the U.S. National Academy of Sciences) as the distribution of individual participant-level clinical trial data to researchers based outside the original study investigator team to enable independent use for scientific purposes[1]. Examples of organizations making recommendations to facilitate data sharing include the National Academy of Medicine[1], the World Health Organization (http://www.who.int/medicines/ebola-treatment/data-sharing_phe/en/), the U.S. National Institutes of Health (NIH)[2], the European Clinical Research Infrastructure Network[3], and the International Committee for Medical Journal Editors[4], as well as both PhRMA and EFPIA (https://www.phrma.org/press-release/joint-efpia-phrma-principles-for-responsible-clinical-trial-data-sharing-become-effective-today), the pharmaceutical trade organizations representing manufacturers in the United States and European Union, respectively.

Given already widespread, and still increasing, support for data sharing[5–7], the challenge facing the field now is to develop fair and sustainable approaches for investigators to access these data and utilize them to advance science. The Yale University Open Data Access (YODA) Project, an initiative housed within Yale University, has been actively working to facilitate access to clinical trial data since 2013. The purpose of this article is to provide an overview of the Project, describe key decisions that were made when establishing data sharing policies, and suggest how our experience and the experiences of our first two data generator partners, Medtronic, Inc. and Johnson & Johnson, can be used to inform and thereby enhance other ongoing or future initiatives.

Results

Overview

In 2011, the YODA Project was founded to promote data sharing among the scientific community and develop a platform that could be used as a means of responsible data sharing[8]. At its inception, the Project established an organizing mission to guide its decision-making: 1) promote the sharing of clinical research data to advance science and improve public health and healthcare; 2) protect the rights of research participants; 3) promote the responsible conduct of research; and 4) ensure good stewardship of clinical research data by external investigators.

The YODA Project began as part of a partnership with Medtronic, Inc. in 2011 with two purposes[9,10]. First, the Project was tasked with soliciting two independent analyses of individual participant-level data (IPD) from all published and unpublished trials relating to its marketed product recombinant human bone morphogenetic protein-2 (rhBMP-2). Second, following completion of these analyses in 2013, the data were to be made available for sharing with the broader scientific community. This partnership (and the data sharing) concluded in 2015.

In 2014, the YODA Project began a partnership with Johnson & Johnson to develop and implement a broadly-encompassing policy to share clinical trial data for all non-Phase I interventional trials of the company's pharmaceutical products; data were first made available to external investigators through this initiative in October 2014[11]. The scope of this agreement was expanded in 2015 to include trials of medical device products and again in 2017 to include trials of consumer products used by health authorities for approval from 2014 onwards; the detailed policy scope is publicly available (http://yoda.yale.edu/policies-procedures-guide-external-investigator-access-clinical-trial-data).

In 2016, the YODA Project began a partnership with SI-BONE, Inc., a smaller medical device company, to share clinical trial data relating to its marketed product, the iFuse sacroiliac joint fusion implant system.

Policy Development Process

The YODA Project data sharing policy was initially established for access to Medtronic's rhBMP-2 clinical trial data, and later for access to Johnson & Johnson's broad portfolio of clinical trial data. The policy established the procedures external investigators were required to follow to gain access to data for independent scientific examination (see Fig. 1 for illustration). Development was iterative and informed by the following:

- Input from partnering companies that have generated the clinical trial data;
- Input from the YODA Project's Steering Committee, an independent group of leaders in the fields of clinical research and biomedical ethics assembled by the YODA Project to provide guidance;
- Input from other experts in the field, industry, regulators, and the general public through a public comment period and personal communication;
- Review of the literature and policies from other organizations engaged in clinical trial data sharing, such as the NIH's Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) within the U.S. National Heart, Lung, and Blood Institute[12,13]; and
- The experience gained by the YODA Project sharing IPD with external investigators.

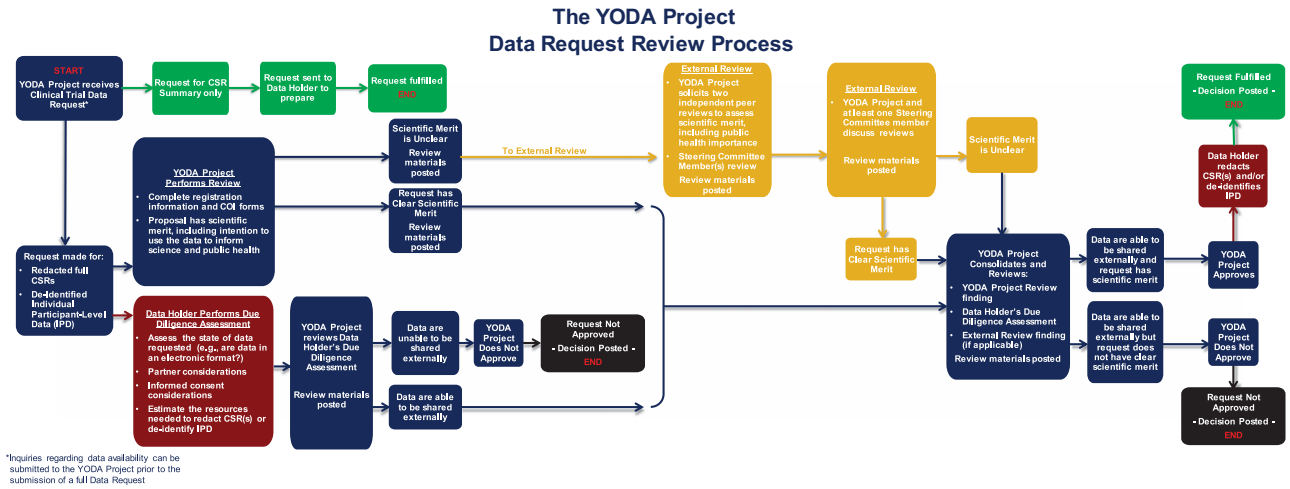


Figure 1. **The YODA Project data request review process.**

Key Aspects of the YODA Project Data Sharing Policy

There are several key aspects that were considered as decisions were made when establishing the Project's data sharing policy. These decisions, and the lessons we learned from their implementation, can be used to inform and enhance other ongoing or future data sharing initiatives.

Commitment to Transparency. Transparency enhances trust in the integrity of the data sharing process and the resulting research, as well as clarity of parties involved. To ensure transparency of the overall effort, the YODA Project makes as much information publicly accessible as possible, including Project personnel; requirements for data access; information on trial data available, such as Clinical Study Report summaries and hyperlinks to registration records on ClinicalTrials.gov and publications on PubMed.gov; information on all submitted proposals for data, such as the number approved and rejected; and approved research proposals in their entirety, including the resulting research upon completion. In addition, all major decisions made by the YODA Project have included an opportunity for public comment, including the finalized policies and procedures for making clinical trial data available.

Full Authority and Independence. Data sharing requires mutual trust and collaboration with each partnering company, while at the same time upholding the independence and impartiality of the data-sharing organization. The YODA Project requires that it has full decision-making authority over the release of the data and serves as an independent intermediary to manage requests and promote data use. The final decisions regarding the design of the data request process, the criteria for access, and approval or rejection of requests all reside with the YODA Project. Maintaining this final authority is intended to build trust in the process and reduces opportunities for real or perceived influence.

Independent Steering Committee. The YODA Project assembled a Steering Committee of external experts, an independent group of leaders in the fields of clinical research and biomedical ethics, to provide guidance as it developed standards, policies, and procedures. The YODA Project was able to call on the Committee's expertise to inform the data release process, including how best to make trial information available, what data request requirements should be established, and what the nature of the data request review process should be. The Committee also provided valuable feedback on other issues, such as the importance of making meta-data available (i.e., information about the trials being shared), including statistical analysis plans, blank case report forms, and study protocols, along with more difficult issues such as what to do when meta-data materials had not been prepared in English. In addition, the Committee assisted in making sure that the procedures were consistent with the standards of ethical research, including avoidance of conflict of interest and protection of patient privacy.

List of Available Trials. Publicly listing trials that are available to external investigators is crucial both to promoting use of the shared data and establishing the transparency of the initiative. However, the proactive preparation of a catalogue of available trials for a large company with a multitude of marketed products is time and resource intensive. Therefore, to ensure the efficient use of resources, at the initiation of our data sharing partnerships, the criteria for defining trials in-scope for sharing were established. Pertinent to this decision is the protection of patient privacy, as data for which the privacy and confidentiality of research participants cannot be protected should not be routinely shared, an important consideration for studies of rare diseases or those that have few participants. Furthermore, for any product, determinations need to be made as to whether the data are owned solely by the company, as

Box 1 | Supporting documentation for each clinical trial made publicly available on the YODA Project webpages.

- Study Title
- Sponsor Protocol Number
- Product Name
- Generic Name
- Therapeutic Area
- Product Class
- Condition(s) Studied
- ClinicalTrials.gov NCT Number and Link to Record
- Link to PubMed Primary Publication Record
- Study Phase
- Enrollment
- Mean/Median Age
- % Female
- % White
- Study Synopsis Link
- Availability (yes/no) of annotated case report forms, data definition specifications, protocol with amendments, analysis datasets, statistical analysis plans, and full clinical study reports

medical products are frequently jointly owned or marketed, requiring consent from both manufacturers before the data could be shared.

For the partnership with Johnson & Johnson, the number of trials in-scope was large, particularly because of the company's commitment to make older trials available. Thus, at the outset, a subset of contemporary trials for commonly used products likely to be of interest to medical researchers were identified and publicly listed. Because this list was not exhaustive of all trials that could be made available, the Project also established a method for external investigators to inquire about the potential availability of other trials that were not listed. As additional trials were requested and made available, they were added to the public listing. Further, to facilitate identification of trials by investigators, the listing was organized into multiple views, including by product, therapeutic area, and condition studied, and also made searchable on key data elements, such as enrollment and trial demographic characteristics.

Supporting Documentation. In order to increase the likelihood that the data can be used for research, interested parties need full understanding of the data resources being made available. To this end, supporting documentation and materials, or meta-data, are described for each available trial, including hyperlinks to the ClinicalTrials.gov registration (or a number from another international registry) and known trial publications through PubMed.gov. Similarly, documentation, such as blank case report forms, clinical study reports, data definition specifications, protocols with amendments, and/or statistical analysis plans, enables investigators to more efficiently and accurately determine for what purpose the data can be best used if access is obtained. Supporting documentation materials made publicly available are listed in Box 1.

Research Proposal Submission and Public Posting. It is essential to demonstrate that research by external investigators making use of data made available by industry is responsibly conducted, since concerns continue to be voiced about the potential for its misuse and misinterpretation[6]. To promote the responsible conduct of research, the YODA Project adopted a controlled access model[14], requiring investigator registration and submission of a proposal to be reviewed prior to approval, which can then be subsequently publicly posted once data access is granted. Specific information must be submitted, including the principal investigator, key project personnel, and the research proposal; requirements are listed in Box 2. Notably, a statistician is not required to be included among key project personnel. Requiring registration and public posting of proposals potentially fosters collaboration and open science, while also making it easier for interested independent scientists to evaluate research using shared data.

Blinded Request Review by the YODA Project. All data requests undergo review by multiple clinical investigator members of the YODA Project, blinded to all identifying details about the investigator, including funding source. Review helps to ensure that the proposal has scientific merit, in that 1) the scientific purpose is clearly described; 2) the data requested will be used to create or materially enhance generalizable scientific and/or medical knowledge to inform science and public health; and 3) the proposed research can be reasonably addressed using the requested data. The assessment of whether the proposed research can be reasonably addressed using the requested data includes evaluating whether the variables needed for the proposed analysis are included in the requested data and whether the question could best be addressed with either individual participant-level or summary-level data. While the YODA Project review is not a detailed technical evaluation of the proposed research *per se*, a high-level review evaluates whether the proposed statistical methods can answer the scientific question. The review process

Box 2 | Information required to be submitted as part of any routine data request.

- o Investigator name, affiliation, key personnel
- o Narrative summary
- o Public abstract
- o Research proposal, which includes a clear description of the project background, research objectives, and proposed methods, such as the study design, pre-specification of the sample inclusion/exclusion criteria, main outcome measures and statistical analysis plan
- o Conflict of interest statement
- o Timeline and dissemination plan

permits peer-review feedback and/or requests for clarification prior to approval determination. All YODA Project reviews are publicly posted.

