

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy
Board on Health Care Services

Changing the Culture of Data Management and Sharing

A Workshop



April 28, 2021
11:00 AM – 4:10 PM ET

April 29, 2021
11:00 AM – 4:00 PM ET

Virtual Meeting
Webcast Link:

<https://www.nationalacademies.org/event/04-29-2021/changing-the-culture-of-data-management-and-sharing-a-workshop>

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy
Board on Health Care Services

**Changing the Culture of Data Management and
Sharing: A Workshop**

April 28-29, 2021

Table of Contents

AGENDA	3
WORKSHOP INFORMATION	13
Statement of Task	14
Planning Committee Roster.....	15
Planning Committee Biographies.....	16
Speaker Biographies	20
Speaker Guidance.....	30
Registered Attendee Information.....	34
National Academies Staff Roster	36
BACKGROUND INFORMATION	37
Links to Background Articles.....	38
Additional Resources.....	41

AGENDA

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy
Board on Health Care Services

Changing the Culture of Data Management and Sharing
A Workshop

April 28-29, 2021

Virtual Workshop

TIMELINE:

April 28, 2021: 11:00 AM – 4:10 PM ET

April 29, 2021: 11:00 AM – 4:00 PM ET

STATEMENT OF TASK: In response to a request from the NIH Office of Science Policy, a planning committee of the National Academies of Sciences, Engineering, and Medicine will convene a two-day virtual public workshop to discuss the challenges and opportunities for researchers, institutions, and funders to establish effective data management and sharing practices. The objective of the workshop is to examine strategies, resources, and promising practices for developing and evaluating data management and sharing plans, as well as to discuss how researchers can effectively share scientific data over the course of the data life cycle.

Input will be sought from a variety of perspectives, including researchers, data repository managers, funding institutions, publishers, research participants, and other stakeholders to include a diversity of biomedical research fields and disciplines. With an emphasis on illustrative case studies, real world examples, and promising practices, potential topics may include:

- Addressing overarching strategies for managing and sharing data, taking into consideration diverse needs (e.g., human vs non-human data, type and size, data generators vs data users);
- Assessing value of shared data and the development and evaluation of data management and sharing plans, which may include discussions of:
 - best practices for repositories to collect the metrics needed to make such assessments,
 - the extent to which data value was anticipated and planned for prior to generating or sharing data, and
 - how this might inform prospective planning for sharing;
- Monitoring and evaluating data management and sharing practices, including discussion of appropriate metrics for timelines of data availability and life cycles of different types of scientific data; and
- Considering educational and other resource needs for responsible data sharing practices.

The workshop planning committee will develop the agenda for the workshop, select and invite speakers and discussants, and moderate or select moderators for the discussions. At the end of each workshop day, key points from individual speakers will be summarized and shared with the audience by a moderator. The workshop website will contain presentations from the speakers who provide the National Academies with permission to share their slides and the video of the webcast from the workshop.

DAY 1: Wednesday, April 28, 2021

11:00 a.m. **Introduction and Charge to the Workshop Speakers and Participants**

MARYANN MARTONE, *Workshop Planning Committee Co-Chair*
Professor, Neurosciences
UCSD

RICHARD NAKAMURA, *Workshop Planning Committee Co-Chair*
(Retired) Former Director, Center for Scientific Review
National Institutes of Health (NIH)

Moderator: Richard Nakamura

11:10 a.m. **Goals for the NIH Data Management and Sharing Policy**

LYRIC JORGENSEN
Deputy Director for the Office of Science Policy
National Institutes of Health (NIH)

11:20 a.m. **Keynote Presentation – What Has the COVID-19 Pandemic Taught Us About Data Sharing and Open Science?**

PATRICIA BRENNAN
Director
National Library of Medicine

11:35 a.m. **Clarifying Questions from the Audience**

11:40 a.m. **Panel Discussion: Perspectives on Data Management and Sharing Across Different Types of Data**

Moderator: Maryann Martone, UCSD

ATUL BUTTE
Priscilla Chan and Mark Zuckerberg Distinguished Professor
UCSF

LARA MANGRAVITE
President
Sage Bionetworks

ALEXANDER ROPELEWSKI
Operations Director
Brain Image Library

JOSHUA WALLACH
Assistant Professor of Epidemiology
Yale School of Public Health

SESSION I. STRATEGIES FOR MANAGING AND SHARING DATA: DIVERSE NEEDS AND CHALLENGES

Moderator: Elaine Mardis, Nationwide Children's Hospital

Session Objectives:

- Explore challenges and potential solutions to effective data management and sharing across a range of scientific disciplines.
- Discuss how data can be effectively shared over the course of the data life cycle.

12:10 p.m. **Data Formatting: Exploring Challenges and Potential Solutions**

DAVID HAUSSLER
Investigator, Howard Hughes Medical Institute
Distinguished Professor, Biomolecular Engineering
UC Santa Cruz

ADAM FERGUSON
Associate Professor
Department of Neurological Surgery, Brain and Spinal Injury Center
UCSF

12:30 p.m. **General Challenges with Data Sharing: Investigator and Repository Perspectives**

REBECCA KOSKELA
Executive Director
Research Data Alliance

RUSSELL POLDRACK
Albert Ray Lang Professor of Psychology
Stanford University

JEREMY WOLFE
Director, Visual Attention Lab
Brigham & Women's Hospital
Professor of Ophthalmology & Radiology
Harvard Medical School

12:55 p.m. **Break**

SESSION II. MONITORING AND EVALUATING DATA MANAGEMENT AND SHARING PRACTICES

Moderator: Wouter Haak, Elsevier

Session Objectives:

- To examine possible approaches for monitoring and measuring the success of data sharing and management across different types of scientific data.

1:20 p.m. **Exploring the Current State of Data Citation Methods and Tools**

DANIELLA LOWENBERG
Director, Make Data Count
California Digital Library

ALBERTO ZIGONI
Market Development Director
Research Data Management Solutions
Elsevier

1:40 p.m. **Brief Discussion with Speakers**

1:45 p.m. **What are the Critical Elements of a Successful Data Sharing Plan?**

ROBERT HANISCH
Director, Office of Data and Informatics
NIST

2:00 p.m. **Brief Discussion with Speakers**

2:05 p.m. **Working to Establish Best Practices for Data Sharing**

ELAINE MARTIN
Director and Chief Administrative Officer, Countway Library
Harvard University; RDMLA

2:20 p.m. **Panel Discussion with the Speakers**

2:35 p.m. **Break**

SESSION III. ENCOURAGING UPTAKE OF DATA SHARING IN THE SCIENTIFIC COMMUNITY

Moderator: Mark Hahnel, Figshare and Daniela Witten, University of Washington

Session Objective: Explore the needs of researchers with respect to implementing the new NIH data sharing policy, and understand what would encourage and promote policy adherence.

2:55 p.m. ***Moderated Panel Discussion***

Questions for the Panelists:

- What does a modern laboratory need (e.g., tools/infrastructure) in order to be prepared to implement the policy?
- What are the needs around training and education for laboratories to successfully implement the new data sharing policy?
- How can we ensure that there will be uptake of the new NIH policy in the scientific community?

Panelists:

TIMOTHY COETZEE
Chief Advocacy, Services, & Science Officer
National Multiple Sclerosis Society

SCOTT FRASER
Provost Professor and Director of Scientific Initiatives
University of Southern California

RICK GILMORE
Professor of Psychology
Penn State University

CAROLE GOBLE
Professor of Computer Science
University of Manchester

SARAH NUSSER
Professor of Statistics
Iowa State University

LETISHA WYATT
Assistant Professor of Neurology, School of Medicine
Director of Diversity in Research, OHSU Research & Innovation
Oregon Health Science University (OHSU)

3:35 p.m. **Q&A with the Speakers and Participants**

3:55 p.m. **Reflections on Day 1 and Preview of Day 2**

MARYANN MARTONE, *Workshop Planning Committee Co-Chair*
Professor, Neurosciences
UCSD

RICHARD NAKAMURA, *Workshop Planning Committee Co-Chair*
(Retired) Former Director, Center for Scientific Review
National Institutes of Health (NIH)

4:10 p.m. **Adjourn Workshop Day 1**

DAY 2: Thursday, April 29, 2021

11:00 a.m. ET **Welcome and Overview of Day 2**

MARYANN MARTONE, *Workshop Planning Committee Co-Chair*
Professor, Neurosciences
UCSD

RICHARD NAKAMURA, *Workshop Planning Committee Co-Chair*
(Retired) Former Director, Center for Scientific Review
National Institutes of Health (NIH)

SESSION IV. VALUE AND COSTS OF MANAGING AND SHARING DATA

Moderator: Christine Borgman, UCLA

Session Objective: To understand from a variety of perspectives what constitutes good practices to ensure that data is reproducible.

11:10 a.m. **Realizing the Value of Data Management from the Laboratory Side**

JOHN BORGHI
Manager of Research and Instruction
Lane Medical Library
Stanford University

ANA VAN GULICK
Government and Funder Lead
Figshare

11:25 a.m. **Data Quality and Other Factors that Make Data More Likely to be Reused**

RAFAEL IRIZARRY
Professor of Applied Statistics
Harvard University

DANIEL GOROFF
Vice President and Program Director
Alfred P. Sloan Foundation

IRENE PASQUETTO
Assistant Professor of Information
University of Michigan

11:55 a.m. **Questions for the Speakers**

12:05 p.m. **Reflections from the Field and Panel Discussion**

JAN BJAALIE
Professor, Infrastructure Development Director
University of Oslo and EU Human Brain Project

12:25 p.m. **Break**

SESSION V. SHAPING A CULTURE OF DATA SHARING – REDUCING BARRIERS AND INCREASING INCENTIVES

Session Objectives:

- Examine legal and ethical issues that can potentially create barriers to large-scale data sharing, including evolving privacy laws, informed consent, and ethical concerns.
- Understand opportunities and barriers to encouraging a culture of data sharing.
- Explore the impact of funder policies (other than sharing mandates) on widespread data sharing.

12:45 p.m. **Overview of Legal Issues around Data Sharing**

KRISTEN ROSATI
Attorney
Coppersmith Brockelman PLC
Past President, American Health Law Association

12:55 p.m. **Seeking “Informed” Consent for Large-Scale Data Sharing**

MARK ROTHSTEIN
Herbert F. Boehl Chair of Law and Medicine
Founding Director of the Institute for Bioethics, Health Policy and Law at the University of Louisville School of Medicine

1:05 p.m. **Creating Good Data Governance: The Evolving Role of the University as Gate Keeper and Cheerleader for Data Sharing**

CORA HAN
Chief Health Data Officer
UC Health, University of California

1:15 p.m. **Implementing the NIH Data Management and Sharing Policy: The Evolving Ethics of Data Sharing**

ANITA ALLEN
Henry R. Silverman Professor
University of Pennsylvania Law School

1:25 p.m. **Discussion/Q&A with the Speakers and Participants** (*Moderator: Kristen Rosati, Attorney, Coppersmith Brockelman PLC*)

1:45 p.m. **Encouraging Data Sharing Outside of Mandates**

NEIL THAKUR
Chief Mission Officer
ALS Association

ASHLEY FARLEY
Associated Officer, Knowledge & Research Services
Bill & Melinda Gates Foundation

MARYROSE FRANKO
Executive Director
Health Research Alliance (HRA)

2:15 p.m. **Discussion/Q&A with the Speakers and Participants** (*Moderator: Neil Thakur, ALS Association*)

2:35 p.m. **Break**

SESSION VI. IMPLEMENTING THE NIH DATA MANAGEMENT AND SHARING POLICY: ARE WE READY FOR 2023?

Session Objectives:

- Examine common themes and takeaways from the earlier workshop sessions.
- Explore possible challenges to the implementation of the NIH data management and sharing policy from a variety of stakeholder perspectives.
- Consider possible actions that can be taken to help different stakeholder groups prepare for and be successful with complying with the NIH data management and sharing policy.

2:55 p.m. **Multi-Stakeholder Panel with Select Speakers**

Moderator:

RICHARD NAKAMURA, *Workshop Planning Committee Co-Chair*
(Retired) Former Director, Center for Scientific Review
National Institutes of Health (NIH)

Panelists:

CHRISTINE BORGMAN
Distinguished Research Professor, Information Studies
UCLA

PHILIP BOURNE
Founding Dean, School of Data Science
Professor of Biomedical Engineering
University of Virginia

SUSANNA-ASSUNTA SANSONE
Associate Professor, Data Readiness
Associate Director, Oxford e-Research Centre

University of Oxford

MARGARET SUTHERLAND
Program Manager
Chan Zuckerberg Initiative

NEIL THAKUR
Chief Mission Officer
ALS Association

LETISHA WYATT
Assistant Professor of Neurology, School of Medicine
Director of Diversity in Research, OHSU Research & Innovation
Oregon Health Science University (OHSU)

3:45 p.m. **Reflections from the Workshop and Final Comments**

MARYANN MARTONE, *Workshop Planning Committee Co-Chair*
Professor, Neurosciences
UCSD

4:00 p.m. **Adjourn Workshop Day 2**

WORKSHOP INFORMATION

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy
Board on Health Care Services

Changing the Culture of Data Management and Sharing
A Workshop

April 28-29, 2021 (virtual)

[Register here](#)

In response to a request from the NIH Office of Science Policy, a planning committee of the National Academies of Sciences, Engineering, and Medicine will convene a two-day virtual public workshop to discuss the challenges and opportunities for researchers, institutions, and funders to establish effective data management and sharing practices. The objective of the workshop is to examine strategies, resources, and promising practices for developing and evaluating data management and sharing plans, as well as to discuss how researchers can effectively share scientific data over the course of the data life cycle.

Input will be sought from a variety of perspectives, including researchers, data repository managers, funding institutions, publishers, research participants, and other stakeholders to include a diversity of biomedical research fields and disciplines. With an emphasis on illustrative case studies, real world examples, and promising practices, potential topics may include:

- Addressing overarching strategies for managing and sharing data, taking into consideration diverse needs (e.g., human vs non-human data, type and size, data generators vs data users);
- Assessing value of shared data and the development and evaluation of data management and sharing plans, which may include discussions of:
 - best practices for repositories to collect the metrics needed to make such assessments,
 - the extent to which data value was anticipated and planned for prior to generating or sharing data, and
 - how this might inform prospective planning for sharing;
- Monitoring and evaluating data management and sharing practices, including discussion of appropriate metrics for timelines of data availability and life cycles of different types of scientific data; and
- Considering educational and other resource needs for responsible data sharing practices.

The workshop planning committee will develop the agenda for the workshop, select and invite speakers and discussants, and moderate or select moderators for the discussions. At the end of each workshop day, key points from individual speakers will be summarized and shared with the audience by a moderator. The workshop website will contain presentations from the speakers who provide the National Academies with permission to share their slides and the video of the webcast from the workshop.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

**Changing the Culture of Data Management and Sharing
A Workshop**

Planning Committee Roster

Co-CHAIRS

Maryann Martone, Ph.D.

Professor Emerita, Neurosciences
Chair of the Academic Senate Committee on
Academic Computing and Communications
University of California, San Diego

Richard Nakamura, Ph.D.

Retired,
Former Director of NIH Center for Scientific
Review

PLANNING COMMITTEE MEMBERS

Christine Borgman, Ph.D.