There have been instances where the YODA Project has also provided feasibility feedback to facilitate the use of data. For instance, one request was received for a methodological study characterizing trial populations' representativeness that would use all trials listed on the YODA Project website (there were 123 at the time); it was determined that preparation of these data would take upwards of 6 months. The YODA Project communicated this to the investigator, who decided to narrow their proposal to use only those trials that were already de-identified and prepared for sharing and planned to subsequently amend their request as more data became available.

Blinded Request Review by Partnering Company. All data requests undergo a due diligence assessment by the partnering company, a blinded assessment of whether the data are already prepared or need to be prepared in a format that is de-identified and can be made available to external researchers, as well as whether the variables of interest are available and to check whether a similar analysis is underway or has already been completed by the company; if so, this information is shared with the investigator. For example, the YODA Project received a data request that proposed characterizing reasons for trial eligibility screening failures in late phase trials in advanced genitourinary cancers. After examining the data, the partnering company advised the YODA Project that reasons for screening failure were not documented for all occurrences. After relaying this information to the investigator, they still chose to proceed with their proposal. All company due diligence assessments are publicly posted.

Opportunity for Collaboration with Partnering Company. Data sharing initiatives create a means for investigators to conduct independent analyses, but they also provide opportunities for new collaborations.[4,15] To foster these collaborations, the YODA Project established a process that allows for and coordinates communication between investigators and the partnering company, when desired by the investigator. This process allows external investigators to be made aware of any similar ongoing research efforts and potentially foster collaboration when there is mutual interest.

Data Use Agreement. All approved data requests require a signed Data Use Agreement (DUA) between Yale University, representing the YODA Project, and the affiliated institution representing the external investigator. The DUA is intended to ensure that the investigator will protect the confidentiality of the data, will not attempt to re-identify trial participants, and will not copy, retransmit, or use the data in any manner other than for the purpose described in the data request. A template is publicly available for investigators to review prior to requesting data (http://yoda.yale.edu/data-use-agreement). The DUA limits data access to one year, ensuring frequent updates and contact between the investigators and the YODA Project, as the agreement can be re-approved for additional time if the research is ongoing. The DUA also specifies that data cannot be used for non-scientific purposes, such as in pursuit of litigation or for commercial interests. While there remains ambiguity in what constitutes non-scientific use, and commercial use in particular, the agreement is intended to promote a good faith approach among researchers. Moreover, the DUA requires investigators to report any notable safety results to the partnering company, as it is the legal responsibility of any manufacturer of all U.S. Food and Drug Administration regulated medical products to report these findings to appropriate regulatory officials. Finally, to ensure that investigators understand the DUA, the YODA Project developed a training module that must be completed prior to submission of an investigator's first data request, emphasizing important points and policies governing access.

Secure Data Access or Transfer. The YODA Project has employed two different models of controlled data access for approved investigators, adopted to ensure the security and prevent wider distribution of shared clinical trial data. In our partnership with Medtronic, password-protected, de-identified data were distributed using secure file transferring services to approved investigators. Certificates of data destruction were required at DUA expiration. This method is not costly and offers greater flexibility to

investigators and enables use of additional software types (so long as the investigator has his/her own license) and linkage to other data sources, but increases the risk of unapproved distribution.

In the YODA Project partnership with Johnson & Johnson, the company entered into licensing agreements with a secure data sharing platform allowing virtual data access to credentialed and approved investigators while precluding data download or distribution. While secure, these platforms are expensive and limited to the licensed analytic tools and software. Moreover, there have been challenges to uploading complementary or other trial data onto the licensed platform, preventing the combination of data from different sources, such as for meta-analyses, so that investigators had to rely on the aggregate-level data. Lastly, users of the platform have also reported it to be difficult to navigate, time intensive to learn, and noted that the programs occasionally closed, causing loss of work and the user to be logged out. Nevertheless, on the whole, these platforms have offered a reasonable technical means for data security and protection against redistribution.

There are certain circumstances that may require access to clinical trial data outside of the secure platform, such as the need to use proprietary software that cannot be placed in the platform or to pool data from other sources that cannot access the platform. However, given the sensitivity of patients' trial data, including the importance of protecting privacy and risk of re-identification, as well as the greater risk for unauthorized data dissemination and analysis, access outside of the secure platform requires clear justification. In collaboration with Johnson & Johnson, the YODA Project developed an exception request process, in which the investigator is required to provide additional information to support his/her request for direct access to the data. This includes 1) strong rationale for why the requested data can uniquely be used to address the proposed project aims, and reasons why other clinical research data are not available or cannot be used; 2) reasons why the platform may not permit the proposed analyses; and 3) a description of the protections in place to ensure data security outside of the platform, including technological and related procedural safeguards. Based on this information, the YODA Project then assesses the need for data access outside of the secure platform. Not all exception requests are approved.

Results Dissemination. Upon project completion, the YODA Project requires public dissemination of research findings, preferentially through the peer-reviewed biomedical literature or at scientific meetings. This explicitly promotes the scientific process and peer review, ensuring the methods used meet the minimal scientific standards prior to dissemination. Once the proposed research has been publicly disseminated through peer-reviewed publication, these findings can be further discussed via non-peer-reviewed forums, including internet posts, newspaper articles, or other means. Regardless of whether the work is published, all results from analyses are required to be reported back to the YODA Project in summary form to be publicly posted at the expiration of the DUA, including whether or not the project was ultimately completed, ensuring public transparency and accountability.

Prior to publication or presentation at a scientific meeting, copies of any abstract or manuscript generated from the data request are required to be shared with the YODA Project. This helps to track whether projects are progressing, clarifies what new scientific information has been generated, and updates our records of the completion and publication of analyses, informing future data sharing efforts.

Data Access Fee. Thus far, the YODA Project and its data partners have not imposed a fee for investigators to access data through the YODA Project. The entire cost has been covered by the industry partners. However, the sustainability of data sharing initiatives, and their current reliance on industry funding, will likely require reconsideration of funding models, especially given the hope that data sharing efforts will broaden to include smaller pharmaceutical, medical device, and biotech companies, as well as academic institutions. A solution may be for investigators to apply for research grants from government agencies or non-profit organizations to support the use of shared data. Similarly, funding to prepare data for external sharing should be built into grants awarded to academic groups when they conduct their own clinical trials. In the current environment, in which funding is already very competitive, whether such funding models would work is uncertain.

Medtronic Experience

As noted above, Medtronic's rhBMP-2 clinical trial data were shared once the two independent systematic reviews and meta-analyses of the data had been published, which occurred in June 2013[16–18]. Concurrently, an online application process was established and a total of 4 requests for the rhBMP-2 trial data were received. All 4 were approved (additional details, including the protocols, can be found at http://yoda.yale.edu/medtronic-past-data-recipients), 2 of which were completed and resulted in peer-reviewed publications[19,20]. While the rhBMP-2 trial data were generated by a single company, individual trials did not adhere to the same data formatting and standards, requiring Medtronic to first invest resources into data preparation for sharing, as well as for investigators to review data files and recode data as needed to allow aggregation and meta-analysis. The remaining 2 were not completed, both due to investigator commitments to other projects and lack of time. No instances of data redistribution were reported and certificates of data destruction were received for the 2 uncompleted projects; the 2 completed projects retain access to the data for 5 years in the event that questions are raised about their published analyses, after which time the DUA expires. Because no requests for the data were received after January 2014, Medtronic discontinued making the data available in summer 2015.

Total inquiries, No.	161
Total inquiries answered to date, No. (%)	159 (98.8%)
Inquiry led to full data request, No. (%)	31 (19.3%)
Median number of days for response to inquiry (Interquartile Range)	15 (7.5–41.5)
Total unique trials requested within answered inquiries, No.	207
Trial data can be made "available" to request, No. (%)	124 (59.9%)
Trial data cannot be made "available" to request, No. (%)	83 (40.1%)
Regulatory approval not yet received, No. (%)	17 (20.5%)
Trial ongoing or completed < 18 months ago, No. (%)	26 (31.3%)
Data cannot be adequately de-identified, No. (%)	0 (0%)
Partner of Data Holder has not agreed to share, No. (%)	11 (13.3%)
Trial is out of scope (i.e., Phase 1, OTC, etc.), No. (%)	25 (30.1%)
Data subject to partner agreement; researcher advised to contact partnering Data Holder, No. (%)	2 (2.4%)
Data cannot be converted to electronic format, No. (%)	1 (1.2%)
Trial materials not available in English, No. (%)	5 (6.0%)

Table 1. Details of YODA Project inquiry process for Johnson & Johnson clinical trials as of August 27, 2018.

Trials Available, No.	270
Products Available, No.	31
Trial Enrollment Size	
Mean	412.7
Median	322
Min	5
Max	2051
Sex, No. (%)	
> 50% Female	101 (37.4%)
≤ 50% Female	131 (48.5%)
[Unknown Sex]	38 (14.1%)
Race, No. (%)	
> 50% White	145 (53.7%)
≤ 50% White	26 (9.6%)
[Unknown Race]	99 (36.7%)
Mean/Median Enrollment Age, No. (%)	
0–19	24 (8.9%)
20–39	73 (27.0%)
40–59	100 (37.0%)
60+	32 (11.9%)
[Unknown Age]	41 (15.2%)
Available Data and Documentation, No. (%)	
Collected datasets	246 (91.1%)
Analysis datasets	5 (1.9%)
[No participant-level data]	19 (7.0%)
Clinical study report (CSR)	252 (93.3%)
Protocol with amendments	256 (94.8%)
Statistical analysis plan	243 (90.0%)
Annotated case report form	224 (83.0%)
Data definition specification	194 (71.9%)
CSR summary available on site	187 (69.3%)
CSR summary not yet prepared	66 (24.4%)

Table 2. Details of Johnson & Johnson clinical trials available to request as of August 27, 2018.

Behaviors and Mental Disorders, No. (%)	106 (39.3)
Muscle, Bone, and Cartilage Diseases, No. (%)	26 (9.6)
Digestive System Diseases, No. (%)	23 (8.5)
Cancers and Other Neoplasms, No. (%)	19 (7.0)
Nutritional and Metabolic Diseases, No. (%)	18 (6.7)
Skin and Connective Tissue Diseases, No. (%)	17 (6.3)
Viral Diseases, No. (%)	14 (5.2)
Blood and Lymph Conditions, No. (%)	13 (4.8)
Nervous System Diseases, No. (%)	12 (4.4)
Immune System Diseases, No. (%)	7 (2.6)
Mouth and Tooth Diseases, No. (%)	4 (1.5)
Urinary Tract, Sexual Organs, and Pregnancy Conditions, No. (%)	4 (1.5)
Respiratory Tract (Lung and Bronchial) Diseases, No. (%)	3 (1.1)
Parasitic Diseases, No. (%)	2 (0.7)
Heart and Blood Diseases, No. (%)	1 (0.4)
Neurosciences, No. (%)	1 (0.4)

Table 3. Therapeutic areas of Johnson & Johnson clinical trials available to request as of August 27, 2018.

Johnson & Johnson Experience

Johnson & Johnson began sharing clinical trial data for all trials of the company's pharmaceutical products in October 2014 and later expanded its scope to include trials of medical device and consumer products. While the company is willing to make all non-Phase I interventional pharmaceutical trials available, at the initiation of the partnership, we proactively identified contemporaneous trials likely to be of greatest interest to the scientific community and listed them on the YODA Project website. At the same time, the inquiry process was critical during the early days of the partnership; as of August 2018, 161 inquiries for more than 200 unique trials have been submitted, identifying 124 that could be made available for sharing (Table 1). Most common reasons for unavailability include that the trial was out-of-scope (i.e., phase 1 healthy volunteer studies), ongoing or completed less than 18 months ago, or that regulatory approval had not yet been received (details are available at: http://yoda.yale.edu/request/summary-data-inquiries-and-requests/details-inquiries-submitted-data-not-yet-available). However, it should be noted that when investigators inquire about the availability of an out-of-scope trial, they are invited to submit an abstract that conveys the scientific importance of the planned research, which is then evaluated by the YODA Project on a case by case basis; thus far, only 1 investigator has followed up with an abstract. As of August 2018, 270 clinical trials are listed on the YODA Project website, along with supporting documentation (Table 2), and additional details, including specific therapeutic areas (Table 3), specific conditions studied (Table 4 (available online only)), specific products (Table 5), and specific product classes (Table 6). Johnson & Johnson started to adopt Clinical Data Interchange Standards Consortium (CDISC, https://www.cdisc.org/) in 2001 and by 2003 study data sets were being routinely reported in CDISC format. By using CDISC standards, it brings benefits in enabling the high reuse of software for data analysis. It also minimizes the time and effort of data preparation before data sharing can commence. For the researcher, it increases familiarity of data structures and helps to ensure consistency between trials (and even sponsors).

As of August 2018, 100 data requests have been received from 89 unique principal investigators for a median of 3 trials per request (Interquartile Range, 1-9), 90 (90.0%) of which have been approved with a DUA signed or in progress, 2 (2.0%) remain under review pending revision, and 8 (8.0%) were withdrawn or closed (Table 7). Withdrawals generally occurred because the due diligence assessment determined that the data needed to address the proposed question were not available (such as requests for pharmacokinetic data or endoscopic video data) or because special statistical software was needed that could not be imported into the secure data sharing platform. Notably, two withdrawn requests resulted in direct collaboration with Johnson & Johnson to pursue the research. No request has been rejected, although 36 (36.0%) of submitted research proposals required revision after YODA Project review for clarification or elaboration, 2 of which were never resubmitted and are now considered withdrawn.

Among the 270 clinical trials currently listed on the YODA Project website, 183(67.8%) have thus far been requested, although 46 have only been available since January 1, 2018. The most common purposes of the proposed projects include (not mutually exclusive) addressing secondary research questions (n = 57; 57.0%), combining data as part of larger meta-analysis (n = 45; 45.0%), and validating previously published studies (n = 24; 24.0%).

RISPERDAL®, No. (%)	25 (9.3)
INVEGA®, No. (%)	23 (8.5)
TOPAMAX®, No. (%)	22 (8.1)
REMICADE®, No. (%)	21 (7.8)
SIMPONI®, No. (%)	21 (7.8)
RAZADYNE®, No. (%)	20 (7.4)
STELARA®, No. (%)	19 (7.0)
INVOKANA®, No. (%)	13 (4.8)
PROCRIT®, No. (%)	13 (4.8)
CONCERTA®, No. (%)	12 (4.4)
INVEGA SUSTENNA®, No. (%)	11 (4.1)
OLYSIO®, No. (%)	11 (4.1)
RISPERDAL CONSTA®, No. (%)	10 (3.7)
YONDELIS®, No. (%)	10 (3.7)
DARZALEX®, No. (%)	5 (1.9)
Other, No. (%)	5 (1.9)
Mouth Rinse, potassium oxalate 1.4%, No. (%)	4 (1.5)
MONONESSA ® ORTHO-CYCLEN ® ORTHO TRI-CYCLEN ® TRINESSA ®, No. (%)	3 (1.1)
PREZISTA®, No. (%)	3 (1.1)
DOXIL®, No. (%)	2 (0.7)
EDURANT®, No. (%)	2 (0.7)
PLIVENSIA™, No. (%)	2 (0.7)
Rogaine 5% Women's Foam, No. (%)	2 (0.7)
SIRTURO®, No. (%)	2 (0.7)
VERMOX®, No. (%)	2 (0.7)
ZYTIGA®, No. (%)	2 (0.7)
INTELENCE®, No. (%)	1 (0.4)
LEVAQUIN®, No. (%)	1 (0.4)
TERAZOL ®, No. (%)	1 (0.4)
THERMOCOOL® SMARTTOUCH™ Catheter, No. (%)	1 (0.4)
TREMFYA®, No. (%)	1 (0.4)

Table 5. Product names of Johnson & Johnson clinical trials available to request as of August 27, 2018.

Of 82 approved requests provided data access, most remain in progress; 11(13.4%) have at least one publication[21–31] (12 publications in total, see Table 8 for publications list) and 7 (8.5%) have at least one article under peer-review, whereas 19 (23.2%) have been discontinued, either because of investigator commitments to other projects and lack of time or because the investigator did not have sufficient statistical expertise to conduct analyses within the secure data sharing platform. All of the projects that have been submitted for publication described analyses representing those specified in the original research proposal. Finally, no instances of data redistribution have been reported.

Discussion

Data sharing and data transparency are quickly becoming the new standard in pharmaceutical and medical device science and in clinical research more broadly. Many national and international organizations are adopting policies to advance scientific and medical knowledge through data availability and transparency that will ultimately improve public health and healthcare delivery, advancing scientific understanding of disease diagnosis and prognosis through the development of novel tools and approaches, while also improving existing knowledge of treatment safety and efficacy.

The early experiences of the YODA Project can be used to inform the field and other data sharing initiatives. Certain decisions contrast with other existing clinical trial data sharing initiatives (including those found at https://biolincc.nhlbi.nih.gov/studies/, https://ClinicalStudyDataRequest.com/, and https://dcri.org, as examples). For instance, the decision to *reactively* de-identify data for sharing in response to requests has meant that 68% of trials listed on the YODA Project website have been used, as opposed to platforms that have *proactively* de-identified data for sharing reporting that approximately 15% of listed

Atypical Antipsychotics, No. (%)	69 (25.6)
Antirheumatic Agents - Biologic Response Modifiers, No. (%)	50 (18.5)
Anticonvulsants, No. (%)	22 (8.1)
Alzheimer's Disease - Cholinesterase Inhibitors, No. (%)	20 (7.4)
Antiviral Agents, No. (%)	17 (6.3)
Antipsoriatics, No. (%)	14 (5.2)
Stimulants/ADHD/Anorexiants, No. (%)	12 (4.4)
Antineoplastic Agents, No. (%)	10 (3.7)
Colony-Stimulating Factors, No. (%)	9 (3.3)
Diabetes Related- Other, No. (%)	9 (3.3)
Monoclonal Antibody, No. (%)	5 (1.9)
Hematologic Agents, No. (%)	4 (1.5)
Mouth Rinse Device, No. (%)	4 (1.5)
Sodium-Glucose Co-Transporter 2 (SGLT2) Inhibitor, No. (%)	4 (1.5)
Immunizations, No. (%)	3 (1.1)
OB/GYN, No. (%)	3 (1.1)
Antimycobacterial Agents, No. (%)	2 (0.7)
Antiparasitics, No. (%)	2 (0.7)
Hormones, No. (%)	2 (0.7)
Oncology - Antibiotic, No. (%)	2 (0.7)
Other, No. (%)	2 (0.7)
Skin & Mucous Membrane Agents, Miscellaneous, No. (%)	2 (0.7)
Cardiovascular Devices, No. (%)	1 (0.4)
Dermatology, No. (%)	1 (0.4)
Quinolones - 3rd gen., No. (%)	1 (0.4)

Table 6. **Product class of Johnson & Johnson clinical trials available to request as of August 27, 2018.**

Trials Available, No.	270
Trials Shared as Part of Approved Requests, No. (%)	183 (67.8)
Complete Data Requests Received, No.	100
Requests Requiring Revision During Review, No. (%)	36 (36.0)
Purpose of Proposed Research	
New research on treatment effectiveness or safety, No. (%)	57 (57.0)
Meta-analysis, No. (%)	45 (45.0)
Validating previous research on treatment effectiveness or safety, No. (%)	24 (24.0)
Research on clinical prediction or risk prediction, No. (%)	20 (20.0)
Develop or refine statistical methods, No. (%)	13 (13.0)
Research on clinical trial methods, No. (%)	12 (12.0)
Preliminary research for a grant proposal, No. (%)	10 (10.0)
Research on comparison group, No. (%)	6 (6.0)
Data Requests Approved, No. (%)	90 (90.0)
Data Requests Under Review, No. (%)	2 (2.0)
Data Requests Withdrawn/Closed, No. (%)	8 (8.0)
Requested data could not be used to address research question, No.	2
Data could not be downloaded as requested by investigator, No.	3
Investigator did not respond to YODA Project request for additional clarification, No.	2
Investigator withdrew approved request prior to signing DUA due to lack of resources, No.	1
Requests with Data Access [DUA signed by both parties], No. (%)	82 (82.0)
Requests with Publications, No.	11

Table 7. **Details of data requests received for Johnson & Johnson clinical trials as of August 27, 2018.**

First Author	Publication Title	Journal	Year	Publication ID	Cited by:
Fu, R	Effectiveness and harms of recombinant human bone morphogenetic protein-2 in spine fusion: a systematic review and meta-analysis.	*Ann Intern Med*	2013	doi:10.7326/0003-4819-158-12-201306180-00006	299
Simmonds, MC	Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual-participant data.	*Ann Intern Med*	2013	doi:10.7326/0003-4819-158-12-201306180-00005	233
Laurie, AL	Meta-analysis of the Impact of Patient Characteristics on Estimates of Effectiveness and Harms of Recombinant Human Bone Morphogenetic Protein-2 in Lumbar Spinal Fusion.	*Spine*	2016	doi:10.1097/BRS.0000000000001580	3
Noshchenko, A	What Is the Clinical Relevance of Radiographic Nonunion After Single-Level Lumbar Interbody Arthrodesis in Degenerative Disc Disease? A Meta-Analysis of the YODA Project Database.	*Spine*	2016	doi:10.1097/BRS.0000000000001113	5
Mospan, GA	5-Day versus 10-Day Course of Fluoroquinolones in Outpatient Males with a Urinary Tract Infection (UTI).	*J Am Board Fam Med*	2016	doi:10.3122/jabfm.2016.06.160065	4
Storgaard, H	Benefits and Harms of Sodium-Glucose Co-Transporter 2 Inhibitors in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis.	*PLoS One*	2016	doi:10.1371/journal.pone.0166125	37
Gay, HC	Feasibility, Process, and Outcomes of Cardiovascular Clinical Trial Data Sharing: A Reproduction Analysis of the SMART-AF Trial.	*JAMA Cardiol*	2017	doi:10.1001/jamacardio.2017.3808	6
Corbett, M	Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation.	*Health Technol Assess*	2017	doi:10.3310/hta21560	4
Mbuagbaw, L	Review of available evidence on the use of bedaquiline for the treatment of multidrug-resistant tuberculosis: Data analysis report; Appendix to A 2016 review of available evidence on the use of bedaquiline in the treatment of multidrug-resistant tuberculosis.	World Health Organization	2017	Report No. WHO/HTM/TB/2017.01	2
Wang, R	Comparative Efficacy of Tumor Necrosis Factor-alpha Inhibitors in Ankylosing Spondylitis: A Systematic Review and Bayesian Network Metaanalysis.	*J Rheumatol*	2018	doi:10.3899/jrheum.170224	1
Schneider-Thoma J	Second-generation antipsychotic drugs and short-term mortality: a systematic review and meta-analysis of placebo-controlled randomised controlled trials.	*Lancet Psychiatry*	2018	doi: 10.1016/S2215-0366(18)30177-9	1
Singh, S	Impact of Obesity on Short- and Intermediate-Term Outcomes in Inflammatory Bowel Diseases: Pooled Analysis of Placebo Arms of Infliximab Clinical Trials.	*Inflamm Bowel Dis*	2018	doi:10.1093/ibd/izy135	
Singh, S	No Benefit of Concomitant 5-Aminosalicylates in Patients With Ulcerative Colitis Escalated to Biologic Therapy: Pooled Analysis of Individual Participant Data From Clinical Trials.	*Am J Gastroenterol*	2018	doi:10.1038/s41395-018-0144-2	
Singh, S	Obesity and Response to Infliximab in Patients with Inflammatory Bowel Diseases: Pooled Analysis of Individual Participant Data from Clinical Trials.	*Am J Gastroenterol*	2018	doi:10.1038/s41395-018-0104-x	
Zou, X	The role of PANSS symptoms and adverse events in explaining the effects of paliperidone on social functioning: a causal mediation analysis approach.	*NPJ Schizophrenia*	2018	doi:10.1038/s41537-018-0054-8	
Spertus, J	Risk of weight gain for specific antipsychotic drugs: a meta-analysis.	*NPJ Schizophrenia*	2018	doi:10.1038/s41537-018-0053-9	

Table 8. Publications using data made available through the YODA Project.

trials having been used[32,33]. Similarly, while the direct, secure data transfer used for Medtronic rhBMP-2 trials was simple, the secure platform used for Johnson & Johnson's trials is more challenging for investigators, particularly those without advanced statistical expertise. Nevertheless, investigator experience with this platform mirrors the experience of those using data shared through the U.S. National Heart, Lung, and Blood Institute BioLINCC repository[34]. Lastly, Johnson & Johnson makes all non-Phase I interventional pharmaceutical trials available, including older trials, whereas many current initiatives and company policies are focused on sharing clinical trial data as of a specific date going forward, limiting the availability of trials that examined medical products commonly being used by patients today.