Professor and Presidential Chair in Information
Studies
University of California, Los Angeles

Elaine Mardis, Ph.D.

Co-Executive Director of the Institute for Genomic
Medicine, Steve and Cindy Rasmussen Endowed
Chair in Genomic Medicine
Nationwide Children Hospital
Professor of Pediatrics
The Ohio State University College of Medicine

Wouter Haak

Vice President
Research Data Management Solutions
Elsevier

Kristen Rosati, J.D.

Attorney
Coppersmith Brockelman PLC
Past President,
American Health Law Association

Mark Hahnel, Ph.D.

CEO and Founder
Figshare

Neil Thakur, Ph.D.

Chief Mission Officer
ALS Association

Nick Lindsay

Director for Journals and Open Access
MIT Press

Daniela Witten, Ph.D.

Professor of Statistics & Biostatistics
Dorothy Gilford Endowed Chair
University of Washington

Workshop Sponsor: U.S. National Institutes of Health; Office of Science Policy

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

Changing the Culture of Data Management and Sharing
A Workshop

April 28-29, 2021

Planning Committee Member Biographies

Maryann Martone, Ph.D., (co-chair) received her B.A. from Wellesley College in Biological Psychology and Ancient Greek and her Ph. D. in Neuroscience from the University of California, San Diego. She is a professor Emerita at UCSD, but still maintains an active laboratory and currently serves as the Chair of the University of California Academic Senate Committee on Academic Computing and Communications. She started her career as a neuroanatomist, specializing in light and electron microscopy, but her main research for the past 15 years focused on informatics for neuroscience, i.e., neuroinformatics. She led the Neuroscience Information Framework (NIF), a national project to establish a uniform resource description framework for neuroscience, and the NIDDK Information Network (dknet), a portal for connecting researchers in digestive, kidney and metabolic disease to data, tools, and materials. She just completed 5 years as Editor-in-Chief of Brain and Behavior, an open access journal, and has just launched a new journal as Editor in Chief, NeuroCommons, with BMC. Dr. Martone is past President of FORCE11, an organization dedicated to advancing scholarly communication and e-scholarship. She completed two years as the chair of the Council on Training, Science and Infrastructure for the International Neuroinformatics Coordinating Facility and is now the chair of the Governing Board. Since retiring, she served as the Director of Biological Sciences for Hypothesis, a technology non-profit developing an open annotation layer for the web (2015-2018) and founded SciCrunch, a technology start up based on technologies developed by NIF and dkNET.

Richard Nakamura, Ph.D., (co-chair) retired as Director of the NIH Center for Scientific Review in April 2018. He is now a volunteer at the Center, where he is working to help CSR complete and publish research related to CSR reviews. On December 3, 2012, Dr. Richard Nakamura was named Director of the Center for Scientific Review (CSR) at the National Institutes of Health (NIH). He leads CSR's 450 scientists and administrative staff, overseeing their efforts to manage 80,000 incoming NIH grant applications a year and review the majority of them in CSR peer review groups. Dr. Nakamura has had a 32-year tenure at the National Institute of Mental Health (NIMH), where he has served as both Scientific Director and Deputy Director of the institute, and he served as Acting Director from 2001 to 2002. During his time at NIMH, he received a number of leadership awards, including the Presidential Rank Award for outstanding leadership. He came to NIMH in 1976 as a postdoctoral fellow. In the mid-80's he coordinated NIMH's Biobehavioral Program and later was Chief of its Integrative Neuroscience Research Branch. Between 1997 and 2007, he served as the institute's Deputy Director. From 2007 to 2011 he has been institute Scientific Director. While at NIMH, he also has held other positions, including Associate Director for Science Policy and Program Planning; Chief, Behavioral and Integrative Neuroscience Research Branch; and Coordinator, ADAMHA Office of Animal Research Issues. Dr. Nakamura attended the Bronx High School of Science and earned his B.A. in psychology from Earlham College in Richmond, Indiana. He received his Ph.D. in psychology from the State University of New York in Stony Brook. Dr. Nakamura has expertise in a number of areas, including cognitive and comparative neuroscience, science policy/funding and ethics in science. He has published 30 peer reviewed scientific journal articles, most related to neurocognition in primates.

Christine Borgman, Ph.D., Distinguished Research Professor of Information Studies at UCLA, conducts research in scientific data practices and information policy. She is the author of more than 250

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

publications in information studies, computer science, communication, and law, which include three books from MIT Press: *Big Data, Little Data, No Data: Scholarship in the Networked World* (2015), winner of the 2015 American Publishers Award for Professional and Scholarly Excellence (PROSE Award) in Computing and Information Sciences; *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* (2007); and *From Gutenberg to the Global Information Infrastructure: Access to Information in a Networked World* (2000). The latter two books won the Best Information Science Book of the Year award from the Association for Information Science and Technology. Professor Borgman is a member of the Library of Congress Scholars Council; a member of the advisory board of the Electronic Privacy Information Center; member of the CLARIAH International Advisory Panel; and is a Fellow of the American Association for the Advancement of Science and of the Association for Computing Machinery. Other honors and awards include the Paul Evan Peters Award from the Coalition for Networked Information, Association for Research Libraries, and EDUCAUSE; Award of Merit and the Research in Information Science Award, both from the Association for Information Science and Technology; and a Legacy Laureate of the University of Pittsburgh. She has keynoted conferences and events in the sciences, social sciences, computer science, data science, medicine, law, and the humanities. Professor Borgman also holds the title of University of California Presidential Chair in Information Studies, Emerita.

Wouter Haak is responsible for research data management at Elsevier, specifically the [Mendeley Data](#) platform. This is an open ecosystem of researcher data tools: a data repository, an electronic lab notebook, a data search tool, and a data project management tool. Aside from his work for Elsevier, Wouter is part of several open data community initiatives; for example he co-chairs the RDA-WDS Scholix working group on data-article linking; he is part of the JISC Data2paper advisory board; and his group participates in the NIH Data Commons pilot project. Prior to Elsevier, Wouter worked in online product and strategy roles. He has worked at eBay Classifieds, e.g. Marktplaats.nl, Kijiji.it – in roles varying from business development to overall responsibility for the classified's businesses in Italy, France, Belgium and Turkey. Furthermore, he has worked for the Boston Consulting Group.

Mark Hahnel, Ph.D., is the CEO and founder of Figshare, which he created whilst completing his PhD in stem cell biology at Imperial College London. Figshare currently provides research data infrastructure for institutions, publishers and funders globally. He is passionate about open science and the potential it has to revolutionize the research community. For the last eight years, Mark has been leading the development of research data infrastructure, with the core aim of reusable and interoperable academic data. Mark sits on the board of DataCite and the advisory board for Directory of Open Access Journals (DOAJ). He was on the judging panel for the National Institutes of Health (NIH), Wellcome Trust Open Science prize and acted as an advisor for the Springer Nature master classes.

Nick Lindsay is the Director for Journals and Open Access at MIT Press where he spends most of his time working on client relations, new journal acquisitions and strategic partnerships, and other outward looking activities. In addition to his work at MIT Press, Mr. Lindsay is a Publications Committee Member in the Association for the Sciences of Limnology and Oceanography as well as a former member of the Board of Directors of the Open Access Network. Mr. Lindsay received a B.A. in English from Dalhousie University in 1992 and a certificate from the Summer Publishing Institute in NYU in 2000. Prior to his current position, he has also served as the Journals Marketing and Circulation Manager at University of California Press.

Elaine Mardis, Ph.D., is co-Executive Director of the Institute for Genomic Medicine at Nationwide Children's Hospital and the Steve and Cindy Rasmussen Endowed Chair in Genomic Medicine. She also

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

is Professor of Pediatrics at The Ohio State University College of Medicine. Dr. Mardis joined Nationwide Children's Hospital in 2016. She was educated at the University of Oklahoma with a B.S. in Zoology and a Ph.D. in Chemistry and Biochemistry, Dr. Mardis did postgraduate work in industry at BioRad Laboratories. She was a member of the faculty of Washington University School of Medicine from 1993-2016. Dr. Mardis has authored over 350 articles in prestigious peer-reviewed journals and has written book chapters for several medical textbooks. She serves as an associate editor for three peer-reviewed journals (Disease Models and Mechanisms, Molecular Cancer Research, and Annals of Oncology) and is Editor-in-Chief of Molecular Case Studies, published by Cold Spring Harbor Press. Dr. Mardis has given lectures at scientific meetings worldwide, and was awarded the Morton K Schwartz award from the American Association for Clinical Chemistry in 2016. She has been listed since 2013 as one of the most highly cited researchers in the world by Thompson Reuters. Dr. Mardis has been a member of the American Association for Cancer Research (AACR) since 2007, was the program committee chair for the 2018 AACR Annual Meeting, and is the AACR President-elect.

Kristen Rosati, J.D., is considered one of the nation's leading "Big Data" and HIPAA compliance attorneys. She also has deep experience in data breaches, health information exchange, data sharing for research and clinical integration initiatives, clinical research compliance, and biobanking and genomic privacy. Ms. Rosati is a sought-after national speaker on these issues and has been active in national healthcare policy. She is Past President (2013-2014) of the American Health Law Association (AHLA), the nation's largest health care legal organization. Ms. Rosati received her B.A. and J.D. from the University of Michigan in 1987 and 1990, respectively.

Neil Thakur, Ph.D., has more than two decades of experience as a public health expert. He has led The ALS Association's mission programs – research, care services, and advocacy – since 2018. Prior to joining the Association, Dr. Thakur served in the National Institutes of Health (NIH) Office of the Director, where he supported NIH governance and helped make NIH research more open and less burdensome. He managed the world's largest policy to make biomedical research papers publicly accessible and co-chaired the White House taskforce that lead to the requirement that all federal science agencies adopt similar policies. He also spent a year on detail to the US Senate Special Committee on Aging, raising awareness about quality issues in long-term health care, particularly around Alzheimer's care and pharmaceuticals. Prior to his time at NIH, Dr. Thakur worked with health systems in many capacities. He was Assistant Director of Health Services Research and Development at the Department of Veterans Affairs (VA), leading an evaluation service for the VA health system and represented the VA research service in setting clinical performance measures. In his post-doctoral-fellowship, he studied the interactions between jails, Medicaid and behavioral health care, and how changes in health financing impacted people's utilization of these systems. During graduate school, he worked throughout the Connecticut behavioral health system, helping to implement managed care and health information systems, and raise tens of millions of dollars in competitive grants. Dr. Thakur won many awards for his government service, including several NIH Director's Awards, and the Secretary for Health and Human Services' award for Meritorious Service, the second highest award that the Secretary can bestow. He holds a Ph.D. in Health Policy from Yale University School of Public Health and completed a NIMH postdoctoral fellowship in mental health services research at the Cecil G. Sheps Center for Health Services Research at the University of North Carolina at Chapel Hill.

Daniela Witten, Ph.D., has done research involving the development of statistical machine learning methods for high dimensional data, with applications to genomics and other fields. Dr. Witten is a co-

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

author (with Gareth James, Trevor Hastie, and Rob Tibshirani) of the very popular textbook "Introduction to Statistical Learning". She is the recipient of a number of honors, including an NIH Director's Early Independence Award, a Sloan Research Fellowship, an NSF CAREER Award, and a Simons Investigator Award. Her work has been featured in the popular media: among other forums, in Forbes Magazine (three times), Elle Magazine, on KUOW radio, and as a PopTech Science Fellow. Dr. Witten completed a B.S. in Math and Biology with Honors and Distinction at Stanford University in 2005, and a Ph.D. in Statistics at Stanford University in 2010. Since 2018, Dr. Witten is the Dorothy Gilford Endowed Chair in Mathematical Statistics and a Professor of Statistics and Biostatistics at University of Washington.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

Changing the Culture of Data Management and Sharing
A Workshop

April 28-29, 2021

Speaker Biographies

Anita L. Allen, J.D., Ph.D., is the Henry R. Silverman Professor of Law and Professor of Philosophy. A graduate of Harvard Law School with a PhD from the University of Michigan in Philosophy, Allen is internationally renown as an expert on philosophical dimensions of privacy and data protection law, ethics, bioethics, legal philosophy, women's rights, and diversity in higher education. She was Penn's Vice Provost for Faculty from 2013-2020, and chaired the Provost's Arts Advisory Council. Allen is an elected member of the National Academy of Medicine, the American Law Institute and a fellow of the American Academy of Arts and Sciences. In 2018-19 she served as President of the Eastern Division of the American Philosophical Association. From 2010 to 2017, Allen served on President Obama's Presidential Commission for the Study of Bioethical Issues. She was presented the Lifetime Achievement Award of the Electronic Privacy Information Center in 2015, and chaired its Board, 2019-2022. Allen has served on the faculty of the School of Criticism and Theory at Cornell, for which she is an advisor. A two-year term as an Associate of the Johns Hopkins Humanities Center concluded in 2018. She has been a visiting Professor at Tel Aviv University, Waseda University, Villanova, the University of Arizona, Harvard and Yale, and a Law and Public Affairs Fellow at Princeton. She was awarded an honorary Doctorate from Tilburg University (Netherlands) in 2019. Allen has given hundreds of talks all over the world and appeared on television, radio and written for major media. She currently serves on the Board of the National Constitution Center, and has served on numerous other boards and professional advisory boards, including the Pennsylvania Board of Continuing Judicial Education, the Association for Practical and Professional Ethics, the Bazelon Center for Mental Health Law, the AALS Executive Committee, the Maternity Care Coalition and the West Philadelphia Alliance for Children. She is a member of the Pennsylvania and New York bars, and formerly taught at Georgetown University Law Center for ten years and the University of Pittsburgh, after practicing briefly at Carvath, Swaine & Moore.

Jan Bjaalie, M.D., Ph.D., is Professor at the Institute of Basic Medical Sciences, University of Oslo, Norway, Infrastructure Director of the EU Human Brain Project, leader of the EBRAINS Data services, and Head of the Norwegian Neuroinformatics Node. He received his M.D. and Ph.D. degrees in 1986 and 1990, respectively, and was appointed Associate Professor at the University of Oslo in 1992 and full Professor in 1997. His research group joined the Centre for Molecular Biology and Neuroscience, a Norwegian Centre of Excellence appointed by The Research Council of Norway, in 2002. His laboratory has discovered fundamental principles of sensory map transformations in large projection systems of the brain and has performed novel regional and whole brain atlasing and histological mapping. The laboratory has used 'gold standard' anatomical methods in combination with computerized methods for visualization and quantitative analyses, electrophysiology, and *in vivo* imaging, for studying systems level organization in the brain. In the context of this research, the group has developed tools for 3-D reconstruction and advanced visualization of neuronal organization, database applications for neuroanatomical image data, and digital brain atlases. The tools and databases of the laboratory are made available via the Rodent Brain Workbench, <http://rbwb.org>, and recently through the EBRAINS web portal, <https://ebrains.eu>. Professor Bjaalie has been partner and coordinator of several EU funded

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

projects and has collaborated extensively with leading laboratories in many countries. He is a member of the editorial board of several journals. He has served as co-Chair of the International Brain Initiative since 2019, Head of the Institute of Basic Medical Sciences at the University of Oslo (2009 - 2016), Chair of the Governing Board (2013 - 2015) and founding Executive Director (2006 - 2008) of the International Neuroinformatics Coordinating Facility (INCF), and member of the Neuroinformatics Committee of the Society for Neuroscience (2004 - 2009).

John Borghi, Ph.D., is the manager of research and instruction at the Lane Medical Library at Stanford University, where he coordinates much of the library's outreach and education-related activities. A cognitive neuroscientist by training, he is focused on making scientific research more accessible and transparent.

Philip E. Bourne, Ph.D., leads a range of initiatives to encourage and facilitate the use of big data in large-scale research across the scientific and technological disciplines, with special emphasis on structural bioinformatics and systems pharmacology. He is the Founding Dean of the School of Data Science and Professor of Biomedical Engineering. From 2014-2017, Phil was the Associate Director for Data Science at the National Institutes of Health. In this role he led the Big Data to Knowledge Program, coordinating access to and analyzing biomedical research from across the globe and making it available to scientists and researchers. While there, he was also responsible for governance and strategic planning activities for data and knowledge management, and established multiple trainings in data science. He has done exceptional work to make biomedical research accessible, as well as to advance the field of data science. Prior to his time at the NIH, Phil spent 20 years on the faculty at the University of California-San Diego, eventually becoming Associate Vice Chancellor of Innovation and Industrial Alliances. He is a highly respected and oft-cited scholar who brings a wealth of experience to UVA.

Patricia Brennan, RN, Ph.D., is the Director of the National Library of Medicine (NLM) at the National Institutes of Health (NIH), where she oversees the world's largest biomedical library. She has positioned the Library to be the hub for biomedical data science at NIH and across the globe. Dr. Brennan holds an appointment as associate investigator in the National Institute of Nursing Research Division of Intramural Research, where she directs the Advanced Visualization Laboratory. Before joining NIH, she was the Lillian L. Moehlman Bascom Professor in the School of Nursing and College of Engineering at the University of Wisconsin-Madison. Dr. Brennan is a member of the National Academy of Medicine. She is a fellow of American Institute for Medical and Biological Engineering (AIMBE), the American Academy of Nursing, the American College of Medical Informatics, and the New York Academy of Medicine.

Atul Butte, M.D., Ph.D., is the Priscilla Chan and Mark Zuckerberg Distinguished Professor and inaugural Director of the Bakar Computational Health Sciences Institute (bchsi.ucsf.edu) at the University of California, San Francisco (UCSF). Dr. Butte is also the Chief Data Scientist for the entire University of California Health System, with 20 health professional schools, 6 medical schools, 5 academic medical centers, 10 hospitals, and over 1000 care delivery sites. Dr. Butte has been continually funded by NIH for 20 years, is an inventor on 24 patents, and has authored over 200 publications, with research repeatedly featured in the New York Times, Wall Street Journal, and Wired Magazine. Dr. Butte was elected into the National Academy of Medicine in 2015, and in 2013, he was recognized by the Obama Administration as a White House Champion of Change in Open Science for promoting science through publicly available data. Dr. Butte is also a founder of three investor-backed data-driven companies: Personalis (IPO, 2019),

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

providing medical genome sequencing services, Carmenta (acquired by Progenity, 2015), discovering diagnostics for pregnancy complications, and NuMedii, finding new uses for drugs through open molecular data. Dr. Butte trained in Computer Science at Brown University, worked as a software engineer at Apple and Microsoft, received his MD at Brown University, trained in Pediatrics and Pediatric Endocrinology at Children's Hospital Boston, then received his PhD from Harvard Medical School and MIT.

Tim Coetze, Ph.D., serves as the National MS Society's Chief Advocacy, Services and Science Officer. In this capacity, he leads the Society's work in the areas of state and federal advocacy, delivery of services and connection programs for people with MS, healthcare professional engagement and training, as well as the Society's global research programs. Most recently, he served as the President of Fast Forward, a venture philanthropy of the National Multiple Sclerosis Society where he was responsible for strategic funding of biotechnology and pharmaceutical companies as well as partnerships with the financial and business communities. Prior to Fast Forward, he led the Society's global research initiatives on nervous system repair and protection in multiple sclerosis as well as the Society's fellowship and faculty award programs. He is a member of the Society's CEO Leadership team, the International Progressive MS Alliance's Scientific Steering Committee and the International Advisory Committee on Clinical Trials in MS. In addition, Tim serves on the National Academy of Medicine's Forum on Neuroscience and Nervous System Disorders, and co-chairs the National Academy of Medicine's Forum on Regenerative Medicine. Prior to joining the Society, Tim held faculty appointments at the University of Connecticut Health Sciences Center where he conducted research into the structure and function of myelin. Tim received his Ph.D. in molecular biology from Albany Medical College in 1993 and has since been involved in the field of multiple sclerosis research. He has been with the National MS Society since the fall of 2000.

Ashley Farley is a Program Officer of Knowledge and Research Services at the Bill and Melinda Gates Foundation. In this capacity, she focuses on the foundation's Open Access Policy's implementation and associated initiatives. This includes leading the work of Gates Open Research, a transparent and revolutionary publishing platform. Other core activities involve supporting the strategic and operational aspects of the foundation's library. This work has sparked a passion for open access, believing that freely accessible knowledge has the power to improve and save lives.

Adam Ferguson, Ph.D., is an associate professor of Neurological Surgery at University of California, San Francisco (UCSF) and a principal investigator in the Brain and Spinal Injury Center (BASIC) at San Francisco General Hospital (SFGH). His research focuses on the mechanisms of recovery after neurological trauma. Injuries to the brain and spinal cord invoke a number of complex biological processes that work in concert to determine the extent of tissue repair and functional recovery. To further complicate matters, some biological processes have contradictory effects when present at different stages of neurological recovery. For example, mechanisms of synaptic regulation can contribute to cell death in the early phases of recovery but may promote plasticity and restoration of the function at later stages. Understanding the mechanisms of repair in the complex microenvironment of the injured central nervous system (CNS) requires a large-scale integration of complex biological information and functional outcomes. Dr. Ferguson's work uses a combination of molecular and cell biology, behavioral neuroscience, and statistical modeling to provide an information-rich picture of the holistic syndrome produced by CNS trauma in translational models. The long-term goal of this research is to provide system-level therapeutic targets for enhancing recovery of function after brain and spinal injury.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

Maryrose Franko, Ph.D., is Executive Director of the Health Research Alliance. Working closely with the organization's board, she sets its strategic priorities, advances its members' objectives through key programs, and develops tools for the nonprofit biomedical community. Dr. Franko's background includes over 20 years of program management at the Howard Hughes Medical Institute (HHMI), including strategic planning as well as creating, implementing, and managing over a dozen programs and initiatives. These include graduate, medical student and postdoctoral research fellowships, and an innovative and groundbreaking joint initiative with the National Institute of Biomedical Imaging and Bioengineering at the National Institutes of Health. Dr. Franko also ran both the graduate and undergraduate programs at HHMI's state-of-the-art research facility, Janelia Research Campus, and created professional development opportunities for Janelia's postdocs. Dr. Franko's collaboration with the Burroughs Wellcome fund to develop a residential Lab Leadership and Management course, led to the creation of Making the Right Moves: A Practical Guide to Scientific Management for Postdocs and New Faculty, and the companion guide, Training Scientists to Make the Right Moves, which were joint efforts of HHMI and Burroughs Wellcome Fund. Dr. Franko received her PhD in molecular genetics from University of Southern California and did a post-doctoral fellowship at the National Institutes of Health before joining HHMI. During her time at HHMI, Franko was a founding board member of HRA, serving from 1995 to 2012. While a member of HRA, she initiated and led the Early Career Scientist Working group which is now the Research Workforce and Early Career Development working group. She serves on many boards, including the Center for Open Science, and Northern Virginia's new interactive science museum – the Children's Science Center.

Scott E. Fraser, Ph.D., is a biophysicist and Provost Professor of Biological Sciences and Biomedical Engineering at the University of Southern California (USC). He is also the Elizabeth Garrett Chair in Convergent Bioscience and Director of Science Initiatives, where he is helping to launch USC's Initiative in Convergent Bioscience. In addition, he holds joint appointments in the Departments of Physiology and Biophysics, Stem Cell Biology and Regenerative Medicine, Pediatrics, Radiology, and Ophthalmology. Fraser and his colleagues are known for their development of light and magnetic resonance imaging (MRI) microscopy techniques for imaging the dynamics of embryonic development. More recently his research team has taken these imaging techniques into disease models and clinical medicine, in areas ranging from eye disease to cancer. Fraser began his scientific career studying Physics (B.S. with honors, Harvey Mudd College, 1976) and Biophysics (Ph.D. with distinction, Johns Hopkins University, 1979) before joining the faculty at the University of California, Irvine in 1980, where he eventually became Chairman of the Department of Physiology and Biophysics. In 1990, Fraser moved to the California Institute of Technology (Caltech) to serve as the Anna L. Rosen Professor of Biology, Professor of Engineering and Applied Science, and the Director of the Biological Imaging Center at the Beckman Institute. He was also the Founding Director of the Caltech Brain Imaging Center from 2002 to 2008, a founding member of the Kavli Nanoscience Institute, and served as the Director of the Rosen Center for Biological Engineering from 2008 to 2012. In 2012, Fraser moved to USC to take a Provost Professorship in the USC Dornsife College of Letters, Arts and Sciences, the Children's Hospital Los Angeles, Keck School of Medicine, and the Viterbi School of Engineering.

Rick Gilmore, Ph.D., is Professor of Psychology at Penn State, where he studies the development of perception and action using computational, neuroscience, and behavioral methods. He is the co-founder and co-director of Databrary.org, a data library specialized for storing and sharing video, audio, and

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

associated sensitive and identifiable data. Gilmore earned his bachelor's degree magna cum laude from Brown University, and his Ph.D. from Carnegie Mellon University.

Carole Goble CBE, FREng., is a Full Professor in the School of Computer Science where she leads a team of researchers and software developers building e-infrastructure for researchers. She applies technical advances in knowledge technologies, distributed computing, workflows and social computing to solve information management problems for Life Scientists, Biodiversity, Chemistry, and Health informatics. Her research interests include: reproducible research, computational workflows, semantic interoperability, knowledge exchange between scientists and new models of scholarly communication. She is a co-founder of the UK's Software Sustainability Institute. Carole is one of the many authors of the influential FAIR Data Principles Nature paper, is active in national and European policy making for data management and data sharing, and currently serves as the UK expert representative on the G7 Working Group on Open Science. She is deeply involved with several pan-European Research Infrastructures (RIs) for Life Sciences and Biodiversity part of the European Open Science Cloud. She is Head of the UK Node of the ELIXIR RI for Life Science Data, co-leads the ELIXIR Interoperability work stream, and co-leads the RDMkit – a toolkit for Research Data Management – and the WorkflowHub.eu, a registry for sharing computational workflows. FAIRDOM.org is a pan-national initiative she has coordinated for over a decade which aims to support projects manage and share their Research Objects – its platform underpins WorkflowHub and over 140 other Hubs including the Leipzig Health Atlas. She co-leads the researchobject.org community who have devised a web-smart approach to exchange Research Objects between infrastructure platforms. In 2008 she was awarded the Microsoft Jim Gray award for outstanding contributions to e-Science.

Daniel Goroff, M.A., M.Phil., Ph.D., is Vice President and Program Director at the Alfred P. Sloan Foundation, a private charity that supports breakthroughs in science, technology, and economics. He is currently on temporary and part-time loan to the National Science Foundation (NSF) serving as Division Director for Social and Economic Sciences. He is also Professor Emeritus of Mathematics and Economics at Harvey Mudd College in Claremont, where he served as Dean of the Faculty and Vice President for Academic Affairs. During twenty years before that at Harvard University, he rose in rank from Assistant Professor of Mathematics to Professor of the Practice and Associate Director of the Derek Bok Center for Teaching and Learning. A winner of the Phi Beta Kappa Award—the highest recognition for educational excellence in Harvard's Faculty of Arts and Science—Goroff not only developed and taught courses in mathematics, but also in physics, economics, engineering, and history of science, as well as a pioneering course on “Decisions, Games, and Negotiations” that was popular online, too. The Masters Program he founded and directed at Harvard on “Mathematics for Teaching” still enrolls dozens of degree candidates each year. Daniel Goroff's research interests include optimization over time, decision-making under uncertainty, the mathematics of privacy, and the economics of science. He has held extended visiting positions at Bell Laboratories in New Jersey, the Institut des Hautes Études Scientifiques in Paris, the Mathematical Sciences Research Institute in Berkeley, the Dibner Institute at MIT, Columbia University's Teachers College, and the Bellagio Residency Program for Academic Writing in Italy. Books he edited include one on Science and Engineering Careers with Richard Freeman and another three-volume translation with an extended introduction for Henri Poincaré's *Les Méthodes Nouvelles de la Mécanique Céleste*. Daniel Goroff earned an B.A.-M.A. summa cum laude in Mathematics as a Borden Scholar at Harvard in 1978, an M.Phil. in Economics as a Churchill Scholar at Cambridge University in 1979, a Masters in Mathematical Finance as an HMC Scholar at Boston University in 2008, a Ph.D. in

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

Mathematics at Princeton University as a Danforth Fellow in 1984, and completed an Executive Education Program for Nonprofit Leaders at Stanford’s Graduate School of Business in 2013.

Cora Han, J.D., is the Chief Health Data Officer at the University of California, where she has focused her work on implementing strategies for leveraging health data in a responsible and innovative way. Ms. Han joined UC Health from the Federal Trade Commission’s Division of Privacy and Identity Protection where she played a leading role on health privacy matters for the Commission in both the enforcement and policy arenas. Ms. Han is a graduate of the University of Chicago Law School.

Robert J. Hanisch, Ph.D., is the Director of the Office of Data and Informatics, Material Measurement Laboratory, at the National Institute of Standards and Technology in Gaithersburg, Maryland. He is responsible for improving data management and analysis practices and helping to assure compliance with national directives on open data access. Prior to coming to NIST in 2014, Dr. Hanisch was a Senior Scientist at the Space Telescope Science Institute, Baltimore, Maryland, and was the Director of the US Virtual Astronomical Observatory. For more than twenty-five years Dr. Hanisch led efforts in the astronomy community to improve the accessibility and interoperability of data archives and catalogs.

David Haussler, Ph.D., is a professor of biomolecular engineering at University of California, Santa Cruz, an investigator at the Howard Hughes Medical Institute, scientific director of the UC Santa Cruz Genomics Institute, scientific co-director of the California Institute for Quantitative Biosciences, and a consulting professor at Stanford University School of Medicine and the UC San Francisco Biopharmaceutical Sciences Department. David Haussler develops innovative methods in computational genomics to accelerate our understanding of molecular function, evolution, and disease process. Working at the interface of mathematics, computer science, and molecular biology, Haussler and his team integrate cross-species comparative and high-throughput genomics data to study gene structure, mechanism, and regulation. They also collaborate with researchers across the country to discover molecular causes of neurodevelopmental diseases and cancer. Haussler is empowering biomedical researchers to generate, access, and share genomic data, ensuring that they can quickly test new ideas and build on discoveries from other labs.