Outstanding issues remain for the field to address, including how to make older trial data available in a contemporary technology format for use today. Further, a sustainable model that covers the cost of data sharing is needed, as efforts are currently being paid exclusively by industry and, in some instances, the U. S. federal government. In addition, while much of the focus on data sharing has thus far been on industry[35,36], many other entities, particularly academia, also generate clinical research data. Lastly, there is a need going forward for systematic adoption of data format standards[37], expectations for how long shared data will be made available, along with informed consent language, to facilitate data sharing. Publicly-available informed consent templates that explicitly allow for the sharing of data with external researchers are already available (http://mrctcenter.org/projects/informed-consent/).

The goal of data sharing initiatives should be to ensure that the data are used to conduct high-quality and rigorous research that honors the voluntary efforts of patients that participated in the trials and serves the best interests of science and public health. The research community and society are likely to greatly benefit from these secondary research efforts. With the continuous advancement of data sharing efforts, the YODA Project's experience and the experiences of its first two data generator partners, Medtronic, Inc. and Johnson & Johnson, can be used to enhance other ongoing or future initiatives.

Methods

We provide an overview of the history of the YODA Project, including a review of the policy and procedures iteratively developed to guide granting qualified public access to clinical trial data provided by partnering data generators. We base the review on the experience with the first two partners in the Project, Medtronic, Inc. and Johnson & Johnson. This policy and set of procedures address the research proposal requirements, data receipt, data analysis, and dissemination of results (http://yoda.yale.edu/policies-procedures-guide-external-investigator-access-clinical-trial-data). Specifically, the policy guides the procedures that are used to make clinical trial data (including both Clinical Study Reports [CSRs] and participant-level trial data) available to external investigators for independent scientific examination. Key aspects of the policy and the underlying decisions were informed by the following:

- The YODA Project's core principles of fairness and transparency;
- The YODA Project's review of the literature and policies from other organizations engaged in clinical trial data sharing;
- Recommendations from the YODA Project Steering Committee, an independent group of leaders in the fields of clinical research and biomedical ethics assembled by the YODA Project to provide guidance;
- Recommendations from other experts in the field, general public, and industry partners through a public comment period and personal communication; and
- The YODA Project's accumulated experience with sharing participant-level clinical trial data.

Where appropriate, descriptive statistics were used to characterize the status of data requests and approvals.

References

1. Institute of Medicine of the National Academies of Science. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks.* (National Academies Press, 2015).
2. Hudson, K. L. & Collins, F. S. Sharing and reporting the results of clinical trials. *JAMA.* **313,** 355–356 (2015).
3. Ohmann, C. *et al.* Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open.* **7,** e018647 (2017).
4. Taichman, D. B. *et al.* Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors. *Ann. Intern. Med.* **164,** 505–506 (2016).
5. Hopkins, C. *et al.* UK publicly funded Clinical Trials Units supported a controlled access approach to share individual participant data but highlighted concerns. *J. Clin. Epidemiol.* **70,** 17–25 (2016).
6. Rathi, V. *et al.* Sharing of clinical trial data among trialists: a cross sectional survey. *BMJ.* **345,** e7570 (2012).
7. Tudur Smith, C. *et al.* Sharing individual participant data from clinical trials: an opinion survey regarding the establishment of a central repository. *PLoS One.* **9,** e97886 (2014).
8. Krumholz, H. M. & Ross, J. S. A model for dissemination and independent analysis of industry data. *JAMA.* **306,** 1593–1594 (2011).
9. Krumholz, H. M. *et al.* A historic moment for open science: the Yale University Open Data Access Project and Medtronic. *Ann. Intern. Med.* **158,** 910–911 (2013).
10. Kuntz, R. E. The changing structure of industry-sponsored clinical research: pioneering data sharing and transparency. *Ann. Intern. Med.* **158,** 914–915 (2013).
11. Krumholz, H. M. & Waldstreicher, J. The Yale Open Data Access (YODA) Project--A Mechanism for Data Sharing. *N. Engl. J. Med.* **375,** 403–405 (2016).
12. Coady, S. A. & Wagner, E. Sharing individual level data from observational studies and clinical trials: a perspective from NHLBI. *Trials.* **14,** 201 (2013).
13. Giffen, C. A. *et al.* Providing Contemporary Access to Historical Biospecimen Collections: Development of the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC). *Biopreserv. Biobank.* **13,** 271–279 (2015).
14. Mello, M. M. *et al.* Preparing for responsible sharing of clinical trial data. *N. Engl. J. Med.* **369,** 1651–1658 (2013).
15. Longo, D. L. & Drazen, J. M. Data Sharing. *N. Engl. J. Med.* **374,** 276–277 (2016).
16. Laine, C. *et al.* Closing in on the truth about recombinant human bone morphogenetic protein-2: evidence synthesis, data sharing, peer review, and reproducible research. *Ann. Intern. Med.* **158,** 916–918 (2013).
17. Fu, R. *et al.* Effectiveness and harms of recombinant human bone morphogenetic protein-2 in spine fusion: a systematic review and meta-analysis. *Ann. Intern. Med.* **158,** 890–902 (2013).
18. Simmonds, M. C. *et al.* Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual-participant data. *Ann. Intern. Med.* **158,** 877–889 (2013).
19. Laurie, A. L., Chen, Y., Chou, R. & Fu, R. Meta-analysis of the Impact of Patient Characteristics on Estimates of Effectiveness and Harms of Recombinant Human Bone Morphogenetic Protein-2 in Lumbar Spinal Fusion. *Spine.* **41,** E1115–E1123 (2016).
20. Noshchenko, A., Lindley, E. M., Burger, E. L., Cain, C. M. & Patel, V. V. What Is the Clinical Relevance of Radiographic Nonunion After Single-Level Lumbar Interbody Arthrodesis in Degenerative Disc Disease? A Meta-Analysis of the YODA Project Database. *Spine.* **41,** 9–17 (2016).
21. Mospan, G. A. & Wargo, K. A. 5-Day versus 10-Day Course of Fluoroquinolones in Outpatient Males with a Urinary Tract Infection (UTI). *J. Am. Board Fam. Med.* **29,** 654–662 (2016).
22. Storgaard, H. *et al.* Benefits and Harms of Sodium-Glucose Co-Transporter 2 Inhibitors in Patients with Type 2 Diabetes: A Systematic Review and Meta-Analysis. *PLoS One.* **11,** e0166125 (2016).
23. Corbett, M. *et al.* Certolizumab pegol and secukinumab for treating active psoriatic arthritis following inadequate response to disease-modifying antirheumatic drugs: a systematic review and economic evaluation. *Health Technol. Assess.* **21,** 1–326 (2017).
24. Gay, H. C., Baldridge, A. S. & Huffman, M. D. Feasibility, Process, and Outcomes of Cardiovascular Clinical Trial Data Sharing: A Reproduction Analysis of the SMART-AF Trial. *JAMA Cardiol.* **2,** 1375–1379 (2017).
25. Wang, R., Dasgupta, A. & Ward, M. M. Comparative Efficacy of Tumor Necrosis Factor-alpha Inhibitors in Ankylosing Spondylitis: A Systematic Review and Bayesian Network Metaanalysis. *J. Rheumatol.* **45,** 481–490 (2018).

26. Mbuagbaw, L. *Review of available evidence on the use of bedaquiline for the treatment of multidrug-resistant tuberculosis: Data analysis report; Appendix to A 2016 review of available evidence on the use of bedaquiline in the treatment of multidrug-resistant tuberculosis*. Report No. WHO/HTM/TB/2017.01 (World Health Organization, 2017).
27. Singh, S., Proudfoot, J., Xu, R. & Sandborn, W. J. Impact of Obesity on Short- and Intermediate-Term Outcomes in Inflammatory Bowel Diseases: Pooled Analysis of Placebo Arms of Infliximab Clinical Trials. *Inflamm. Bowel. Dis.* **24,** 2278–2284 (2018).
28. Singh, S. *et al.* No Benefit of Concomitant 5-Aminosalicylates in Patients With Ulcerative Colitis Escalated to Biologic Therapy: Pooled Analysis of Individual Participant Data From Clinical Trials. *Am. J. Gastroenterol.* **113,** 1197–1205 (2018).
29. Singh, S., Proudfoot, J., Xu, R. & Sandborn, W. J. Obesity and Response to Infliximab in Patients with Inflammatory Bowel Diseases: Pooled Analysis of Individual Participant Data from Clinical Trials. *Am. J. Gastroenterol.* **113,** 883–889 (2018).
30. Zou, X. *et al.* The role of PANSS symptoms and adverse events in explaining the effects of paliperidone on social functioning: a causal mediation analysis approach. *NPJ Schizophr.* **4,** 13 (2018).
31. Spertus, J., Horvitz-Lennon, M., Abing, H. & Normand, S.-L. Risk of weight gain for specific antipsychotic drugs: a meta-analysis. *NPJ Schizophr.* **4,** 12 (2018).
32. Navar, A. M., Pencina, M. J., Rymer, J. A., Louzao, D. M. & Peterson, E. D. Use of Open Access Platforms for Clinical Trial Data. *JAMA.* **315,** 1283–1284 (2016).
33. Vaduganathan, M. *et al.* Availability and Use of Shared Data From Cardiometabolic Clinical Trials. *Circulation.* **137,** 938–947 (2017).
34. Ross, J. S. *et al.* Data sharing through an NIH central database repository: a cross-sectional survey of BioLINCC users. *BMJ Open.* **6,** e012769 (2016).
35. Krumholz, H. M. *et al.* Sea change in open science and data sharing: leadership by industry. *Circ. Cardiovasc. Qual. Outcomes.* **7,** 499–504 (2014).
36. Ross, J. S. & Krumholz, H. M. Ushering in a new era of open science through data sharing: the wall must come down. *JAMA.* **309,** 1355–1356 (2013).
37. Berlin, J. A. *et al.* Bumps and bridges on the road to responsible sharing of clinical trial data. *Clin. Trials.* **11,** 7–12 (2014).

Acknowledgements

The authors would like to thank the following individuals who serve, or served, on the YODA Project Steering Committee since 2011, several of whom provided comments on an earlier version of this manuscript: Allan Coukell, BScPharm; Ezekiel Emanuel, MD, PhD (Chair); Ralph Horwitz, MD; James Neaton, PhD; Steven Pearson, MD, MSc; Bruce Psaty, MD, PhD; Rita Redberg, MD, MSc; Dennis Thompson, PhD; and Deborah Zarin, MD. Steering Committee members were offered an honorarium for their time. The authors would also like to thank the following individuals who served on the YODA Project Clinical Steering Committee as part of the Medtronic partnership from 2011-2013: Kevin J. Bozic, MD, MBA; Mark Litwin, MD, MPH; Sohail K. Mirza, MD, MPH; Regis J. O'Keefe, MD, PhD; Daniel K. Resnick, MD; and Marc F. Swiontkowski, MD. Finally, the authors would like to thank Emily Finn, MPH, who contributed to earlier versions of this manuscript while affiliated with the Center for Outcomes Research and Evaluation, Yale New Haven Hospital, from 2013-2015. This research was the direct result of experience gained from research agreements through Yale University from Medtronic and Johnson & Johnson (Janssen) to develop methods for clinical trial data sharing.

Author Contributions

Dr. Ross and Ms. Ritchie drafted the paper. All authors were involved in this collaborative project, revised the paper for important intellectual content and reviewed and approved the submitted version. Dr. Ross is the guarantor and affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Additional Information

Table 4 is only available in the online version of this paper.

Competing interests: In the past 36 months, Drs. Ross, Krumholz, and Ms. Gamble have received research support through Yale University from Medtronic, Inc. and the Food and Drug Administration (FDA) to develop methods for postmarket surveillance of medical devices (U01FD004585). Drs. Ross, Krumholz, and Desai have received research support from the Centers of Medicare and Medicaid Services (CMS) to develop and maintain performance measures that are used for public reporting (HHSM-500-2013-13018I). Dr. Ross and Ms. Ritchie have received research support through Yale University from the Food and Drug Administration to establish Yale-Mayo Clinic Center for Excellence in Regulatory Science and Innovation (CERSI) program (U01FD005938). Dr. Ross, Ms. Ritchie, and Ms. Gamble have received research support through Yale University from the Blue Cross Blue Shield Association to better understand medical technology evaluation. Dr. Ross has received research support through Yale University from the Agency for Healthcare Research and Quality (R01HS022882), from the National Heart, Lung and Blood Institute of the National Institutes of Health (NIH) (R01HS025164), and from the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International and to establish the Collaboration for Research Integrity and Transparency (CRIT) at Yale. Drs. Waldstreicher and Berlin, Mr. Bamford, Ms. Childers, Mr. Lins, and Dr. Morris are employees of Johnson & Johnson. Dr. Kuntz is an employee of Medtronic, Inc. All other authors declare no competing interests.