Rafael Irizarry, Ph.D., received his Bachelor’s in Mathematics in 1993 from the University of Puerto Rico and went on to receive a Ph.D. in Statistics in 1998 from the University of California, Berkeley. His thesis work was on Statistical Models for Music Sound Signals. He joined the faculty of the Department of Biostatistics in the Johns Hopkins Bloomberg School of Public Health in 1998 and was promoted to Professor in 2007. He is now Professor of Biostatistics and Computational Biology at the Dana-Farber Cancer Institute and a Professor of Biostatistics at Harvard School of Public Health. Since 1999, Rafael Irizarry’s work has focused on Genomics and Computational Biology problems. In particular, he has worked on the analysis and signal processing of microarray, next-generation sequencing, and genomic data. He is currently interested in leveraging his knowledge in translational work, e.g. developing diagnostic tools and discovering biomarkers.

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

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

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.

Rebecca Koskela, M.S., is the Executive Director of Research Data Alliance US. Prior to this, she was the Executive Director of DataONE at the University of New Mexico. Rebecca has a background in both bioinformatics and high-performance computing. Prior to this position, Rebecca was the Life Sciences Informatics Manager for Alaska INBRE and the Biostatistics and Epidemiology Core Manager for the Center for Alaska Native Health Research at the University of Alaska Fairbanks. At the Arctic Region Supercomputing Center she led the team evaluating the use of the semantic web, including focused ontologies for use as data integration tool for biological data including microarray data, molecular pathways, and protein profiling data at the University of Alaska Fairbanks. Prior to that, she was a member of the senior management team of the Aventis Cambridge Genomics Center and manager of the Scientific Computing group in Cambridge, MA responsible for computing infrastructure, both hardware and software, for the global functional genomics organization including support for high-throughput sequencing and transcriptional profiling, molecular pathway analyses, and data integration. She was a bioinformatics specialist at the Mayo Clinic, director of informatics in the Department of Genetics at Stanford University, and also worked at Cold Spring Harbor Laboratory with the Dana Consortium for the Genetic Basis of Manic-Depression Illness. In addition to her bioinformatics experience, Rebecca specialized in system performance and analysis at Sandia National Laboratories, Los Alamos National Laboratory, Cray Research, Intel, and IBM.

Daniella Lowenberg is based at the California Digital Library, a part of the University of California Office of the President, where she leads services and initiatives around open data publishing. She is the Director and Principal Investigator for Make Data Count (<https://makedatacount.org>), a global initiative focused on the development of responsible research data assessment metrics. She is also the Product Manager for Dryad (<https://datadryad.org>), and founding chair of the FORCE11 Research Data Publishing Ethics working group. Prior to joining UC, Daniella worked in open access publishing where she implemented and oversaw the PLOS Open Data Policy as well as researched pharmacogenomics pathways at Stanford.

Lara Mangravite, Ph.D., is president of Sage Bionetworks, an organization that focuses on open practices to advance biomedicine through data-driven science and digital research. Recognizing that all research is limited by restrictions placed on the distribution of information, Sage works closely with institutes, foundations, and research communities to improve information flow, benchmark research practices, and establish research outcomes of sufficient confidence to support translation. Dr. Mangravite obtained a BS in physics from Pennsylvania State University, a PhD in pharmaceutical chemistry from the University of California, San Francisco, and a postdoctoral fellowship in cardiovascular pharmacogenomics at the Children's Hospital Oakland Research Institute.

Elaine R. Martin, MSLS, DA, joined the Countway Library as the Director and Chief Administrative Officer in 2016. Under her direction she oversees and manages a complex organization with one of the largest collections of both current medical research materials and historical and rare collections in the

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

world, holding more than 630,000 volumes. The Countway Library serves both academic and practicing physicians at Harvard Medical School, the Harvard T.H. Chan School of Public Health, the Harvard School of Dental Medicine, and the Massachusetts Medical Society. At Countway, Elaine is responsible for providing leadership in strategic planning, development, and the promotion of library resources and services. She is a strong advocate for the role of the Library in Research Data Management. She is currently serving as the co-leader of an Elsevier-funded project to create a research data management library academy which is a collaboration project to develop advanced data literacy for Librarians and Researchers throughout the world. Elaine is a Fellow of the Medical Library Association and she obtained her Doctorate in Library Information Science from Simmons University.

Sarah M. Nusser, Ph.D., is professor of statistics and affiliated with the Center for Survey Statistics and Methodology (CSSM) at Iowa State University. She is currently visiting professor at University of Virginia’s Social and Decision Analytics Division of the Biocomplexity Institute and senior research fellow with the Association of American Universities (AAU). Dr. Nusser serves as chair of the National Academies Board on Research Data and Information, is a member of the Committee on Transparency and Reproducibility in Federal Statistics, and recently completed her service on the Committee on National Statistics. Dr. Nusser previously served as vice president for research at Iowa State University from 2014 to 2020. In that capacity, she served as co-chair the AAU-APLU Public Access Working Group, chair of the Association of Public and Land-grant Universities (APLU) Council on Research, and currently serves on the project team and chairs the steering committee of the AAU-APLU Accelerating Public Access to Research Data (APARD) initiative. Dr. Nusser’s current research focuses on improving the reusability and impact of publicly accessible research data. Previous research activities involved survey statistics and methodology for land-based and human population surveys, supported by the U.S. Department of Agriculture, National Science Foundation, National Institutes of Health, and federal statistical agencies such as the U.S. Census Bureau and Bureau of Labor Statistics. Dr. Nusser twice served as director of the Center for Survey Statistics and Methodology and is fellow of the American Statistical Association, elected member of the International Statistical Institute, and has served on numerous scientific panels, advisory committees and governing boards. Dr. Nusser received a B.S. in botany from the University of Wisconsin-Madison, M.S. in botany from North Carolina State University, and Ph.D. in statistics from Iowa State University.

Irene Pasquetto, Ph.D., is a scholar in the field of information and communication science. She holds a position as Assistant Professor at the University of Michigan School of Information where she teaches “Information Ethics” and “Digital Curation.” Her most recent research work focuses on issues of scientific mis- and disinformation, open science practices, and public understanding and reuse of science products. From 2018 to 2020, she was a postdoctoral fellow at the Shorenstein Center on Media, Politics, and Public Policy, at the Harvard Kennedy School. At the Kennedy School, Irene co-founded and led for two years the Harvard Kennedy School Misinformation Review. Irene earned a Ph.D. in Information Studies from the University of California, Los Angeles (UCLA), where she also worked as a research assistant at the UCLA Center for Knowledge Infrastructures (CKI) and the UCLA Institute for Society and Genetics. Previously, Irene earned a master’s and a bachelor’s degree from the University of Verona (Italy).

Russell A. Poldrack, Ph.D., is the Albert Ray Lang Professor in the Department of Psychology and Professor (by courtesy) of Computer Science at Stanford University, and Director of the Stanford Center for Reproducible Neuroscience. His research uses neuroimaging to understand the brain systems underlying decision making and executive function. His lab is also engaged in the development of neuroinformatics tools to help improve the reproducibility and transparency of neuroscience, including the Openneuro.org and Neurovault.org data sharing projects and the Cognitive Atlas ontology.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

Alexander Ropelewski cultivated his 30+ year professional career at the Pittsburgh Supercomputing Center where he directs the Biomedical Applications Group, a group focused on enhancing the use of High-Performance Computing, Networking, and Data Science within the Biomedical Research Community. A computer scientist graduate from the University of Pittsburgh, Ropelewski's HPC work includes the creation of parallel codes on a wide-range of computing architectures and major contributions to architectural frameworks for data-intensive projects. Ropelewski is currently PI and Operations Director for the Brain Image Library (BIL), an NIH funded national public resource enabling researchers to deposit, analyze, mine, share and interact with large brain image datasets. Other data intensive projects Ropelewski currently contributes to include the AUROA-US Breast Cancer Data Coordinating Center and the Infrastructure and Engagement component of the National Institutes of Health HuBMAP project. In addition to those data intensive projects, Ropelewski co-directs the training and dissemination components of the National Center for Multiscale Modeling of Biological Systems. In the recent past, he led the PSC's NIH funded MARC program, a multi-institutional collaborative bioinformatics training effort involving scientists and educators at several Minority Serving Institutions.

Mark A. Rothstein, J.D., is the Herbert F. Boehl Chair of Law and Medicine and Director of the Institute for Bioethics, Health Policy, and Law at the University of Louisville School of Medicine. He received a B.A. from the University of Pittsburgh and a J.D. from Georgetown University. Professor Rothstein has concentrated his research on health privacy, research ethics, genetics, and public health. He is a past president of the American Society of Law, Medicine and Ethics, an elected member of the American Law Institute, and an elected fellow of the Hastings Center. From 1999-2008, he served as Chair of the Subcommittee on Privacy and Confidentiality of the National Committee on Vital and Health Statistics (NCVHS), the statutory federal advisory committee to the Secretary of Health and Human Services and Congress on health information policy. From 2011-2019, he was Associate Editor for Public Health Ethics and Law of the *American Journal of Public Health*. Since 2000, he has written a regular column on bioethics for the *Journal of Law, Medicine and Ethics*. Professor Rothstein is the author or editor of 19 books and over 300 book chapters and articles.

Susanna Sansone, M.Sc., DIC., Ph.D., works in the areas of data interoperability and reproducibility, research integrity, and the evolution of scholarly publishing, collaborating with researchers, service providers, journal publishers, library science experts, funders and learned societies in academic, commercial and government settings alike. Her Data Readiness group at the University of Oxford researches and develops new methods and tools to make digital research objects FAIR.

Margaret Sutherland, Ph.D., is currently a Program Manager at the Chan Zuckerberg Initiative. Previously, she has served as Program Director of the National Institute of Neurological Disorders and Stroke where she managed NIH extramural activities including program planning, evaluation, review assessment, and monitoring of major basic, clinical and/or applied research supported by the NINDS that have national and international scope and impact. Earlier in her career, she has also served as an assistant professors at the Children's National Medical Center, George Washington University, and at Vanderbilt University. Dr. Sutherland received a Ph.D. in molecular neuroscience at The Open University in 1993.

Ana Van Gulick, Ph.D., is the Government and Funder Lead at Figshare where she manages research repository projects for clients including US Federal agencies and nonprofit research funders as well as leading Figshare's data curation services. She also holds a visiting faculty appointment at the Carnegie Mellon University Libraries, where was a faculty member for 4 years prior to joining Figshare in 2020. She received a PhD in Psychology and Cognitive Neuroscience from Vanderbilt University in 2014 and has worked on data management, data sharing, and open science practices and infrastructure for the past 7 years including from the perspectives of the neuroscience research community, an academic library, and a data repository.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

Joshua Wallach, M.S., Ph.D., conducts research focuses on synthesizing, evaluating, and establishing the best evidence to inform research, regulatory, and public health decisions. His primary area of research, known as meta-research (i.e. the study of research itself), includes the key thematic areas of research methods, reporting/transparency, and reproducibility. Dr. Wallach's research interests include meta-analytical methodology, evaluating study biases, clinical trial design/conduct, pharmacoepidemiology, and regulatory science. His work with the Collaboration for Research Integrity and Transparency (CRIT) at Yale focuses on evaluating the tools, standards, and approaches used to assess the safety, efficacy, quality, and performance of FDA-regulated products using epidemiologic and meta-research methods. Dr. Wallach is also a Faculty Affiliate of the Meta-Research Innovation Center at Stanford (METRICS). Dr. Wallach is currently leading or collaborating on numerous studies, including meta-analyses of environmental exposures and clinical interventions, real world data analyses of medications, and meta-research projects with students.

Jeremy Wolfe, Ph.D., is Professor of Ophthalmology and Professor of Radiology at Harvard Medical School. He is Director of the Visual Attention Lab at Brigham and Women's Hospital. Wolfe received an AB in Psychology in 1977 from Princeton and his PhD in Psychology in 1981 from MIT. His research focuses on visual search and visual attention with a particular interest in socially important search tasks in areas such as medical image perception (e.g. cancer screening), security (e.g. baggage screening), and intelligence. His lab has been funded since 1982 by NIH (NEI, NIMH, NCI), NSF, AFOSR (Air Force), ONR (Navy), ARO (Army), Homeland Security, and the Nat. Geospatial Agency as well as by IBM, Google, Toshiba, Hewlett-Packard, & GE. Wolfe taught Intro. Psychology and other courses for 25 years, mostly at MIT. Leadership: Past President or Chair: Federation of Associations in Behavioral and Brain Sciences (FABBS), Psychonomic Soc, APA Division 3, Eastern Psychological Assoc, NAS Panel on Soldier Systems. Boards: Vision Sciences Society, APA Div 1, 6. Founding Editor-in-Chief of Cognitive Research: Principles and Implications (CRPI). Past-Editor of Attention, Perception, and Psychophysics. Wolfe also serves on the Oversight Committee of the North American Board of the Union for Reform Judaism. He was elected to American Academy of Arts and Sciences in 2019.

Letisha R. Wyatt, Ph.D., is an Assistant Professor of Neurology and the Director of Diversity in Research in the OHSU Research and Innovation Office. She currently oversees the development and implementation of training programs for scientists from underrepresented backgrounds. She is a former bench neuropharmacologist with a strong record of mentorship in the laboratory and classroom. Dr. Wyatt has also served as faculty in the OHSU Library and the Center for Cancer Early Detection Advanced Research (CEDAR), working together with researchers to support open science practices and data stewardship needs.

Alberto Zignoni has worked for over three years as the Market Development Director in the Research Data Management Solutions team. In his role, he has been responsible for the go-to-market of Elsevier's Research Data Management solutions across the globe. Prior to that, Alberto has worked as a Senior Consultant in the Research Intelligence sales team for South Europe, Middle East and Africa. In this capacity he has led various large scale projects for academic institutions, government bodies and funding agencies, including several national research assessment exercises.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy
Board on Health Care Services

Changing the Culture of Data Management and Sharing: A Workshop

April 28-29, 2021
Virtual Public Workshop

Speaker Guidance Questions

TO BE CONSIDERED BY ALL SPEAKERS:

1. What are **three key points** that you want the audience to take away from your presentation?

INTRODUCTORY SESSION AND KEYNOTE

Part A:

1. (*Martone, Nakamura*) What is the current state of biomedical research data sharing and the obstacles to meeting the new NIH policy in 2023?
2. (*Jorgenson*) Why is data sharing important from NIH's perspective? What are the end goals of the new policy (e.g., data reuse, integration, reproducibility, speed, verification, interoperability, etc.)?
3. (*Jorgenson*) From your perspective, what will success look like after the new policy goes into effect?
4. (*Brennan*) What did we learn about the importance of data management and sharing during the COVID-19 pandemic? From your vantage point, did data sharing become more common or more efficient?
5. (*Brennan*) Have you seen any exciting innovations around scientific collaboration and/or data sharing during the pandemic? Are there lessons or specific approaches that could be carried forward? How do you see data sharing evolving after the pandemic?

Part B:

6. (*Butte, Mangravite, Ropelewski, Wallach*) If data sharing is the means, what are the end goals? Data re-use? Integration? Reproducibility? Verification? Interoperability? Policy goals? Other?
7. (*Butte, Mangravite, Ropelewski, Wallach*) What do you see as the biggest challenges to implementing the NIH data sharing policy across different types of data? What steps can be taken to mitigate these challenges?
8. (*Butte, Mangravite, Ropelewski, Wallach*) From your perspective, what will successful data sharing look like after the implementation of the NIH policy? Are there specific metrics that would be useful for tracking the impact of the NIH policy?