How to cite this article: Ross, J. S. *et al.* Overview and experience of the YODA Project with clinical trial data sharing after 5 years. *Sci. Data.* 5:180268 doi: 10.1038/sdata.2018.268 (2018).

VIEWPOINT

SHARING MEDICINE

Sharing Clinical Research Data—Finding the Right Balance

Bernard Lo, MD
The Greenwall Foundation, New York, New York.

Steven N. Goodman, MD, PhD
Departments of Medicine and Health Research & Policy, Stanford University, Stanford, California.

Sharing full data from clinical trials has been extensively advocated to better understand the harms and benefits of current treatments, generate new hypotheses, and maximize knowledge gained through trial participants' altruism. Several pharmaceutical companies and the European Medicines Agency, which licenses drugs in Europe, are now sharing clinical trial data.[1] An Institute of Medicine report[1] presented a framework and the International Committee of Medical Journal Editors issued a draft proposal for clinical trial data sharing. The National Institutes of Health have expanded requirements for registration of clinical trials, reporting of summary results, and data management plans.[2] The Patient Centered Outcomes Research Institute has been working on a data sharing policy almost since its inception. Beyond clinical trials, researchers can study the effectiveness or safety of therapies via observational data collected in electronic health records within the US Food and Drug Administration (FDA) Sentinel Program, within research-oriented health care systems, and through the Patient-Centered Outcomes Research Network. In various precision medicine initiatives, patients are sharing with researchers data from personal devices and genomic sequencing of biospecimens.

However, there continues to be a lag between data-sharing intentions and the implementation of policies to make sharing happen. Relatively few researchers have requested access to newly available clinical trial data sets. To date, few important results have been published from secondary analyses of shared clinical trial data[3] even though examples exist where such availability either could, or did, change conclusions. Identifying and accessing such data sets can be difficult because sponsor platforms are not discoverable, searchable, or interoperable. Resistance to data sharing from clinical trialists has become more apparent,[3] mostly based on assertions of data ownership and academic incentives for publishing multiple articles from single studies. Academic incentives should reward data sharing that leads to secondary publications by others and that encourage collaboration between the researchers who generate the data and secondary users.[4] Fundamental issues that still need to be addressed include costs, consent, privacy, and data security.

Corresponding Author: Bernard Lo, MD, The Greenwall Foundation, One Penn Plaza, Ste 4726, New York, NY 10019 (bernardlo@greenwall.org).

Costs of Data Sharing

Many have argued for the possible benefits of sharing full data from clinical trials, but the cost of sharing research data has received less attention, even though they are of central concern to both funders and researchers. Deidentification, data curation and storage, and responding to data requests could require resources extending over many years. Pharmaceutical companies that share data from clinical trials currently bear all these costs but have indicated they cannot do so indefinitely.[1] A necessary first step is an analysis of the costs of sharing clinical trial data and of the options for sustainable and equitable funding.[1] Such information can provide a foundation for discussing how to allocate fairly the costs of data sharing.

Front-end costs could be reduced through use of common data elements and standardized formats for collecting and managing health care and research data.[1] This would also make shared data sets more interoperable and useful. A common platform for sponsors and funders to upload data and for data users to request data—or at least consolidating data sharing platforms—would also reduce infrastructure costs, facilitate data searching and access, and thereby increase the potential benefits of data sharing.

Much of the advocacy for data sharing has not grappled with its financial costs or, from the funder perspective, the tradeoff between the benefits of data sharing and the opportunity costs of funding fewer research projects. Experience to date suggests that only a small fraction of studies will have sharing requests, and for even fewer will the request yield an important scientific advance or modification of published claims. Benefits that can only be assessed when sharing is more widespread are the yield from individual-patient data meta-analyses, better assessment of safety across multiple studies and observational data sets, and possibly improved data management and analysis knowing that others might attempt replicate the analyses. Future research should analyze carefully both these benefits and costs of data sharing. If the costs are indeed high in comparison with benefits, requirements for sharing might be calibrated, with greater sharing obligations if it would be costly or difficult to obtain similar data, if the trial was directly relevant to clinical practice, or if the stakes for erroneous analysis are judged to be high.

Consent for Data Sharing

Under US federal regulations, using and sharing deidentified health data for research does not require the consent of patients. The implicit rationale is that if data cannot be reidentified, the risk to persons whose data are shared is no greater than the risks accepted in daily life.[5]

However, in the big data era, this regulatory exception to consent is outmoded because no data can be accurately characterized as "deidentified." Identifiability is not an inherent property of a data set but depends on

what other data can be combined with it.[1,6] As big data grow, reidentification becomes ever more feasible.

Can sharing data without explicit consent be justified without using an outdated concept of deidentification? Physicians and health care organizations have a moral obligation to improve clinical outcomes, and patients also have a moral obligation to allow data in the electronic health records collected during routine care to be used and shared in observational studies and some very low-risk clinical trials, with appropriate privacy safeguards.[7]

Strengthening Privacy and Security of Shared Data

Breaches of personal data held by retailers, websites, financial institutions, government agencies, and health care organizations are everyday news. How can medical and research data that are shared be better protected? First, identifiable health data should have privacy and security safeguards regardless of who holds them. Congress should extend the health privacy and security protections to all parties that collect or hold such data, including Internet service providers, websites, and mobile application and device developers.[8]

Second, even without federal requirements, data sharing should use state of the art protections, such as 256-bit encryption, virtual private networks, and testing for network security threats. Methods exist whereby data can be made available to authorized secondary users for analysis without allowing them to download it. Furthermore, distributed or federated networks can aggregate health data held by several institutions more securely than a centralized site holding data from many researchers and institutions.[1,] The FDA Sentinel project and a confederation of integrated health care systems use distributed data networks in which individual patient-level observational data never leave the site of clinical care.

Third, technical approaches to protecting privacy should be developed and adopted. In differential privacy, some values in a data set are altered so that the data set remains useful for group analyses while better protecting individuals from reidentification. In the altered data set, the risk of reidentification for an individual is no greater if the individual is included in the data set or excluded, and the usefulness of the data set is reduced by no more than a small prespecified amount.[9] This approach has been studied by computer scientists, but should be tested and if the findings are promising used more broadly on large health data sets.

Fourth, organizations that collect, store, and use large data sets containing health information for research should appoint a data access and oversight committee that includes patient or public representatives.[1,7] The committee should be empowered to identify and address important public concerns.

More needs to be learned about the societal benefits and costs of different data sharing models. The recent contest and prize sponsored by a medical journal and the National Institutes of Health (NIH) for the best secondary analysis of one large NIH-sponsored clinical trial might be broadened to a larger number of trials, similar to XPRIZE competitions.[10] These are competitions to solve formidable challenges in diverse areas, such as inventing handheld devices to diagnose disease, developing highly efficient automobiles, and cleaning up oil spills. Medical journals, professional societies, governmental agencies, and nonprofit organizations could provide forums to recommend how to study the societal benefit of funders' investments in data sharing. Such efforts could help fulfill the promise of data sharing by providing sounder evidence on how to optimize the balance between investing in data sharing, funding new research, and maintaining scientists' incentives to conduct research requiring primary data collection.

ARTICLE INFORMATION

Published Online: July 17, 2017. doi:10.1001/jamainternmed.2017.1926

Correction: This article was corrected on August 28, 2017, for an incorrect author affiliation.

Conflict of Interest Disclosures: None reported.

REFERENCES

1. Institute of Medicine. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. Washington, DC: The National Academies Press; 2015. doi:10.17226/18998

2. National Institutes of Health. HHS take steps to provide more information about clinical trials to the public. https://www.nih.gov/news-events/news-releases/hhs-take-steps-provide-more-information-about-clinical-trials-public?utm_source=twitterfeed&utm_medium=twitter. Published September 16, 2016. Accessed June 1, 2017.

3. Drazen JM, Morrissey S, Malina D, Hamel MB, Campion EW. The importance–and the complexities–of data sharing. *N Engl J Med*. 2016;375(12):1182-1183.

4. Lo B, DeMets DL. Incentives for Clinical Trialists to Share Data. *N Engl J Med*. 2016;375(12):1112-1115.

5. Department of Health and Human Services. Protection of Human Subjects 45 CFR Part 46. http://www.dhhs.gov/ohrp/humansubjects/guidance/45cfr46.html. Accessed June 1, 2017.

6. Barocas S, Nissenbaum H. Big Data's End Run around Anonymity and Consent. In: Lane J, Stodden V, Bender S, Nissenbaum H, eds. *Privacy, Big Data, and the Public Good*. New York: Cambridge University Press; 2014:44-75.

7. Faden RR, Kass NE, Goodman SN, Pronovost P, Tunis S, Beauchamp TL. An ethics framework for a learning health care system: a departure from traditional research ethics and clinical ethics. *Hastings Cent Rep*. 2013;(Spec No):S16-S27.

8. US Department of Health and Human Services. Examining Oversight of thr Privacy & Security of Health Data Cellected by Entities Not Regulated by HIPAA. https://www.healthit.gov/sites/default/files/non-covered_entities_report_june_17_2016.pdf. Updated June 17, 2016. Accessed June 1, 2017.

9. Dwork C, Roth A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theorerical Computer Science*. 2014;9:211-407.

10. XPrize. http://www.xprize.org. Accessed March 21, 2017.

searchers, clinicians, regulatory agencies, patients, and other interested parties as a means to achieve this goal.

An audio interview with Jeffrey Drazen is available at NEJM.org

Disclosure forms provided by the authors are available at NEJM.org.

From the Center for Data Intensive Science, University of Chicago, Chicago (R.L.G., A.P.H.); the Ontario Institute for Cancer Research, Toronto (V.F.); Weill Cornell Medicine, Cornell University, New York (H.E.V.); and the National Cancer Institute, Bethesda, MD (D.R.L., W.A.K., L.M.S.).

1. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014;505:495-501.
2. Shaw AT, Ou S-HI, Bang Y-J, et al. Crizotinib in *ROS1*-rearranged non–small-cell lung cancer. N Engl J Med 2014;371:1963-71.
3. Larkin J, Ascierto PA, Dréno B, et al. Combined vemurafenib and cobimetinib in *BRAF*-mutated melanoma. N Engl J Med 2014;371:1867-76.
4. Wilson WH, Young RM, Schmitz R, et al. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. Nat Med 2015;21:922-6.
5. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data — a proposal from the International Committee of Medical Journal Editors. N Engl J Med 2016;374:384-6.

DOI: 10.1056/NEJMp1607591

Incentives for Clinical Trialists to Share Data

Bernard Lo, M.D., and David L. DeMets, Ph.D.

Sharing of data from clinical trials benefits patients by enabling new discoveries, meta-analyses, and confirmation of published results. As the table shows, the European Medicines Agency (EMA), a number of drug companies, and one other trial funder have already implemented data sharing. A comprehensive Institute of Medicine (IOM) report recommends the sorts of data that should be shared, how long after a trial, and under what conditions.[1] The International Committee of Medical Journal Editors (ICMJE) proposes that the analytic data set supporting a published article be shared no later than 6 months after publication.[2] Others propose longer periods of exclusive data access for trialists.[3] The challenges now are to share data effectively and to minimize disruptions to the clinical trials system, including those affecting trialists who devote years to designing, conducting, and analyzing trials.

Many academic clinical trialists have deep concerns about data sharing.[3] They fear that other investigators will gain unfair rewards from their work and that coinvestigators and mentees will no longer have preferred access to data sets in return for working on the trial. But many trials are never published, and many secondary analyses never get done. Data sharing allows other investigators to carry out these analyses, providing the public with new knowledge gleaned from the contributions of trial participants.

Clinical trialists have practical know-how about a data set that facilitates valid secondary analyses. One of us recently returned to a large data set to carry out a secondary analysis 10 years after trial completion. Although his group had acted as the trial's biostatistics center and documented the data set extensively, including keeping statistical programs, he found he had forgotten important features of the data set, such as the rationale for defining derived variables and censoring rules. Reproducing the published results exactly was challenging because the final data set had been slightly updated from the data set used for publication — some additional events had been discovered during trial closeout. Key staff members were no longer available to provide needed details. Fortunately, with sufficient effort, the new analysis was completed successfully.[4]

Trialists tend to document what they need for their own immediate use and not consider what will be needed for later secondary analyses, perhaps even their own. If a statistical center finds it hard to reanalyze its own data set, it would be even more difficult for secondary users working with an unfamiliar data set. Collaborating with the original trialists would help other investigators derive new knowledge from shared data.