SESSION I. STRATEGIES FOR MANAGING AND SHARING DATA: DIVERSE NEEDS AND CHALLENGES

1. What practical costs are associated with implementing data management and sharing? How can investigators get to the point where they can share useful data?
2. From your vantage point, what are the challenges associated with data formatting? Are you aware of successful efforts related to data harmonization and the development of standards that could be applied in a more general way?
3. Can you tell us about approaches that you are aware of that could make data management and sharing easier? (e.g., automated systems for importing data into repositories)
4. Are repositories prepared from a financial perspective and with respect to space needed to host data? What metrics should repositories be collecting?

SESSION II. MONITORING AND EVALUATING DATA MANAGEMENT AND SHARING PRACTICES

1. (*Lowenberg, Zigoni*) Please provide some examples of effective data citations, linking, and measuring. In your view, what is the current state of data citation methods and tools? Is the field in a good place, and if not, what can be improved?
2. *Lowenberg, Zigoni*) If approaches to data citation have been figured out conceptually, why is data citation not happening more frequently in practice? What stands in the way of wider implementation and acceptance of these practices?
3. (*Hanisch*) What level of specificity is important to include in a successful data sharing plan? What have you learned from your experience with developing the [NIST Research Data Framework \(RDaF\)](#) that might be applicable to the NIH data management and sharing policy?
4. (*Hanisch*) What elements do investigators need to keep in mind in order to comply with the NIH data management and sharing policy?
5. (*Hanisch*) Are there standardization efforts underway that could enable data sharing and data reuse?
6. (*Martin*) What role can university libraries and their staff play in helping with implementation of the NIH data management and sharing policy? Can you briefly describe the RDMLA efforts and adoption, and how this has helped with data sharing? Can this be leveraged for the NIH data management and sharing policy?
7. (*Martin*) Are there efforts underway in the university library community to help establish best practices for data management and sharing? What training is required from the librarian perspective and also from the researcher perspective?

SESSION III: ENCOURAGING UPTAKE OF DATA SHARING IN THE SCIENTIFIC COMMUNITY

Key Questions:

1. How can we ensure that there will be uptake of the new NIH policy in the scientific community? From your perspective, what will success look like after the policy goes into effect?
2. What does a modern laboratory need (e.g., tools/infrastructure) in order to be prepared to implement the policy?
3. What are the needs around training and education for laboratories to successfully implement the new data sharing policy?

SESSION IV: VALUE AND COSTS OF MANAGING AND SHARING DATA

Key Questions:

1. (*Van Gulick, Borghi*) What are the current perceptions among researchers about the value of data management and sharing practices? How do researchers think about the value of data management and sharing for themselves/their labs, and when others in the field use their data? Does the researchers' perceived value of data sharing align with the realized value after sharing?
2. (*Van Gulick, Borghi*) What challenges are researchers encountering with data management and sharing and how have they attempted to overcome those challenges?
3. (*Irizarry, Goroff, Pasquetto*) What challenges exist with data re-use with regard to data quality? How can researchers improve the quality of the data they are sharing so that it can be more effectively re-used to advance science?
4. (*Bjaalie*) What is the current status of data curation across different disciplines in research? What are the potential barriers to successful data curation and are you aware of successful efforts to overcome those barriers?
5. What type of feedback and/or metrics around data management and sharing would be attractive to investigators?

SESSION V: SHAPING A CULTURE OF DATA SHARING – REDUCING BARRIERS AND INCREASING INCENTIVES

Key Questions:

1. (*Rosati*) What legal obstacles arise in the context of the NIH data sharing policy, and how can they be mitigated by researchers and institutions?
2. (*Rosati*) How are de-identification standards evolving?
3. (*Rosati*) What issues related to data “ownership” may arise in the context of data sharing?
4. (*Rothstein*) What should researchers be considering when they seek informed consent for large-scale data sharing efforts? What are best practices in explaining potential future risks for large-scale data sharing, especially for non-genetic data (given that NIH has [available resources](#) for genetic data)? Are there approaches that can be taken to mitigate these risks?

5. *(Rothstein)* What major components of consent for data sharing should be the focus in the next 2 years? (e.g., different views on language from the perspective of all stakeholders; coded versus non-coded consents; opting in/out for future research) Are there other elements of the consent that should be considered?
6. *(Han)* What are some key considerations with regard to data governance processes, including control of de-identified information and University permission for faculty data sharing?
7. *(Han)* What steps can researchers and universities take to be good stewards of de-identified data? What are the specific training needs around de-identification of data?
8. *(Allen)* What are the major ethical issues related to data sharing that NIH funded investigators need to take into account as policy implementation takes place? How will the ethics of data sharing likely evolve?
9. *(Thakur, Franko, Farley)* How are non-NIH funders promoting data management and sharing practices? Are there lessons learned from non-NIH funders about incentivizing or promoting data management and sharing that could help inform the implementation of the NIH policy? Is there a way to encourage more open data sharing outside of mandates?
10. *(for all speakers)* How can researchers and other data stakeholders help to build and maintain trust from patients and participants around the sharing of their data?

SESSION VI: IMPLEMENTING THE NIH DATA SHARING POLICY: ARE WE READY FOR 2023?

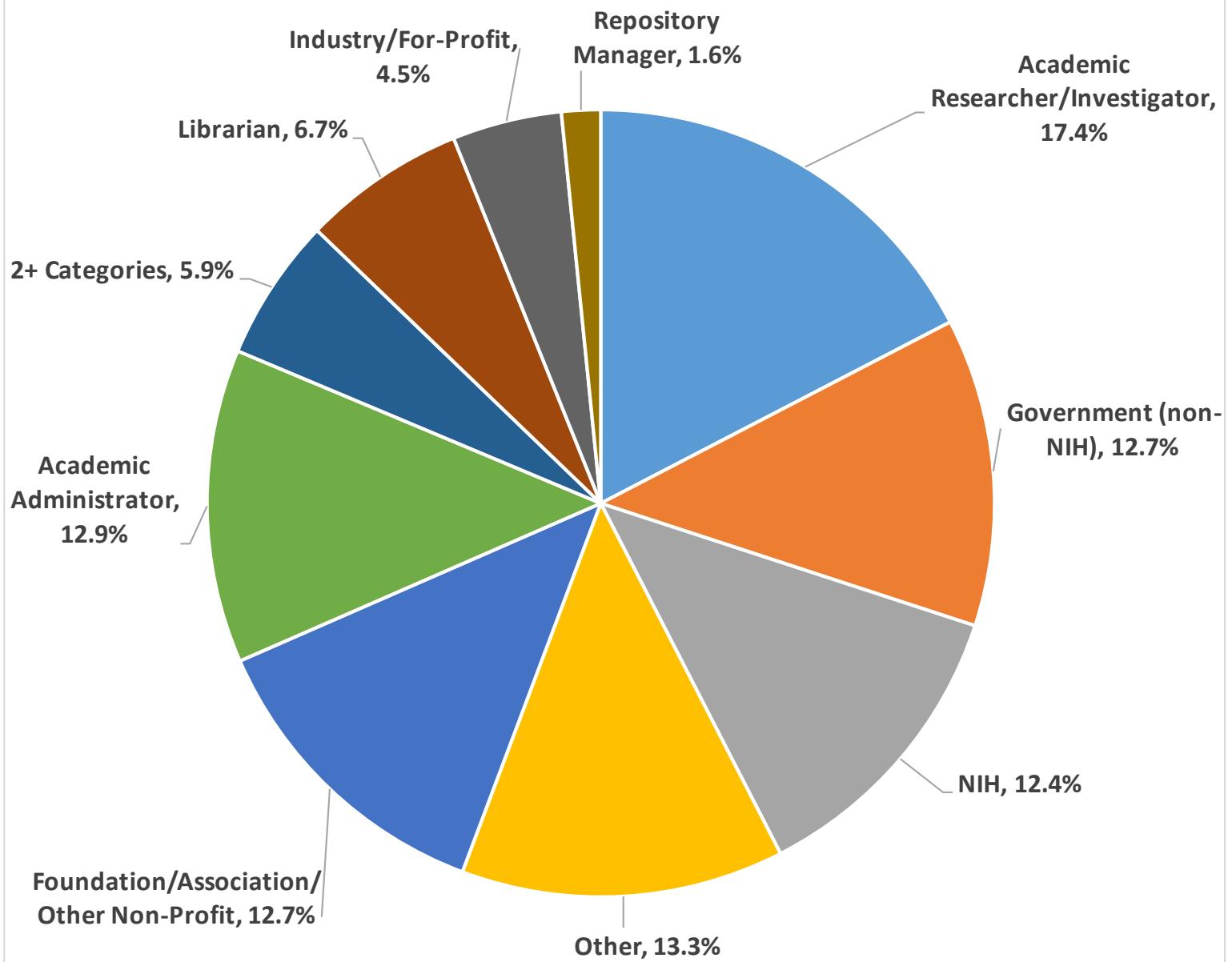
Key Questions:

1. What 1-2 actions would you prioritize in preparation for the implementation of the NIH data management and sharing policy in 2023?
2. What does a research laboratory need to look like by 2023 in order to be positioned for successful compliance with the new NIH data management and sharing policy?
3. What 1-2 themes or takeaways did you hear during the workshop that may require further attention? Did you hear about any promising solutions today that may help to overcome challenges to implementing the NIH data management and sharing policy? What promising next steps could be taken to incentivize data sharing across the biomedical research community?
4. Were there any key challenges and potential solutions to data management and sharing that were not addressed during the workshop?
5. What would successful data sharing look like through the NIH policy? What metrics could help measure success of the NIH policy?

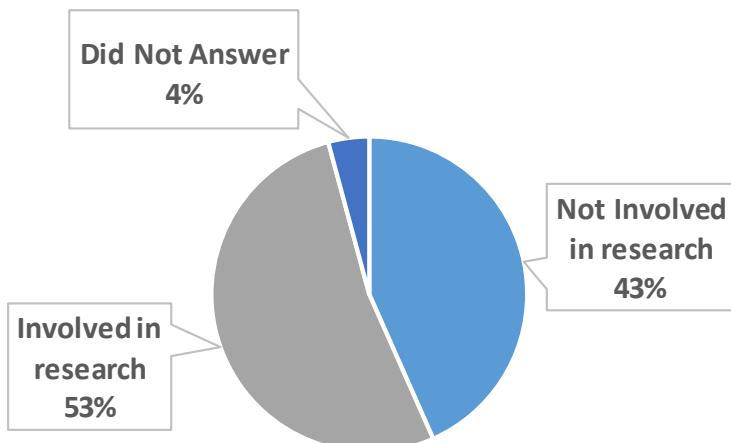
Workshop Registration (as of 04/13/2021)

Total participants: 1312

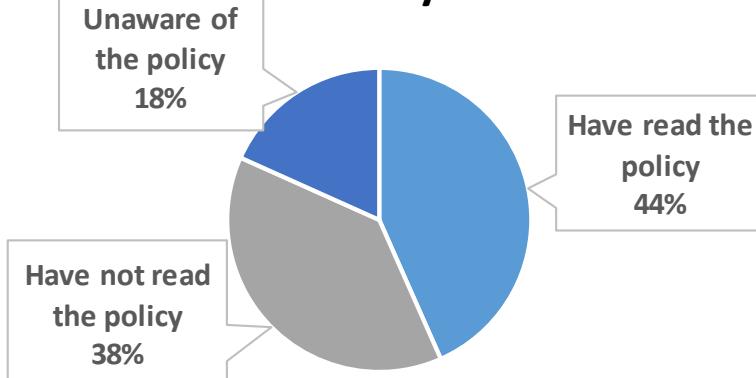
Workshop Participants by Category



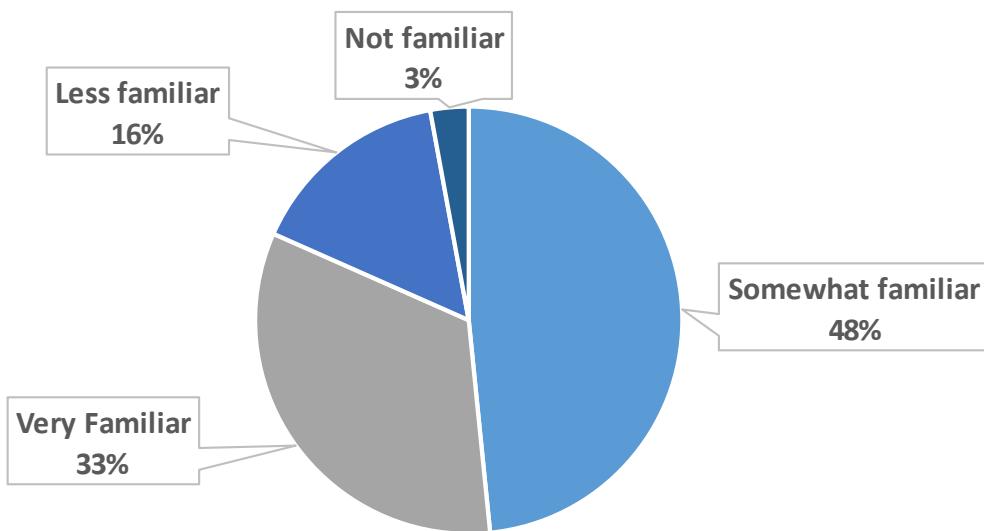
Research Involvement



Familiarity with NIH Data Sharing Policy



Familiarity with Data Sharing and Data Management



The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Board on Health Sciences Policy

Board on Health Care Services

National Academies Staff Roster

Sarah H. Beachy, Ph.D.
Senior Program Officer
Board on Health Sciences Policy

Andrew M. Pope, Ph.D.
Senior Board Director
Board on Health Sciences Policy

Siobhan Addie, Ph.D.
Program Officer
Board on Health Sciences Policy

Bridget Borel
Program Coordinator
Board on Health Sciences Policy

Meredith Hackmann
Associate Program Officer
Board on Health Sciences Policy

Erin Balogh, Ph.D.
Senior Program Officer
Board on Health Care Services

Lydia Teferra
Research Assistant
Board on Health Sciences Policy

BACKGROUND MATERIALS

Links to Additional Resources

OPENING SESSION

Introduction to the NIH Policy on Data Sharing and Management

- Final NIH Data Management and Sharing Policy: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>
- [NOT-OD-21-014](#) – Supplemental Information to the NIH Policy for Data Management and Sharing: Elements of an NIH Data Management and Sharing Plan
- [NOT-OD-21-015](#) – Supplemental Information to the NIH Policy for Data Management and Sharing: Allowable Costs for Data Management and Sharing
- [NOT-OD-21-016](#) – Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research
- NIH/OSP/Data sharing homepage: <https://osp.od.nih.gov/scientific-sharing/nih-data-management-and-sharing-activities-related-to-public-access-and-open-science/>
- Baynes and Hahnel. (2020). Research Practices in the wake of COVID-19: Busting open the myths around open data: <https://www.springernature.com/gp/advancing-discovery/blog/blogposts/research-practices-in-the-wake-of-covid/18256280>
- Butte (2021). Trials and Tribulations—11 Reasons Why We Need to Promote Clinical Trials Data Sharing: <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2775663>
- Norgeot, Glicksberg, and Butte. (2019). A call for deep-learning healthcare: <https://www.nature.com/articles/s41591-018-0320-3>