Clinical trialists would view data sharing more favorably if their concerns were addressed. First, funders and sponsors could provide resources for clinical trial data sharing.

Second, clinical trialists could be given incentives to share data. Trialists could receive appropriate acknowledgment and academic rewards when other researchers use "their" shared data to publish

Proposed Timelines for Sharing Deidentified Clinical Trial Analytic Data Sets.*

Organization	Time from Publication to Sharing of Data Supporting Published Results	Time from Publication to Sharing of Full Data Set
Data sharing enacted		
European Medicines Agency (EMA)	NA	Clinical study reports immediately after regulatory decision by EMA; will require sharing of individual patient data
ClinicalStudyDataRequest.com (consortium of 13 drug companies)	NA	Varies by company, but generally after regulatory approval (or study termination) and manuscript accepted for publication
Johnson & Johnson (Yale Open Data Access [YODA] Project)	NA	Participant-level data after regulatory approval (or study termination), manuscript accepted for publication, and 18 mo after study completion
Merck	NA	Patient-level data after regulatory approval, manuscript accepted for publication, and 18 mo after trial completion
Bristol-Myers Squibb (Supporting Open Access for Researchers [SOAR] initiative)	NA	Patient-level data after regulatory approval (or study termination) and 24 mo after study completion
Gates Foundation	On publication (as of January 2017)	NA
Data-sharing proposals		
Institute of Medicine	No later than 6 mo	18 mo after study completion (or 30 days after regulatory approval or abandonment of development if trial supports regulatory application)
ICMJE	No later than 6 mo	NA
ACCESS CV	12 mo	24 mo after publication of primary results
Consortium on Fairness in Trial Data Sharing	NA	2–5 yr after publication of primary trial results, depending on time to complete trial
Concordat on Open Research Data (U.K. funders, including Medical Research Council and Wellcome Trust, and universities)	On publication	NA

* ACCESS CV denotes Academic Research Organization Consortium for Continuing Evaluation of Scientific Studies — Cardiovascular, ICMJE International Committee of Medical Journal Editors, and NA not applicable. The companies participating in ClinicalStudyDataRequest.com are Astellas, Bayer, Boehringer Ingelheim, Daiichi-Sankyo, Eisai, GSK, Lilly, Novartis, Roche, Sanofi, Takeda, UCB, and ViiV Healthcare.

papers. Such acknowledgment requires several steps.[5] Each data set would need a unique digital identifier. Downloads of the data set and resulting publications would have to be tracked and measured. The commercial Data Citation Index (http://wokinfo.com/products_tools/multidisciplinary/dci) has such functions but tracks only some data repositories and selected biomedical journals. An effective global tracking system would include most repositories containing clinical trial data sets, including data released by pharmaceutical companies, the EMA, the National Institutes of Health (NIH), and funders of global health clinical trials.

Organizations such as Clinical and Translational Science Award (CTSA) sites, the Clinical Trials Transformation Initiative, the Association of American Medical Colleges, the Multi-Regional Clinical Trials Center, and the Wellcome Trust could convene key constituents to establish a system for tracking secondary publications from shared data sets. They should include clinical trialists, secondary investigators, research institutions, funders and sponsors, medical journals, data-sharing platforms and repositories, and patient advocates. The NIH can make data sharing a criterion for grants review, including center, program, training, and CTSA grants. With such NIH incentives in place, universities and research centers may reward data sharing in hiring and promotion decisions. Discussions

should include participants from around the world in order to obtain an appropriate global view.

Third, the responsibilities of secondary investigators analyzing shared clinical trial data need to be clarified. We believe these investigators should provide a research question and data-analysis plan when requesting data access, submit their findings to a peer-reviewed journal, and share their own data analyses.[1] These commitments, which deter invalid secondary analyses, can be included in data-use agreements and conditions for submitting manuscripts. Journal editors can require secondary investigators to explain how their analysis differs from that of the original trialists.[2]

One best practice is for secondary users to first verify that they understand the data set by trying to reproduce the published findings using the original methods, much as a basic scientist might reproduce published results before doing additional research with shared materials. Medical journals could require authors of secondary analyses to demonstrate in an appendix that they were able to substantially reproduce the published primary results. If the secondary investigators cannot reproduce those results exactly, they could have a collegial discussion with the original trialists. As described above, the final shared data set may have been updated from the publication data set. Alternatively, the secondary investigators may have misunderstood the data set or the trialists' analytic plan, or an error may have occurred in the original analysis. Such peer-to-peer discussions are a key part of scientific discovery.

Fourth, collaboration between clinical trialists and secondary users of data sets should be promoted. Data sharing is often perceived as a zero-sum game in which the original trialists lose if others perform secondary analyses. But trialists and secondary investigators could be collaborators rather than antagonists, with shared authorship as an incentive and an ideal. The usual expectation might be shared authorship unless the original team declines. Secondary users can better understand the data set if they work closely with trialists who planned the trial and managed the data. Clinical trialists can amplify their efforts and the impact of their data set if they collaborate on additional analyses. Those who do not have the time or resources to do additional analyses themselves could partner with other investigators who take primary responsibility. Trialists can continue to mentor junior colleagues on secondary analyses during collaborations with other groups, particularly if funding is available. The prospect of coauthorship also gives trialists an incentive to document the data set in sufficient detail to assist with secondary analyses that might be done years later.

Data sharing is often perceived as a zero-sum game. But trialists and secondary investigators could be collaborators rather than antagonists, with shared authorship as an incentive and an ideal.

Clinical trial teams often plan to undertake some secondary analyses, such as subgroup analyses or mechanism studies. The original trialists could, perhaps in an appendix to the primary article or as part of a data-sharing plan at trial registration, stake a time-limited claim on leading secondary analyses, describing the planned analyses, who will lead them, and expected finish dates. The various parties mentioned above would need to reach agreement on reasonable standards for claims for priority. Other investigators would most likely respect reasonable claims because trialists and mentees can begin secondary analyses after submitting the primary manuscript before the full analyzable data set needs to be shared. The most efficient way to glean additional knowledge from completed trials in a timely manner would be to fund coinvestigators to carry out important secondary analyses.

Changing the culture of clinical trials requires a multipronged approach. With mandates for data sharing, clinical trialists and sponsors will no longer have exclusive long-term control over secondary analyses. Most trialists regard their mission as increasing knowledge about the benefits and risks of therapies, thereby helping patients in need. This professionalism will be strengthened if their concerns regarding data sharing are addressed.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Greenwall Foundation, New York (B.L.); and the Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison (D.L.D.).

1. Institute of Medicine. Sharing clinical trial data: maximizing benefits, minimizing risk. Washington, DC: National Academies Press, 2015 (http://www.nap.edu/catalog/18998/sharing-clinical-trial-data-maximizing-benefits-minimizing-risk).
2. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data — a proposal from the International Committee of Medical Journal Editors. N Engl J Med 2016;374:384-6.
3. The International Consortium of Investigators for Fairness in Trial Data Sharing. Toward fairness in data sharing. N Engl J Med 2016;375:405-7.
4. Bristow MR, Saxon LA, Feldman AM, Mei C, Anderson SA, DeMets DL. Lessons learned and insights gained in the design, analysis, and outcomes of the COMPANION trial. JACC Heart Fail 2016;4:521-35.
5. National Research Council. For attribution — developing data attribution and citation practices and standards. Washington, DC: National Academies Press, 2012.

DOI: 10.1056/NEJMp1608351

Data Sharing at a Crossroads

Frank Rockhold, Ph.D., Perry Nisen, M.D., Ph.D., and Andrew Freeman, B.Sc.

Sharing patient-level data from clinical trials can improve the quality of research and our understanding of disease and medical treatments. Various concerns have been voiced about data sharing; they involve privacy, consent, intellectual property, costs, infrastructure, data standards, free-riding researchers, and potentially erroneous conclusions. Many of these concerns cannot be totally eliminated, but they can be mitigated and managed.

The clinical research community is at an important crossroads. We believe that sharing data is the right thing to do and that we need to find the best ways to realize the benefits while minimizing the risks. Multiple different approaches and systems may be creating a fragmented, complex, and confusing landscape in which data sharing's full benefits will not be realized.

GlaxoSmithKline (GSK) took a step in 2013 with the aim for there to be a single system through which clinical trial data could be easily shared by sponsors.[1] Initial signs were encouraging. The request site we launched was relaunched in January 2014 (https://clinicalstudydatarequest.com), now including studies from other sponsors or data generators. Today, there are more than 3000 trials listed from 13 industry sponsors.

The costs and required resources for data sharing have presented a major barrier for academic and smaller sponsors. An investment of about $30,000 to $50,000 per year is needed for an academic sponsor to list up to 20 studies on the request site and for up to 10 research projects to be undertaken using data in the secure access site. Additional costs for requested studies include those for administering requests, collating data sets and relevant study documentation, anonymizing data and documents and loading them onto the access site, and providing support for researchers. The overall costs can seem disproportionately high for sponsors or investigators with few trials.

From the start, we believed that proposals for research conducted with patient-level data should be reviewed for scientific merit as a condition of access and that such review could be conducted objectively only by a panel independent of study sponsors. We appointed the original independent review panel (IRP). Sponsors checked research proposals to ensure that they were complete and met conditions for data access, then sent them to the IRP. Sponsors communicated the outcomes to researchers. To strengthen independence, in 2015 the Wellcome Trust began managing proposal review, interacting with sponsors, the IRP, and researchers. The Trust has appointed a new IRP with no sponsor involvement, which has been operating since December 2015 (its members are listed, along with their charter, at https://clinicalstudydatarequest.com).

The original panel's default approach was to approve proposals and permit access unless there was a compelling reason not to do so.[2] More than 200 research proposals have been submitted to date. Of those that have met requirements and not been withdrawn, less than 10% have been rejected or have resulted in the researchers being advised to resubmit the proposal. There are currently nearly 100 ongoing research projects that are using requested data. To date, however, only four analyses have been published using these data. In the first year, the majority of proposed research aimed to ask

From the Greenwall Foundation, New York (B.L.); and the Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison (D.L.D.).

1. Institute of Medicine. Sharing clinical trial data: maximizing benefits, minimizing risk. Washington, DC: National Academies Press, 2015 (http://www.nap.edu/catalog/18998/sharing-clinical-trial-data-maximizing-benefits-minimizing-risk).
2. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data — a proposal from the International Committee of Medical Journal Editors. N Engl J Med 2016;374:384-6.
3. The International Consortium of Investigators for Fairness in Trial Data Sharing. Toward fairness in data sharing. N Engl J Med 2016;375:405-7.
4. Bristow MR, Saxon LA, Feldman AM, Mei C, Anderson SA, DeMets DL. Lessons learned and insights gained in the design, analysis, and outcomes of the COMPANION trial. JACC Heart Fail 2016;4:521-35.
5. National Research Council. For attribution — developing data attribution and citation practices and standards. Washington, DC: National Academies Press, 2012.

DOI: 10.1056/NEJMp1608351

Data Sharing at a Crossroads

Frank Rockhold, Ph.D., Perry Nisen, M.D., Ph.D., and Andrew Freeman, B.Sc.

Sharing patient-level data from clinical trials can improve the quality of research and our understanding of disease and medical treatments. Various concerns have been voiced about data sharing; they involve privacy, consent, intellectual property, costs, infrastructure, data standards, free-riding researchers, and potentially erroneous conclusions. Many of these concerns cannot be totally eliminated, but they can be mitigated and managed.

The clinical research community is at an important crossroads. We believe that sharing data is the right thing to do and that we need to find the best ways to realize the benefits while minimizing the risks. Multiple different approaches and systems may be creating a fragmented, complex, and confusing landscape in which data sharing's full benefits will not be realized.

GlaxoSmithKline (GSK) took a step in 2013 with the aim for there to be a single system through which clinical trial data could be easily shared by sponsors.[1] Initial signs were encouraging. The request site we launched was relaunched in January 2014 (https://clinicalstudydatarequest.com), now including studies from other sponsors or data generators. Today, there are more than 3000 trials listed from 13 industry sponsors.

The costs and required resources for data sharing have presented a major barrier for academic and smaller sponsors. An investment of about $30,000 to $50,000 per year is needed for an academic sponsor to list up to 20 studies on the request site and for up to 10 research projects to be undertaken using data in the secure access site. Additional costs for requested studies include those for administering requests, collating data sets and relevant study documentation, anonymizing data and documents and loading them onto the access site, and providing support for researchers. The overall costs can seem disproportionately high for sponsors or investigators with few trials.

From the start, we believed that proposals for research conducted with patient-level data should be reviewed for scientific merit as a condition of access and that such review could be conducted objectively only by a panel independent of study sponsors. We appointed the original independent review panel (IRP). Sponsors checked research proposals to ensure that they were complete and met conditions for data access, then sent them to the IRP. Sponsors communicated the outcomes to researchers. To strengthen independence, in 2015 the Wellcome Trust began managing proposal review, interacting with sponsors, the IRP, and researchers. The Trust has appointed a new IRP with no sponsor involvement, which has been operating since December 2015 (its members are listed, along with their charter, at https://clinicalstudydatarequest.com).

The original panel's default approach was to approve proposals and permit access unless there was a compelling reason not to do so.[2] More than 200 research proposals have been submitted to date. Of those that have met requirements and not been withdrawn, less than 10% have been rejected or have resulted in the researchers being advised to resubmit the proposal. There are currently nearly 100 ongoing research projects that are using requested data. To date, however, only four analyses have been published using these data. In the first year, the majority of proposed research aimed to ask

new questions using data from multiple studies rather than reanalyzing single studies.[2] This finding is consistent with a review of proposals including those submitted to two other data-sharing systems, the Yale University Open Data Access (YODA) project and the Supporting Open Access Research (SOAR) initiative.[3]

Should more research have been conducted, and more articles been published, with the available data? Greater awareness of data availability and expertise in using data from clinical trials in pooled and meta-analyses are likely to increase use. Since the research is conducted and published independently of the study sponsors, we don't know why there are so few publications. The yield may reflect the time required for preparing and submitting publications, but also perhaps difficulties in conducting analyses of data from trials that used different structures and standards for data and metadata. Monitoring of research outputs can demonstrate the benefits of data sharing but also inform adjustment of systems, processes, and mechanisms and encourage consistency of data standards.

Several factors were important in developing the system. GSK encouraged other sponsors to join and benefit from the infrastructure. An industry commitment to data sharing was important,[4] as was sponsors' ability to list studies for data sharing according to individual company policies. Sponsors have collaborated, respecting different perspectives in finding solutions.

Not surprisingly, such collaboration can be challenging. For example, some sponsors wanted the option to deny requests in cases of potential conflict of interest or competitive risk. Others didn't require this safety net, which could be seen as compromising independence. The Wellcome Trust agreed to accommodate this approach to enable more sponsors to share data through a single system, which benefits researchers by increasing the amount of data available from as many sponsors as possible. The safety net has not been used to date, and as sponsors gain experience, they may also gain confidence that it isn't needed.

Our vision of a simple single system may be challenging to achieve. But there's a risk that if myriad systems emerge, the benefits will be limited by the complexity of obtaining data.

Our vision of a simple single system may be challenging to achieve. Some industry sponsors such as Johnson & Johnson and Pfizer have chosen to set up their own systems, the Duke Clinical Research Institute has announced that its patient data related to cardiovascular disease will be made available through SOAR (http://soar.dcri.org), and there are other disease-based systems such as the Virtual International Stroke Trials Archive (http://vistacollaboration.org) and Project Data Sphere (https://projectdatasphere.org).

Though these commitments and the proposal of the International Committee of Medical Journal Editors to make data sharing a condition of publication[5] are welcome, there's a risk that if myriad systems emerge, the benefits will be limited by the complexity of obtaining data. The Multi-Regional Clinical Trials (MRCT) Center of Brigham and Women's Hospital and Harvard University, with whom we consulted, is leading an initiative to create a single portal through which data from multiple sponsors and systems can be shared. It is critical that this portal be focused and fit for its purpose and that it work in concert with existing systems.

One possible approach is for everyone to move to a single system whereby sponsors or investigators send study details, data, or both to an independent custodian who manages scientific review, privacy, and other aspects. This approach would require sponsors to give up their own effective systems but would realize economies of scale, helping to address cost barriers. Alternatively, the provision of data-sharing services for some sponsors could be combined with a federated model offering a central portal linking to other systems.

This portal would have to be more than a directory of systems. For receiving and reviewing data requests, it makes sense to have a common proposal form and data-sharing agreement, along with a mutual-recognition approach, whereby approval by an IRP linked to one system is considered sufficient by sponsors who

routinely rely on a different IRP. For accessing data, it may be difficult or impossible for a single project to use data from diverse secure systems. One possible solution is for the researcher (rather than the sponsor) to provide the secure environment into which anonymized data from different sponsors can be securely downloaded. That approach would reduce sponsors' costs and enable researchers to use software other than that provided in secure access systems. Since it could increase privacy risks and misuse of data, however, it might have to be combined with a researcher-accreditation system. As part of maintaining accreditation for data use, the researcher's secure system and the research conducted could be subject to independent audit.

We believe the clinical research enterprise needs to come together to build on what exists and create a simple one-stop shop for clinical trial data sharing. If we get this system right, it could provide a basis for sharing other types of data, such as preclinical data and real-world epidemiologic data. If we allow inevitable differences in systems or processes to produce a fragmented, uncoordinated approach, we will miss the opportunity to realize great value for patients.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

From the Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC (F.R.); the Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA (P.N.); and Global Medical, GSK, Middlesex, United Kingdom (A.F.).

1. Nisen P, Rockhold F. Access to patient-level data from GlaxoSmithKline clinical trials. N Engl J Med 2013;369:475-8.
2. Strom BL, Buyse M, Hughes J, Knoppers BM. Data sharing, year 1 — access to data from industry-sponsored clinical trials. N Engl J Med 2014;371:2052-4.
3. Navar AM, Pencina MJ, Rymer JA, Louzao DM, Peterson ED. Use of open access platforms for clinical trial data. JAMA 2016;315:1283-4.
4. Bierer BE, Li R, Barnes M, Sim I. A global, neutral platform for sharing trial data. N Engl J Med 2016;374:2411-3.
5. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data — a proposal from the International Committee of Medical Journal Editors. N Engl J Med 2016;374:384-6.

DOI: 10.1056/NEJMp1608086

VIEWPOINT

Finding Means to Fulfill the Societal and Academic Imperative for Open Data Access and Sharing

Eric D. Peterson, MD, MPH
Division of Cardiology, Department of Medicine, Duke Clinical Research Institute, Duke University School of Medicine, Durham, North Carolina.

Frank W. Rockhold, PhD
Biostatistics and Bioinformatics, Duke Clinical Research Institute, Duke University School of Medicine, Durham, North Carolina.

Overview

Calls for greater data sharing and transparency in clinical research have intensified in recent years. The potential benefits of data sharing are multifaceted. Increased data transparency can allow research findings to be independently verified; be a valuable resource to address multiple secondary questions; allow combinations with other study data, thereby enabling meta-analytic insights; and facilitate better planning of future research. Data sharing is also consistent with the fundamental principle that science should be a collaborative process used to its maximal common benefit.

In June 2017, these altruistic ideals began to have real operational consequences. The International Committee of Medical Journal Editors (ICMJE) stated that clinical trials submitted to ICMJE-compliant journals after July 1, 2018, must include an explicit data-sharing statement, and by January 2019, all trials must include a data-sharing plan at trial registration.[1] While the ICMJE fell short of requiring full data sharing, the editors commented that it was an "ethical obligation to responsibly share data generated by interventional clinical trials because trial participants have put themselves at risk."[1(p2491)] This drive toward increased trial transparency mirrors the ambitions already set forth by the National Institutes of Health[2] and the National Academy of Sciences.[3]

While the ethical reasons why data should be shared are clear, making data sharing more practical presents several operational challenges (**Table**). These include issues with informed consent, questions about how to protect patient privacy, the need for safe, user-friendly, and efficient data platforms, a lack of information on where research data exist and in what form, a lack of common data standards and definitions, and issues with how to credit the investigators who originally collected the data. Thankfully, progress is being made on each of these topics.

The Existing Evidence

The Office for Human Research Protections recently clarified that the use of deidentified patient data was allowed for future research, as long as it is not expressly prohibited in the informed consent procedure for the study from which the data originated.[5] Most patients are willing to have their clinical study data used to benefit future patients[4]; however, they want this done in a manner that also assures that their privacy is protected to the greatest extent possible.[4]

Protecting patient privacy remains a huge issue, but it can be facilitated by a combination of better deidentification and safer data platforms. The European Medicines Agency has formed an advisory group called the Technical Anonymisation Group[6] to establish clear guidelines for patient data deidentification while retaining data usefulness. Additionally, a number of secure, multisponsor data platforms are now available that can give researchers data access and allow analysis, but confine data use to specific questions and behind safe firewalls. These protect patient privacy and limit nonauthorized use of clinical data.

Secure data platforms also often feature advanced search tools that make it easier for a researcher to understand what data are available and what specific clinical data are collected within each study dataset. These search engines can also often identify specific patient types across study cohorts, further facilitating the work of external investigators. As an example, a search engine being developed by Vivli (http://www.vivli.org) aims to provide more advanced search tools and increased data utility through the use of standards in the description of clinical studies and the storage of the data themselves.[7] Similarly, the American Heart Association Precision Medicine Platform (https://precision.heart.org/) is a cloud-based data resource that allows the research community access to multiple trial and observational databases, along with search capacities supported by Amazon Web Services.

Data standards and common data models exist (eg, those available from the Clinical Data Interchange Standards Consortium at https://www.cdisc.org/standards) but they have not been routinely applied across all clinical studies to date. Lacking such standards across studies, investigators and sponsors must invest a significant amount of post hoc informatics resources to facilitate data usability. This can be a time-consuming and expensive process. In the future, studies may preemptively use more standardized approaches to data collection, which will allow for more straightforward data access once completed.

The final issue is providing appropriate credit to the investigators who originally conducted a study. Carrying out a large observational study or clinical trial requires significant effort on the part of the investigative team. Thus, there is a need to ensure that all involved in the original study are adequately recognized and rewarded in subsequent published analyses. To some extent, existing mechanisms allow external investigators to ask questions submitted to study publication committees. For example, the Global Utilization of Streptokinase and Tissue Plasminogen Activator for Occluded Coronary Arteries (GUSTO) Investigators generated nearly 100 publications and included almost 1000 investigators around the world (written communication, R Califf, MD MACC, Month Year). However, this degree of access and data use is atypical; most studies pro-

Corresponding Author: Eric D. Peterson, MD, MPH, 2400 Pratt St, Ste 7009, Duke Box 3236, Durham, NC 27705 (eric.peterson@duke.edu).

jamanetwork/2018/car/jun2018/hvp180001 PAGE: right 1 SESS: 16 OUTPUT: Feb 19 11:22 2018

Table. Remaining Barriers to Data Sharing and Their Solutions

Issue	Actions Taken or in Progress
Informed consent	Patients generally support the sharing of their data.[4] The Office for Human Research Protections has issued a statement that if standard informed consent forms are silent on data sharing, this will be taken to imply agreement to data sharing on the part of the signer.
Patient privacy	Many acceptable deidentification protocols exist. Given that a 100% guarantee of protection is impossible, various privacy regulations, such as the US Health Insurance Portability and Accountability Act and the EU General Data Protection Regulation are defining acceptable standards for deidentification.
Need for efficient platforms	Data platforms need to be broad-based portals that enhance the ability of researchers to utilize available data to make meaningful contributions. These will require access to analytical software, sophisticated search capabilities that can locate trials, and accessible individual participant data.
Data standards	Data standards exist and represent a resource and cost issue for many smaller institutions. It is recognized this is a challenge, but implementation of data standards or interoperability is critical for ultimate transparency of disclosed information.
Credit for data originators	As recommended in the Institute of Medicine report[2] and numerous other publications, there is need to promote collaboration between data originators and secondary researchers, create a citation tool for credit to data originators when their data are used, and shift the academic promotion system to accommodate the sharing of credit between data originators and users.

duce few secondary papers, if any, beyond those generated by the original investigators.[8] As opposed to seeing the multiplicative value of open access, investigators often see sharing data as a zero-sum game in which they lose if others perform additional analyses. Researchers also express concerns that others outside the original trial may not fully understand the intricacies of the study design or data and that this could affect the validity of these analyses.