SESSION I: STRATEGIES FOR MANAGING AND SHARING DATA: DIVERSE NEEDS AND CHALLENGES

- Gorgolewski et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4978148/>
- Gorgolewski and Poldrack (2014). Making big data open: Data sharing in neuroimaging: <https://www.nature.com/articles/nn.3818>
- Perrier, Blondal, and MacDonald (2020). The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229182>
- Sage Bionetworks- Landscape analysis of approaches available for management of data sharing practices <https://sage-bionetworks.github.io/governanceGreenPaper/>

SESSION II: MONITORING AND EVALUATING DATA MANAGEMENT AND SHARING PRACTICES

- Research Data Management Librarian Academy (RDMLA) Factsheet: <https://rdmla.github.io/>

SESSION III: ENCOURAGING UPTAKE OF DATA SHARING IN THE SCIENTIFIC COMMUNITY

- Association of American Universities and Association of Public Land-Grant Universities (2019). Accelerating Public Access to Research Data: Workshop Report: <https://www.aplu.org/projects-and-initiatives/research-science-and-technology/public-access/workshop-on-public-access-report-aplu-aau-2019.pdf>
- Ayris et al. (2018). Open Science and its role in universities: a roadmap for cultural change: <https://www.leru.org/publications/open-science-and-its-role-in-universities-a-roadmap-for-cultural-change>
- Coetzee et al. (2021). Data Sharing Goals for Nonprofit Funders of Clinical Trials: <https://jopm.jmir.org/2021/1/e23011>
- Community toolkit for life scientists: <https://rdmkit.elixir-europe.org/>
- Databrary- Video data library for behavioral scientist: <https://databrary.org/>
- Gilmore and Adolph. (2017). Video can make behavioural science more reproducible: <https://www.nature.com/articles/s41562-017-0128>
- Gilmore, Kennedy, and Adolph. (2018). Practical Solutions for Sharing Data and Materials From Psychological Research: <https://journals.sagepub.com/doi/10.1177/2515245917746500>
- Jetten et al. (2021). Professionalising data stewardship in the Netherlands. Competences, training and education. Dutch roadmap towards national implementation of FAIR data stewardship: <https://zenodo.org/record/4486423#.YHR1XehKg2w>
- Mayor et al. (2021). Implementing FAIR data management within the German Network for Bioinformatics Infrastructure (de.NBI) exemplified by selected use cases: <https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbab010/6135008>
- NASEM Data Sharing Scrolling Page: <https://www.nap.edu/resource/25838/interactive/>
- NASEM Consensus Study- Open Science by Design: Realizing a Vision for 21st Century Research (2018): <https://www.nap.edu/catalog/25116/open-science-by-design-realizing-a-vision-for-21st-century>
- Play & Learning Across a Year (PLAY) project: <https://www.play-project.org/>
- Report of the 2020 Science Europe High Level Workshop on ERA: <https://scienceeurope.org/media/2ykjc3uz/20201118-high-level-workshop-report.pdf>
- Science Europe (2020). Implementing Research Data Management Policies across Europe: https://www.scienceeurope.org/media/jikjl2g/se_rdm_best_practices.pdf

- Science Europe (2019). Practical Guide to the International Alignment of Research Data Management - Extended Edition:
https://www.scienceeurope.org/media/4brkxxe5/se_rdm_practical_guide_extended_final.pdf

SESSION IV: VALUE AND COSTS OF MANAGING AND SHARING DATA

- Borghi and Van Gulick. (2018). Data management and sharing in neuroimaging: Practices and perceptions of MRI researchers:
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0200562>
- Borghi and Van Gulick. (2020). Data Management and Sharing: Practices and Perceptions of Psychology Researchers. <https://psyarxiv.com/7g3ae/>
- Borghi et al. (2018). Support Your Data: A Research Data Management Guide for Researchers: https://www.researchgate.net/publication/325048194_Support_Your_Data_A_Research_Data_Management_Guide_for_Researchers
- Danchev et al. (2021). Evaluation of Data Sharing After Implementation of the International Committee of Medical Journal Editors Data Sharing Statement Requirement:
<https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2775667>
- Pasquetto, Borgman, and Wofford. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage: <https://hdsr.mitpress.mit.edu/pub/jduhd7og/release/8>

SESSION V: SHAPING A CULTURE OF DATA SHARING – REDUCING BARRIERS AND INCREASING INCENTIVES

- Rothstein (2015). Ethical Issues in Big Data Health Research:
<https://journals.sagepub.com/doi/pdf/10.1111/jlme.12258>
- Rothstein (2010). Is Deidentification Sufficient to Protect Health Privacy in Research?
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3032399/>
- Rothstein (2017). Structural Challenges of Precision Medicine: Currents in Contemporary Bioethics: <https://journals.sagepub.com/doi/10.1177/1073110517720655>

Final NIH Policy for Data Management and Sharing

Notice Number:

NOT-OD-21-013

Key Dates

Release Date:

Effective Date:

October 29, 2020

January 25, 2023

Related Announcements

[NOT-HG-21-023](#) - Notice Announcing NHGRI Guidance for Third-Party Involvement in Extramural Research

[NOT-HG-21-022](#) - Notice Announcing the National Human Genome Research Institute's Expectation for Sharing Quality Metadata and Phenotypic Data

[NOT-OD-21-014](#) – Supplemental Information to the NIH Policy for Data Management and Sharing: Elements of an NIH Data Management and Sharing Plan

[NOT-OD-21-015](#) – Supplemental Information to the NIH Policy for Data Management and Sharing: Allowable Costs for Data Management and Sharing

[NOT-OD-21-016](#) – Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research

[NOT-OD-20-013](#) - Request for Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance

Issued by

Office of The Director, National Institutes of Health ([OD](#))

Purpose

Summary

The National Institutes of Health (NIH) is issuing this final NIH Policy for Data Management and Sharing (DMS Policy) to promote the management and sharing of scientific data generated from NIH-funded or conducted research. This Policy establishes the requirements of submission of Data Management and Sharing Plans (hereinafter Plans) and compliance with NIH Institute, Center, or Office (ICO)-approved Plans. It also emphasizes the importance of good data management practices and establishes the expectation for maximizing the appropriate sharing of scientific data generated from NIH-funded or conducted research, with justified limitations or exceptions. This Policy applies to research funded or conducted by NIH that results in the generation of scientific data.

Background

Sharing scientific data accelerates biomedical research discovery, in part, by enabling validation of research results, providing accessibility to high-value datasets, and promoting data reuse for future research studies.¹⁴ As a steward of the nation's investment in biomedical research, NIH has long championed policies that make research available to the public to achieve these goals. For example,

the [2003 NIH Data Sharing Policy](#) reinforced NIH's commitment to data sharing by requiring investigators to address data sharing in applications for large research awards. [NIH's 2014 Genomic Data Sharing \(GDS\) Policy](#), initially preceded by the [2008 Genome-Wide Association Studies Policy](#), set the expectation that researchers share large-scale genomic data, regardless of species, to enable the combination of large and information-rich datasets. In 2016, the [NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information \(Clinical Trials Policy\)](#) further reinforced NIH's commitment to research participants and the research community by making the results of clinical trials accessible in a timely fashion.

NIH recognizes that its data sharing policy efforts must flexibly evolve to keep pace with scientific and technological opportunities and notes that researchers' ability to generate, store, share, and combine data has never been greater. To capitalize on these advancements, NIH initiated the development of a more comprehensive data sharing policy alongside its efforts to modernize data sharing infrastructure in its [2015 Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research](#). With policy and infrastructure modernization efforts working in tandem, NIH initiated a stepwise process for seeking feedback from the community to develop a robust data sharing policy capable of reflecting the diversity of its community's data sharing needs. In 2016, NIH requested public comments on data management and sharing strategies and priorities ([NOT-OD-17-015](#)). In 2018, NIH solicited public input on proposed key provisions that could serve as a foundation for a future NIH policy for data management and sharing ([NOT-OD-19-014](#)). Using public feedback to inform its thinking, in 2019 NIH released a draft proposal for a future data management and sharing policy in the Federal Register ([84 FR 60398](#)).

Along with the Draft Policy proposal, NIH sought feedback on supplemental materials that could help researchers integrate effective data management and sharing practices into research, including "[Elements of an NIH Data Management and Sharing Plan](#)" and "[Allowable Costs for Data Management and Sharing](#)." We note that a third document, "Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research," was developed in response to public comments received on both the Draft Policy and the "Request for Public Comments on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research," which was released for public comment by the White House Office of Science and Technology Policy (OSTP) to promote consistency across federal agencies and reduce researcher burden ([85 FR 3085](#)).

In respect and recognition of Tribal sovereignty, NIH also initiated Tribal Consultation on its Draft Policy proposal, in accordance with the [HHS Tribal Consultation Policy](#) and the [NIH Guidance on the Implementation of the HHS Tribal Consultation Policy](#). The [NIH Tribal Consultation Report – NIH Draft Policy for Data Management and Sharing](#) provides more detail on the Tribal Consultation process relative to the development of the final DMS Policy and NIH's response. Briefly, three themes emerged from Tribal Nations' input: 1) Strengthen engagement built on trust between researchers and Tribal Nations; 2) Train researchers to responsibly and respectfully manage and share American Indian and Alaska Native (AI/AN) data; and 3) Ensure research practices are aligned with the laws, policies, and preferences of AI/AN community partners. NIH intends to continue discussions to ensure appropriate implementation of the DMS Policy as it relates to these communities, and details about some of the implementation planning follows in the discussion below.

Overview of Public Comments

NIH incorporated feedback over the course of several years to develop a data management and sharing policy proposal and released its Request for Comments on the Draft NIH Policy for Data Management and Sharing and Draft Supplemental Guidance on November 8, 2019 ([84 FR 60398](#), comment period closing on January 10, 2020). NIH held a public webinar on December 16, 2019, with over 580 people participating. In response to the Draft Policy, NIH received 203 responses from both domestic and international stakeholders, and the comments are publicly available.¹² The largest group of respondents reported affiliation with universities, followed by nonprofit research organizations, professional associations (tied with "other"), as well as small percentages of respondents affiliated with government agencies, healthcare delivery organizations, and patient advocacy organizations. Respondents typically identified themselves as scientific researchers, while another sizeable section self-identified as "other."

Remaining respondents identified as institutional officials, with smaller percentages self-identified as bioethicists or social science researchers, government officials, patient advocates, and members of the public. NIH considered all feedback in the development of the final DMS Policy, and a discussion of the public comments on topics follows below.

Discussion of Public Comments on the Draft NIH Policy for Data Management and Sharing

Clarifying Expectations for Sharing Scientific Data

Draft Policy: The Draft Policy did not explicitly set a default expectation of data sharing. Rather, it focused on requiring submission of and compliance with a Data Management and Sharing Plan (Plan) that outlines how data will be managed and shared. The Draft Policy also included recognition of that fact that certain factors (i.e., legal, ethical, or technical) may limit the ability to preserve and share data.

Public Comments: While commenters were generally supportive of the overall scope of the Draft Policy, many requested NIH make an explicitly stronger commitment to expecting data sharing from the research community. Suggestions included requiring data sharing and indicating that data sharing should be the default, with well justified exceptions being permitted.

Final Policy: The final DMS Policy does not create a uniform requirement to share all scientific data. Unlike a requirement for submission of Plans, which can be implemented across various funding mechanisms and types of research with little variation, appropriate data sharing is likely to be varied and contextual. Through the requirement to submit a Plan, researchers are prospectively planning for data sharing, which we anticipate will increasingly lead researchers to integrate data sharing into the routine conduct of research. Accordingly, we have included in the final DMS Policy an expectation that researchers will maximize appropriate data sharing when developing Plans. The final DMS Policy retains the Draft Policy's factors (i.e., ethical, legal, or technical) that may necessitate variations in the extent of scientific data preservation and sharing, and researchers should convey such factors in their Plans. The final DMS Policy has also been modified to clarify these factors are not limited to data derived from human research participants. We believe this will provide the necessary flexibility for researchers to accommodate the substantial variety in research fields, projects, and data types that this expectation will encompass.

Definition of "Scientific Data"

Draft Policy: The scope of which data will be shared relies on the definition of "scientific data." This term was defined in the Draft Policy as: "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."

Public Comments: Commenters focused on a variety of aspects of the definition of "scientific data." They suggested that the concept of data quality be included, as data that may otherwise meet the definition but, if uninterpretable, are not of value. Commenters also suggested the definition address null or negative findings (and indicate that these data should be shared). Commenters requested clarification about the sentence that NIH expects reasonable efforts will be made to digitize all scientific data, including whether NIH would cover costs to digitize data that are not collected in digital form.

Final Policy: The final DMS Policy defines Scientific Data as: "The recorded factual material commonly accepted in the scientific community as of sufficient quality 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." We agree that data quality is an important concept to convey to ensure that scientific data are useful and to prevent data sharing from becoming a perfunctory administrative requirement, but rather one that should be done with the understanding that these data are intended to be used by others. Therefore, we have added to the definition that the data should be of sufficient quality to validate and

replicate research findings. Even those scientific data not used to support a publication are considered scientific data and within the final DMS Policy's scope. We understand that a lack of publication does not necessarily mean that the findings are null or negative; however, indicating that scientific data are defined independent of publication is sufficient to cover data underlying null or negative findings.

We also note that while the final DMS Policy states that scientific data are those of sufficient quality to "validate and replicate," we anticipate that shared scientific data will be used for a variety of purposes (consistent with applicable laws, policies, and limitations) including subsequent analyses, as suggested in the Purpose section of the final DMS Policy. Therefore, the concepts of validation and replication provide a standard for determining what constitutes scientific data and are not intended to limit uses of shared data.

Finally, we have removed the expectation for digitizing scientific data. We encourage reasonable efforts to digitize data, recognizing that digitizing data may be a technical factor that may limit the sharing of data.

Timing of Submission of Data Management and Sharing Plans

Draft Policy: The Draft Policy proposed the submission of Plans at Just-in-Time for grants.

Public Comments: While we received a range of comments about timing of Plan submission, the majority were opposed to or requested further clarification about Just-in-Time Plan submission. Commenters were concerned about not having sufficient time to develop Plans and expressed concerns about the Plan revision process leading to delays in issuing awards. Others indicated that institutions would want to review Plans because they would ultimately be responsible for compliance, but a Just-in-Time Plan submission would not afford institutions sufficient time. A key practical concern with Just-in-Time Plan submission was difficulty submitting a budget at application that included requests for allowable data management and sharing costs prior to actually drafting the Plan. Commenters who favored submitting Plans at Just-in-Time frequently cited decreased burden on applicants, because with Just-in-Time, only those applicants likely to be funded would be required to submit Plans, rather than all applicants.

Final Policy: The final DMS Policy requires submission of a Plan for extramural grants at application. This approach is more conducive to achieving NIH's goal of promoting a culture in which data management and sharing are recognized to be an integral component of a biomedical research project, rather than an administrative or additive one. While NIH is aware that this approach places the requirement on the general pool of grant applicants rather than on those likely to be funded, it is precisely this approach of prospective planning for data management and sharing that NIH hopes to promote and that a number of commenters suggested is crucial for ensuring more regular planning for data management and sharing. We were swayed by the logistical concerns expressed in comments, namely how applicants could submit budgets appropriately reflective of data management and sharing when not yet required to submit the Plan that is intended to help them consider these issues. In addition, the concerns about institutions having sufficient time to review Plans and potential logistical challenges in issuing timely awards was persuasive. This approach is also consistent with the 2018 Request for Information on Proposed Provisions of a Draft Data Management and Sharing Policy for NIH Funded or Supported Research, which proposed Plans be submitted with extramural grant applications. The responses to that proposal generally favored Plan submission at the time of application.

Assessment of Plans

Draft Policy: The Draft Policy proposed that NIH Program Staff in the funding NIH ICO assess Plans from extramural grants.

Public Comments: Many commenters supported peer review of Plans, noting their skill and that peer review of Plans would promote a cultural shift in favor of data sharing. Commenters also suggested that NIH Program Staff review may lead to more consistent Plan assessment and decrease peer reviewer burden.

Final Policy: The final DMS Policy maintains NIH Program Staff assessments of Plans' merits. However, peer reviewers may comment on the proposed budget for data management and sharing, although these

comments will not impact the overall score. This approach balances the benefit of consistency afforded by NIH Program Staff review of Plans, review of updates, and compliance monitoring, with the opportunity for peer reviewers to comment on the requests for data management and sharing costs. Over time, and through these reviews, we hope to learn more about what constitutes reasonable costs for various data management and sharing activities across the NIH portfolio of research.

NIH ICO Consistency of Data Sharing Expectations

Draft Policy: The Draft Policy noted that NIH ICOs may supplement the Policy's expectations for Plans with their own complementary requirements to further advance their specific program or research goals. In addition, the Draft Policy stated the funding NIH ICO may request additional or specific information to be included within Plans to meet expectations for data management and sharing in support of programmatic priorities or to expand the utility of the scientific data generated from the research.

Public Comments: In light of various existing NIH ICO data sharing policies, commenters expressed confusion around having potentially varying expectations in data sharing policy implementation across NIH. There were concerns about insufficient direction to NIH ICOs and around a potentially uncoordinated variety of approaches. Commenters suggested guidance to facilitate NIH ICO consistency and suggested that NIH provide a centralized location of NIH ICO-specific expectations to help researchers navigate variations, particularly when subject to more than one NIH ICO's data sharing policies.

Final Policy: While the final DMS Policy's language on this issue has not substantively changed from that of the Draft Policy, we have heard the concerns and intend to address them during the period of implementation planning prior to the DMS Policy's Effective Date. NIH ICOs can, within certain bounds, meet their scientific, policy, and programmatic goals in different ways. As such, this Policy affords NIH ICOs the opportunity to meet the goals of this Policy in ways that enhance their respective science. However, we intend to promote consistency on some key tenets of the final DMS Policy, such as the requirement for submission of Plans and the timing of their submission. The DMS Policy represents the minimum requirements for the NIH, but NIH ICOs may expect more specificity in Plans. For example, NIH ICOs and Programs may wish to promote, via specific Funding Opportunity Announcements (FOAs) or across their research portfolios, the use of particular standards to enable interoperability of datasets and resources. We are appreciative of the suggestion about how to organize NIH ICO-specific expectations and will be working to ensure clear implementation materials for applicants and awardees.

Data Derived from Human Participants

Draft Policy: The Draft Policy acknowledged the applicability of laws, regulations, guidance, and policies that govern the conduct of research with human participants and how data derived from human participants should be used. It also described that Plans should indicate how human participants and data derived from them would be protected. Finally, the Draft Policy acknowledged that certain factors may limit the ability to share data and proposed that these factors be described in the Plan. Importantly, the Draft Policy did not propose any new expectations for the conduct of research with human participants.

Public Comments: Commenters expressed concerns about how to safeguard participant privacy and confidentiality when sharing data, with some requesting information on de-identification practices. Commenters also requested guidance on best practices in communicating data sharing in informed consent. They also stressed the importance of data sharing to maximize the contributions of those who volunteer to participate in NIH-funded studies. Some pointed to special populations with preferences on data sharing issues, such as AI/AN populations, and asked how sharing of data from these participant populations is expected to be handled.

In addition to the public comments submitted during the comment period, NIH received input from the Secretary's Advisory Committee on Human Research Protections (SACHRP).¹² SACHRP provided a set of recommendations relating to applying the DMS Policy to research with human participants, some of which we have incorporated into the final DMS Policy and are discussed below.

AI/AN communities provided input through various channels, including through letters sent to NIH as part of government-to-government communications. The Tribal Consultation process also led to valuable input that is informing NIH's implementation efforts, described further below.

Final Policy: As with the Draft Policy, the final DMS Policy does not introduce new requirements for protections for research with human participants. Existing laws (e.g., Certificates of Confidentiality), regulations (e.g., the Common Rule), and policies (e.g., the NIH Genomic Data Sharing Policy) continue to apply. However, through this Policy and associated supplemental information and other activities, NIH promotes thoughtful practices regarding the treatment of data derived from human participants.

In response to public comments and SACHRP's recommendations on the Draft Policy, we have included in the final DMS Policy three concepts that we believe are important to emphasize for investigators as they think through how to engage prospective participants regarding what is expected to happen with the data they contribute and, downstream, how best to respect these contributions. First, we encourage investigators to consider, while developing their Plans, how to address data management and sharing in the informed consent process, such that prospective participants will understand what is expected to happen with their data. This planning will serve investigators as they develop their Plans, because some of the Plan elements prompt investigators to outline anticipated factors that might affect the ability to share and preserve scientific data, such as any limitations arising from the informed consent process. NIH also intends to develop resources to help researchers and institutions in communicating the intent to share data with prospective research participants. Second, we note that any limitations on subsequent use of data (which may apply to non-human data as well) should be communicated to those individuals or entities preserving and sharing the scientific data. This ensures that factors that may affect subsequent use of data are properly communicated and will travel with the data. Finally, we highlight the importance of researchers considering whether, in choosing where and how to make their data available (if not already specified by an FOA or funding NIH ICO expectation), access to scientific data derived from humans should be controlled, even if de-identified and lacking explicit limitations on subsequent use.

We note that data carrying explicit limitations on subsequent use require access controls to manage such limitations. This approach honors the wishes and autonomy of the participants who contributed their data and is important to uphold, even if the data are de-identified. In addition, investigators should consider whether access to data even without such limitations should be controlled. SACHRP identified concerns regarding re-identification of otherwise de-identified data, and indeed technological advances and increasing interoperability among data resources, while providing opportunities for new analyses, present identifiability concerns that are widely acknowledged. In response to concerns expressed in public comments and by SACHRP, NIH may support development of resources to assist researchers and institutions in determining how to appropriately de-identify data from human participants, as well as for communicating data sharing in informed consent.

The final DMS Policy does not preclude the open sharing of data from human participants in ways that are consistent with consent practices, established norms, and applicable law. For example, open sharing of a compilation of a population's genotype at a particular locus may be an acceptable and established practice if consistent with informed consent. And importantly, we are aware that some patient communities prioritize openness to speed scientific progress and discovery. Nothing in the final DMS Policy is intended to prevent these approaches, as long as participants are appropriately informed and prospectively agree to them.

We emphasize that respecting participant autonomy and maintaining privacy of participants and confidentiality of their data can be consistent with data sharing. Through the final DMS Policy, we outline a balance that accommodates various responsible approaches that meet data sharing expectations and honor appropriate limitations in sharing. In addition, while the DMS Policy sets the expectation that, through their Plans, researchers maximize the appropriate sharing of scientific data (acknowledging factors that may limit such sharing, as discussed above), the DMS Policy does not expect that the informed consent given by participants to be obtained in any particular way, such as through broad consent.

In response to input from Tribal Nations, the final DMS Policy clarifies agency respect for Tribal sovereignty in the absence of written Tribal laws or policies. To address some of the other themes and comments we heard from both AI/AN communities as well as public commenters who expressed interest in agency efforts to promote responsible and respectful engagement of AI/AN populations, we are developing supplemental information for researchers who wish to work with AI/AN communities. Such guidance is expected to encourage researchers to (among other topics): thoughtfully consider the unique

data sharing concerns of AI/AN communities; respectfully negotiate agreements for data use with Tribal Nations; and enhance researcher awareness of processes Tribal Nations use to review prospective research. NIH will seek input from AI/AN communities on the development of the guidance, to ensure it serves the goals of guiding researchers while taking into account Tribal preferences and values.

When Data Are Expected To Be Shared

Draft Policy: The Draft Policy proposed that shared scientific data should be made accessible in a timely manner for use by the research community and the broader public.

Public Comments: While commenters appreciated the flexibility afforded by this approach, they also expressed concern about its ambiguity. Some suggested timing of data sharing be connected to publication. Commenters also suggested NIH should specify outer bounds for timing of data sharing in the absence of a publication. Overall, commenters expressed the desire for more clarity.

Final Policy: The final DMS Policy states that “[s]hared scientific data should be made accessible as soon as possible, and no later than the time of an associated publication, or the end of the award/support period, whichever comes first.” This statement provides more clarity than the Draft Policy through outer bounds to guide researchers in when to make the scientific data available. It clarifies that publication triggers release of the data that underlie that publication (indeed, publishers often require the same). But it also recognizes that research does not always lead to a publication that would itself trigger the release of data. Importantly, the final DMS Policy is designed to increase the sharing of scientific data, regardless of whether a publication is produced. Important research may never be published for a variety of reasons, not least of which because the results did not prove the hypothesis. However, we believe the scientific data underlying all NIH-funded research to be of importance, particularly to serve the purposes of accountability and transparency. Data that do not form the basis of a publication produced during the award period should be shared by the end of the award period. A single research project may take advantage of both approaches. Namely, researchers may share data underlying publication during the period of award but may share other data that have not yet led to a publication by the end of the award period.

How Long Data Should Be Available

Draft Policy: The Draft Policy stated that “NIH encourages shared scientific data to be made available as long as it is deemed useful to the research community or the public.”

Public Comments: Commenters expressed uncertainty about how the concept of usefulness would be determined, and who would determine usefulness.

Final Policy: We have indicated a framework for helping researchers think through a minimum time period for data availability. Providing this framework is anticipated to help researchers both develop Plans and also budget accordingly for data management and sharing costs, when needed. Existing requirements and expectations set forth through, for example, applicable record retention requirements, repository policies, and journal policies may guide researchers as they seek to define minimal periods for data availability. However, we encourage researchers to propose longer time periods that may be informed by other factors, such as anticipated value of the dataset for the scientific community and the public.

Where to Share Scientific Data

Draft Policy: The Draft Policy stated that “NIH encourages the use of established repositories for preserving and sharing scientific data.”

Public Comments: Commenters supported the use of established repositories for preserving and sharing scientific data.

Final Policy: The final DMS Policy strongly encourages the use of established repositories to the extent possible. This reflects NIH’s preference that scientific data be shared and preserved through repositories, rather than kept only by the researcher or institution and provided on request, with the recognition that this is not always a practical or even a preferred approach. For example, we recognize and respect that AI/AN communities, in particular, may wish to manage, preserve, and share their own data. We support

efforts that enable AI/AN communities to prioritize research opportunities and to ensure sufficient protections on scientific data generated from such research. In addition, we have released the Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research, which will aid researchers as they choose suitable repositories for the preservation and sharing of data. This supplemental information is discussed in more detail below.

Discussion of Public Comments on the Draft Supplemental Information: Elements of an NIH Data Management and Sharing Plan

Page Limit and Template for Plans

Draft Supplemental Information: The Draft Supplemental Information suggested a limit for Plan length of two pages or less. It did not indicate whether template Plans would be provided.

Public Comments: Commenters expressed that two pages is insufficient to describe approaches for data management and sharing, particularly for larger, more complicated projects, such as those involving consortia. In addition, commenters suggested that NIH provide a template for Plans, with Plans being machine-readable.

Final Supplemental Information: We understand the concern about describing plans for data management and sharing in two pages. In the final supplemental information, we have noted the elements to be addressed in two pages or less, indicating that these descriptions need not be long narratives. In addition, short Plans are anticipated to limit researcher burden.

The Acceptability of “To Be Determined” as a Response to Plan Elements

Draft Supplemental Information: The Draft Supplemental Information proposed that if certain elements of a Plan have not been determined by the time of Plan submission, an entry of “to be determined” may be acceptable if a justification is provided along with a timeline or appropriate milestone at which a determination will be made.

Public Comments: Commenters disagreed with allowing responses of “to be determined” at initial Plan submission.

Final Supplemental Information: The final Supplemental Information eliminates the language that a response of “to be determined” is acceptable. We do not expect researchers to necessarily have all details at the application stage, but we encourage researchers to fill out Plans to the best of their knowledge and ability, so the Plans may be appropriately assessed. We also note that adherence with NIH ICO-approved Plans is a requirement of the final DMS Policy. As indicated in the final DMS Policy, researchers will have opportunities to update their Plans throughout the course of their awards, subject to NIH ICO approval.

The Use of Persistent Unique Identifiers (PIDs)

Draft Supplemental Information: The Draft Supplemental Information asked for researchers to indicate how data will be findable and whether a persistent unique identifier or other standard indexing tools will be used.

Public Comments: Commenters expressed support for PIDs, explaining that researchers are incentivized to use PIDs because they enable effective citation. They also noted PIDs are a way to track data sharing compliance.

Final Supplemental Information: The final Supplemental Information asks researchers to describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools. This wording change is meant to highlight the importance of using a PID or other standard indexing tool so the data are findable, which is a key component of the FAIR (Findable, Accessible, Interoperable, and Reusable) Principles. PIDs are also listed as a desirable characteristic of data repositories in the Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research.

Data Security

Draft Supplemental Information: The Draft Supplemental Information proposed that researchers address provisions for maintaining the security and integrity of the scientific data, such as through encryption and back-ups. It also noted that data sharing should be consistent with security as well as other factors.

Public Comments: Commenters emphasized the importance of data security.

Final Supplemental Information: We have removed the prompt for researchers to address provisions related to the security of scientific data. While we agree with the importance of appropriate data security measures, we believe that technical provisions regarding data security are more appropriately addressed by the institutions and repositories preserving and sharing the scientific data. The Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research (discussed in more detail below) outlines characteristics of suitable repositories, and we do not wish to burden the funded community with describing in-depth the data security processes of the data repositories preserving and sharing the data generated by their research. While data may remain with an institution prior to submission to a data repository, the DMS Policy is not designed to set any new standards for institutional data security practices.

Discussion of Public Comments on the Draft Supplemental Information: Allowable Costs for Data Management and Sharing

Timelines for Using Funds for Data Management and Sharing Activities

Draft Supplemental Information: The Draft Guidance noted that budget requests to the NIH may include costs for preserving and sharing data through repositories that charge recurring fees, however it did not specify timelines by which funds allotted for data management and sharing must be spent or how to account for paying fees to data repositories storing data after the end of the performance period.

Public Comments: Commenters generally supported the proposal but sought clarification on whether funds may be used to pre-pay fees for long-term data availability. Commenters also asked whether these funds could cover personnel expenses.