While these concerns are valid, they can be addressed in several ways. The National Academy of Medicine and journal editors have first acknowledged the need for an exclusivity period for up to 2 years after study completion before broader access is offered to allow the original researchers to publish primary and major secondary manuscripts. Additionally, external researchers are encouraged to include original study investigators as co-authors to help assure appropriate analysis and interpretation. Third, the effort of original investigators to generate data could be directly referenced in any subsequent publications from these data. Once the criteria for data authorship references are established, these could be used as traditional author citations are used for faculty promotion and tenure consideration.[2,9,10]

Conclusions

The data access effort should be viewed as an imperative journey that benefits patients as well as society and science in general. Today, requirements from journals as well as regulators and funders are rapidly making data sharing a mandate. As noted in this article, there remain many operational challenges to data sharing. However, these issues are being addressed as we move toward the final frontier of ensuring high quality data, and the right balance between patient privacy, data utility, and appropriate credit for data originators and users. Collaboration between original investigators and secondary users of data will continue to increase as these operational issues are overcome. Ultimately, these processes will promote greater data access to all researchers and better science for all.

ARTICLE INFORMATION

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. The authors are advisors to the Supporting Open Access for Researchers (SOAR) and Vivli data sharing platforms. Dr Peterson reports grants from Bristol-Meyers Squibb and is the Executive Director of Duke Clinical Research Institute, which runs many large clinical trials. Dr Rockhold sits on the European Medical Agency Technical Anonymisation Group; he also reports personal fees from GlaxoSmithKline, EMD Serono, Janssen, UCB Biosciences, NovoNordisk, Amgen, and AbbVie, and grants from Janssen and AstraZeneca. No other disclosures were reported.

REFERENCES

1. Taichman DB, Sahni P, Pindborg A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *JAMA*. 2017;317(24):2491-2492.

2. Institute of Medicine. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. Washington, DC: National Academies Press; 2015.

3. Hudson KL, Lauer MS, Collins FS. Toward a new era of trust and transparency in clinical trials. *JAMA*. 2016;316(13):1353-1354.

4. Drazen J. NEJM: Bringing together ideas for sharing clinical trial data (interview with Jeffrey Drazen, M.D., Editor-in-Chief, *The New England Journal of Medicine*). http://libraryhub.nejm.org/article-hub/2017/07/nejm-bringing-together-ideas-sharing-clinical-trial-data/. July 11, 2017. Accessed February 12, 2018.

5. Medikoff J. Letter to ICMJE secretariat. http://icmje.org/news-and-editorials/menikoff_icmje_questions_20170307.pdf. Published March 7 2017; Accessed February 6, 2018.

6. European Medicines Agency. Clinical data publication: technical anonymisation group. http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_001880.jsp. Published March 2017. Accessed February 6, 2018.

7. Bierer BE, Li R, Barnes M, Sim I. A global, neutral platform for sharing trial data. *N Engl J Med*. 2016;374(25):2411-2413.

8. Ramanan M, Myburgh J, Finfer S, Bellomo R, Venkatesh B. Publication of secondary analyses from randomized trials in critical care. *N Engl J Med*. 2016;375(21):2105-2106.

9. Lo B, DeMets DL. Incentives for clinical trialists to share data. *N Engl J Med*. 2016;375(12):1112-1115.

10. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. *N Engl J Med*. 2017;376(17):1684-1687.

https://osp.od.nih.gov/2019/11/06/draft-data-management-and-sharing-policy-we-need-to-hear-from-you/

NIH's DRAFT Data Management and Sharing Policy: We Need to Hear From You!

Around this time last year, I wrote about a request for information (RFI) on potential key elements that could comprise a future NIH data management and sharing policy. Not surprisingly, we received a lot of helpful feedback. Most commenters supported data sharing and the importance of prospectively planning for where, when, and how scientific data should be managed and shared. There were, however, concerns about how one policy could fit all sizes and types of data across the biomedical research universe as well as potential burden on the research community.

Over the course of the last year, NIH has been incorporating many of these suggestions into our thinking and continuing to engage the community on their thoughts about data management and sharing. We've also been working with sovereign Tribal Nations through consultation sessions held across the U.S which have been vital in shaping NIH"s perspective on the potentially unique data sharing needs of those communities.

Today, NIH has released for public comment in the Federal Register a *Draft NIH Policy for Data Management and Sharing* along with supplement draft guidance. The draft policy furthers NIH longstanding commitment to making available the results and products of the research we fund and conduct.

To facilitate public comments, NIH has established a web-portal where folks can easily and securely provide their feedback. The portal can be accessed by clicking here. To ensure that your comments are considered, responses must be submitted no later than January 10, 2020.

I recognize that there is a perception that a Draft policy represents a finished product, that NIH has already made up our mind. I want to assure you that this is not the case. It is very important that we hear from the stakeholder community about what you think works and doesn't work with

respect to what we have proposed. We are also eager to hear your thoughts on the utility of the supplemental draft guidance or recommendations for any other guidance materials that would be helpful. I previously wrote a blog on best practices for public comments that you might find useful.

Finally, to further engage stakeholders, NIH will be hosting a webinar on the draft policy in the near future. Please stay tuned for the details. We look forward to hearing what you think about the draft policy and supplemental draft guidance and encourage you to broadly share its availability and our request for comments. With your help, we can ensure that the draft policy maximizes the myriad benefits of data management and sharing while minimizing the burden to the research community.

By admin| November 6, 2019|Blog|0 Comments

DRAFT NIH Policy for Data Management and Sharing

I. Purpose

The NIH Policy for Data Management and Sharing (herein referred to as the Policy) reinforces NIH's longstanding commitment to making the results and outputs of the research that it funds and conducts available to the public. Data sharing enables researchers to rigorously test the validity of research findings, strengthen analyses through combined datasets, reuse hard-to-generate data, and explore new frontiers of discovery. In addition, NIH emphasizes the importance of good data management practices, which provide the foundation for effective data sharing and improve the reproducibility and reliability of research findings. NIH encourages data management and data sharing practices consistent with the NIH Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research and the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles.

To promote effective and efficient data management and data sharing, NIH expects researchers to manage scientific data resulting from NIH-funded or conducted research and prospectively plan for which scientific data will be preserved and shared. Under this Policy, individuals and entities would be required to provide a Data Management and Sharing Plan (Plan) describing how scientific data will be managed, including when and where the scientific data will be preserved and shared, prior to initiating the research study. Shared data should be made accessible in a timely manner for use by the research community and the broader public. This Policy is intended to establish expectations for Data Management and Sharing Plans upon which other NIH Institutes, Centers and Offices (ICO) may supplement as appropriate.

II. Definitions

For the purposes of this Policy, terms are defined as follows:

- **Data Management and Sharing Plan (Plan):** A plan describing how scientific data will be managed, preserved, and shared with others (e.g., researchers, institutions, the broader public), as appropriate.

- **Data Management:** The process of validating, organizing, securing, maintaining, and processing scientific data, and of determining which scientific data to preserve.

- **Data Sharing:** The act of making scientific data available for use by others (e.g., researchers, institutions, the broader public).

- **Metadata:** Data describing scientific data that provide additional information to make such scientific data more understandable (e.g., date, independent sample and variable description, outcome measures, and any intermediate, descriptive, or phenotypic observational variables).

- **Scientific Data:** The recorded factual material commonly accepted in the scientific community as necessary to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses,

completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens. NIH expects that reasonable efforts will be made to digitize all scientific data.

III. Scope

This Policy applies to all research, funded or conducted in whole or in part by NIH, that results in the generation of scientific data. This includes research funded or conducted by extramural grants, contracts, intramural research projects, or other funding agreements regardless of NIH funding level or funding mechanism.

IV. Effective Date(s)

The effective date of this Policy and subsequent implementation deadlines are dependent upon feedback on this proposal. This Policy is proposed to be effective for NIH-funded or conducted research, including:

- Competing grant applications that are submitted to NIH for a future receipt date or subsequent receipt dates (date yet to be determined);
- Proposals for contracts that are submitted to NIH on or after a future date (date yet to be determined);
- NIH Intramural research conducted on or after a future date (date yet to be determined); and
- Other funding agreements (e.g., Other Transactions) that are executed on or after a future date (date yet to be determined), unless otherwise stipulated by NIH.

V. Requirements

This Policy would require:

- Submission of a Data Management and Sharing Plan (Plan) outlining how scientific data will be managed and shared, taking into account any potential restrictions or limitations.
- Compliance with the NIH ICO-approved Plan, prospectively describing effective management and timely sharing of scientific data (as appropriate) and accompanying metadata resulting from NIH-funded or conducted research.

The funding NIH ICO may request additional or specific information to be included within the Plan in order to meet expectations for data management and data sharing in support of programmatic priorities or to expand the utility of the scientific data generated from the research. Costs associated with data management and data sharing may be allowable under the budget for the proposed project (Supplemental DRAFT Guidance: Allowable Costs for Data Management and Sharing).

VI. Data Management and Sharing Plans

Researchers with NIH-funded or conducted research projects resulting in the generation of scientific data are required to submit a Plan to the funding NIH ICO as part of Just-in- Time for extramural awards, as part of the technical evaluation for contracts, as part of the NIH Intramural Annual Report, or prior to release of funds for other funding agreements. Plans should explain how scientific data generated by a research study will be managed and which of these scientific data will be shared. Plans may be updated by researchers (with appropriate NIH ICO approval) during regular reporting intervals if changes are necessary or at the request of the NIH ICO to reflect changes in the previously documented approach to data management and data sharing throughout the research project, as appropriate. NIH encourages shared scientific data to be made available as long as it is deemed useful to the research community or the public. Plans should also identify strategies or approaches to ensure data security and compliance with privacy protections are in place throughout the life of the scientific data. NIH may make Plans publicly available.

NIH prioritizes the responsible management and sharing of scientific data derived from human participants. Applicable Federal, Tribal, state, and local laws, regulations, statutes, guidance, and institutional policies dictate how research involving human participants should be conducted and how the scientific data derived from human participants should be used. Researchers proposing to generate scientific data derived from human participants should outline in their Plans how human participants' privacy, rights, and confidentiality will be protected, i.e., through de-identification or other protective measures. NIH recognizes that certain factors (e.g., legal, ethical, technical) may limit the ability to preserve and share data. Plans should include consideration of these factors, when applicable, in describing the approach to data management and data sharing. NIH encourages the use of established repositories for preserving and sharing scientific data.

Plan Elements: Consider addressing specific elements outlined in Supplemental DRAFT Guidance: Elements of An NIH Data Management and Sharing Plan

Plan Assessment: The funding NIH ICO will assess the Plan, through the following processes:

- Extramural Awards: Plans will undergo a programmatic assessment by NIH staff within the proposed funding NIH ICO. NIH encourages potential awardees to work with NIH staff to address any potential concerns regarding the Plan prior to submission.

- Contracts: Plans will be included as part of the technical evaluation performed by NIH staff.

- Intramural Research Projects: Plans will be assessed by the Scientific Director (or designee) or Clinical Director (or designee) of the researcher's funding NIH ICO.

- Other funding agreements: Plans will be assessed in the context of other funding agreement mechanisms (e.g., Other Transactions).

VII. Compliance and Enforcement

During the Funding or Support Period:

During the funding period, compliance with the Plan will be determined by the funding NIH ICO. Compliance with the Plan, including any Plan updates, will be reviewed during regular reporting intervals (e.g., at the time of annual Research Performance Progress Reports (RPPRs)) at a minimum.

- Extramural Awards: The Plan will become a Term and Condition of the Notice of Award. Failure to comply with the Terms and Conditions may result in an enforcement action, including additional special terms and conditions or termination of the award, and may affect future funding decisions.

- Contracts: The Plan will become a Term and Condition of the Award, and compliance with and enforcement of the Plan will be consistent with the award and the Federal Acquisition Regulations (FAR), as applicable.

- Intramural Research Projects: Compliance with and enforcement of the Plan will be consistent with applicable NIH policies established by the NIH Office of Intramural Research and the applicable NIH ICO.

- Other funding agreements: Compliance with and enforcement of the Plan will be consistent with applicable NIH policies.

Post Funding or Support Period

After the end of the funding period, non-compliance with the NIH ICO-approved Plan may be taken into account by the funding NIH ICO for future funding decisions for the recipient institution (e.g., as authorized in the NIH Grants Policy Statement, Section 8.5, Special Award Conditions, and Remedies for Noncompliance (Special Award Conditions and Enforcement Actions)).