Final Supplemental Information: Personnel costs required to perform the types of data management and sharing activities described in the final Supplemental Information are allowable. Regarding the availability of data beyond the end of the project, which is crucial to achieving the goals of the DMS Policy, the final Supplemental Information clarifies that fees for long-term data preservation and sharing are allowable, but funds for these activities must be spent during the performance period, even for scientific data and metadata preserved and shared beyond the award period. NIH funds cannot legally be spent after the award period.

Discussion of Requests for Additional Guidance and Information

Public commenters requested more clarity not only on information in provided materials, but about issues key to implementation. One common theme was a request for guidance about how to choose a data repository, with some requesting a list of suitable repositories. NIH does not intend to provide a comprehensive list of suitable repositories outside of those supported or stewarded by NIH.¹⁴ However, NIH recognizes the need for providing a way to help researchers determine what characteristics make for a suitable repository for the preservation and sharing of data from NIH-funded research. As such, we are releasing the Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research. This document stems in part from an interagency effort led by the White House OSTP to outline desirable characteristics of preserving and sharing data from federally funded research, released as the Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research ([85 FR 3085](#)). The purpose was also to promote consistency across federal agencies to reduce researcher burden. The public comments on this document also informed the development of the Supplemental Information.

The Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research includes a process to help researchers

determine suitable repositories by providing relevant characteristics, noting that NIH ICOs may have identified preferred repositories in FOAs or through other announcements.

Concluding Points

As the DMS Policy is released, the world is in the midst of the COVID-19 pandemic. The recognition that more open sharing can lead to faster advances and treatments has led to an unprecedented worldwide effort to openly share publications and data related to both SARS-CoV-2 (the novel coronavirus that causes COVID-19) and coronaviruses more generally. While this is a specific example of an urgent public health need, patients, families, and patient advocacy groups consider the diseases and conditions that affect them to be of equal urgency, as do those who research these diseases and conditions and treat affected patients. With public input, NIH has worked to develop and refine this DMS Policy, the goal of which is to increase the sharing of scientific data generated from NIH-funded research to ultimately enhance health, lengthen life, and reduce illness and disability.

In addition to the Supplemental Information discussed here, we intend to provide frequently asked questions and other information to aid in implementation, prior to the DMS Policy's Effective Date. We recognize that some fields and researchers plan for sharing and prepare data for preservation and sharing as a regular practice. For others, these activities may be new. We anticipate a period of learning and an evolution of implementation practices. Further, it is important to acknowledge that NIH recognizes that expectations for robust data management and sharing practices will need to be met with investments in and evolution of accompanying data infrastructure. We look forward to working with applicants and the funded community as they prepare to meet the DMS Policy's requirements and expectations, as we all move toward a future in which data sharing is a community norm.

The final DMS policy is set forth below. Upon its Effective Date, the DMS Policy replaces the 2003 NIH Data Sharing Policy.

NIH Policy for Data Management and Sharing

Section I. Purpose

The National Institutes of Health (NIH) Policy for Data Management and Sharing (herein referred to as the DMS Policy) reinforces NIH's longstanding commitment to making the results and outputs of NIH-funded research available to the public through effective and efficient data management and data sharing practices. Data sharing enables researchers to rigorously test the validity of research findings,¹ 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 FAIR data principles.²

Under the DMS Policy, NIH requires researchers to prospectively plan for how scientific data will be preserved and shared through submission of a Data Management and Sharing Plan (Plan). Upon NIH approval of a Plan, NIH expects researchers and institutions to implement data management and sharing practices as described. The DMS Policy is intended to establish expectations for Data Management and Sharing Plans, which applicable NIH Institutes, Centers and Offices (ICO) may supplement as appropriate.

Section II. Definitions

For the purposes of the DMS Policy, terms are defined as follows:

Scientific Data: The recorded factual material commonly accepted in the scientific community as of sufficient quality 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.

Data Management: The process of validating, organizing, protecting, maintaining, and processing scientific data to ensure the accessibility, reliability, and quality of the scientific data for its users.

Data Sharing: The act of making scientific data available for use by others (e.g., the larger research community, institutions, the broader public), for example, via an established repository.

Metadata: Data that provide additional information intended to make scientific data interpretable and reusable (e.g., date, independent sample and variable construction and description, methodology, data provenance, data transformations, any intermediate or descriptive observational variables).

Data Management and Sharing Plan (Plan): A plan describing the data management, preservation, and sharing of scientific data and accompanying metadata.

Section III. Scope

The DMS 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. The DMS Policy does not apply to research and other activities that do not generate scientific data, including training, infrastructure development, and non-research activities.

Section IV. Effective Date(s)

The effective date of the DMS Policy is January 25, 2023, including for:

- - Competing grant applications that are submitted to NIH for the January 25, 2023 and subsequent receipt dates;
 - Proposals for contracts that are submitted to NIH on or after January 25, 2023;
 - NIH Intramural Research Projects conducted on or after January 25, 2023; and
 - Other funding agreements (e.g., Other Transactions) that are executed on or after January 25, 2023, unless otherwise stipulated by NIH.

Section V. Requirements

The DMS Policy requires:

- - Submission of a Data Management and Sharing Plan outlining how scientific data and any accompanying metadata will be managed and shared, taking into account any potential restrictions or limitations.
 - Compliance with the awardee's plan as approved by the NIH ICO.

The 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 (see Supplemental Information to the NIH Policy for Data Management and Sharing: Allowable Costs for Data Management and Sharing).

Section VI. Data Management and Sharing Plans

Researchers planning to generate scientific data are required to submit a Plan to the funding NIH ICO as part of the Budget Justification section of the application for extramural awards, as part of the technical evaluation for contracts, as determined by the Intramural Research Program for Intramural Research Projects consistent with the objectives of this Policy, or prior to release of funds for other funding agreements. Plans should explain how scientific data generated by research projects will be managed and which of these scientific data and accompanying metadata will be shared. If Plan revisions are necessary (e.g., new scientific direction, a different data repository, or a timeline revision), Plans should

be updated by researchers and reviewed by the NIH ICO during regular reporting intervals or sooner. Plans from NIH-funded or conducted research may be made publicly available and should not include proprietary or private information.¹⁴

Plan Elements: NIH has developed Supplemental Information to the NIH Policy for Data Management and Sharing: Elements of an NIH Data Management and Sharing Plan that describes recommended elements to address in Plans.

Plan Assessment: The NIH ICO will assess the Plan, through the following processes:

- - Extramural Awards: Plans will undergo programmatic assessment by NIH as determined by the proposed NIH ICO. NIH encourages potential awardees to work with NIH staff to address any potential questions regarding Plan development 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 in a manner determined to be appropriate by the Intramural Research Program.
 - Other funding agreements: Plans will be assessed in the context of other funding agreement mechanisms (e.g., Other Transactions).

Section VII. Managing and Sharing Scientific Data

NIH expects that in drafting Plans, researchers will maximize the appropriate sharing of scientific data, acknowledging certain factors (i.e., legal, ethical, or technical) that may affect the extent to which scientific data are preserved and shared. Any potential limitations on subsequent data use should be communicated to individuals or entities (e.g., data repository managers) that will preserve and share the scientific data. The NIH ICO will assess whether Plans appropriately consider and describe these factors.

Considerations for Scientific Data Derived from Human Participants: 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 govern research involving human participants and the sharing and use of scientific data derived from human participants. NIH also respects Tribal sovereignty in the absence of written Tribal laws or polices. The DMS Policy is consistent with federal regulations for the protection of human research participants and other NIH expectations for the use and sharing of scientific data derived from human participants, including the [NIH's 2014 Genomic Data Sharing \(GDS\) Policy](#), [2015 Intramural Research Program Human Data Sharing Policy](#), and [45 CFR 46](#). Researchers proposing to generate scientific data derived from human participants should outline in their Plans how privacy, rights, and confidentiality of human research participants will be protected (i.e., through de-identification, [Certificates of Confidentiality](#), and other protective measures).

NIH strongly encourages researchers to plan for how data management and sharing will be addressed in the informed consent process, including communicating with prospective participants how their scientific data are expected to be used and shared. Researchers should consider whether access to scientific data derived from humans, even if de-identified and lacking explicit limitations on subsequent use, should be controlled.

Data Repository Selection: NIH strongly encourages the use of established repositories to the extent possible for preserving and sharing scientific data.¹⁵ The Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research assists researchers in selecting a suitable data repository(ies) or cloud-computing platform.

Data Preservation and Sharing Timelines: Shared scientific data should be made accessible as soon as possible, and no later than the time of an associated publication, or the end of performance period, whichever comes first. Researchers are encouraged to consider relevant requirements and expectations (e.g., data repository policies, award record retention requirements, journal policies) as guidance for the minimum time frame that scientific data should be made available, which researchers may extend.

Section VIII. Compliance and Enforcement

During the Funding or Support Period

During the funding period, compliance with the Plan will be determined by the NIH ICO. Compliance with the Plan, including any Plan updates, may be reviewed during regular reporting intervals (e.g., at the time of annual Research Performance Progress Reports (RPPRs)).

- - 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, 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 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 NIH 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)).

[1] See also NIH Rigor and Reproducibility efforts at <https://www.nih.gov/research-training/rigor-reproducibility>

[2] Compiled Public Comments on a DRAFT NIH Policy for Data Management and Sharing and Supplemental DRAFT Guidance (February 2020) https://osp.od.nih.gov/wp-content/uploads/RFI_Final_Report_Feb2020.pdf

[3] Attachment A – NIH Data Sharing Policy (September 2020) <https://www.hhs.gov/ohrp/sachrp-committee/recommendations/august-12-2020-attachment-a-nih-data-sharing-policy/index.html>

[4] For an example of NIH-supported or -stewarded repositories see Open Domain-Specific Data Sharing Repositories (September 2020) https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html

[5] NIH Rigor and Reproducibility <https://www.nih.gov/research-training/rigor-reproducibility>

[6] Wilkinson, M., Dumontier, M. et al, The FAIR Guiding Principles for Scientific Data Management and Stewardship (March 2016) <https://www.nature.com/articles/sdata201618>

[7] NIH Grants Policy Statement 2.3.11 Availability and Confidentiality of Information (October 2019) https://grants.nih.gov/grants/policy/nihgps/html5/section_2/2.3.11_availability_and_confidentiality_of_information.htm

[8] NIH Strategic Plan for Data Science (June 2018) https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf

Inquiries

Please direct all inquiries to:

NIH Office of Science Policy
SciencePolicy@od.nih.gov

Does Consent Bias Research?

Mark A. Rothstein, University of Louisville School of Medicine
Abigail B. Shoben, The Ohio State University

Researchers increasingly rely on large data sets of health information, often linked with biological specimens. In recent years, the argument has been made that obtaining informed consent for conducting records-based research is unduly burdensome and results in "consent bias." As a type of selection bias, consent bias is said to exist when the group giving researchers access to their data differs from the group denying access. Therefore, to promote socially beneficial research, it is argued that consent should be unnecessary. After analyzing the biostatistics evidence and bioethics arguments, the article concludes that (1) claims about the amount of consent bias are overstated; (2) commonly used statistical methods usually can account for consent bias; and (3) any residual effects of consent bias are below an acceptable level of imprecision and constitute a reasonable social cost for conducting ethically responsible research.

Keywords: research, selection bias, consent bias, and informed consent

Since the 1947 publication of the Nuremberg Code, informed consent has been the foundational doctrine for the ethical conduct of research with human subjects. The first principle of the Nuremberg Code begins: "The voluntary consent of the human subject is absolutely essential" (Nuremberg Code 1947). In the last several years, the nature of biomedical research has expanded from mostly small-scale, single-site interventional studies to include larger scale, multisite analyses of biospecimens and health records. Some commentators, as well as a report by the Institute of Medicine (IOM; Institute of Medicine 2009), have argued that the original concern of the Nuremberg Code, as well as the Declaration of Helsinki (World Medical Association 2004) and Common Rule (45 CFR Part 45, subpart A), safeguarding the welfare of research subjects, is inapplicable to this newer type of research. In particular, they argue that requiring informed consent for noninterventional studies is burdensome (in effort, time, and cost), hinders enrollment of adequate numbers of research subjects, and skews results due to "consent bias."

Consent bias, a type of selection bias, exists when informed consent for research, especially information-based research, results in an unrepresentative sample.

Selection bias occurs if the individuals who give permission for researchers to access their medical data differ from the group of individuals who are unwilling to give permission for their health information to be used in research. (Institute of Medicine 2009, 209)

Selection bias can lead to inaccurate results and reduced generalizability of research results to the general population (Institute of Medicine 2009, 209).

The authors are indebted to the following individuals for helpful comments on an earlier draft: Rebecca Andridge, Kyle Brothers, Ellen Clayton, and Eric Meslin.

Address correspondence to Mark A. Rothstein, JD, Herbert F. Boehl Chair of Law and Medicine, Director, Institute for Bioethics, Health Policy and Law, University of Louisville School of Medicine, 501 East Broadway #310, Louisville, KY 40202, USA. E-mail: mark.rothstein@louisville.edu

This article analyzes the magnitude and effects of consent bias in noninterventional research. It does not address the other criticisms of informed consent—regulatory burdens and adequacy of accrual of subjects. In particular, the article reviews the literature dealing with consent bias, considers the effectiveness of statistical techniques to minimize and account for bias, and explores consent bias in the larger context of ethical imperatives of responsible research with human subjects. The article concludes that (1) claims about the degree of consent bias are overstated; (2) commonly used statistical methods usually can account for all forms of selection bias; and (3) any residual effects of bias are below an acceptable level of imprecision and constitute a reasonable social cost for conducting ethically responsible research.

THE IMPORTANCE OF INFORMED CONSENT

Informed consent and other elements of modern research ethics implicitly recognize that beneficial scientific ends do not justify oppressive means. Scientific discoveries are simply not worth pursuing if they entail the unethical treatment of research subjects. Specifically addressing research, Hans Jonas stated:

Let us also remember that a slower progress in the conquest of disease would not threaten society, grievous as it is to those who have to deplore that their particular disease be not yet conquered, but that society would indeed be threatened by the erosion of those moral values whose loss, possibly caused by too ruthless a pursuit of scientific progress, would make its most dazzling triumphs not worth having. (Jonas 1969, 245)

By enacting the Common Rule in the United States and comparable laws and regulations in other countries, the leading nations performing biomedical research have acknowledged the importance of protecting human research subjects. The various laws and regulations unquestionably impose additional burdens on researchers, but these are considered by the populace and polity to be necessary and acceptable. Although rigorous efforts are needed to minimize the compliance burden on researchers (Emanuel and Menikoff 2011; Lo and Barnes 2011), reasonable regulatory requirements cannot be traded off against the potential benefits of research—they are prerequisites to research.

The Belmont Report (National Commission 1979) indicated that informed consent is a key application of the principle of respect for persons. More recently, informed consent has been considered as an application of respect for autonomy.

Respect for autonomy obligates professionals in health care and research involving human subjects to disclose information, to probe for and ensure understanding and voluntariness, and to foster adequate decision making. As some contemporary Kantians have argued, the demand that we treat others as ends requires that we assist them in achieving their ends and foster their capacities as agents, not merely that we avoid treating them solely as means to our ends. (Beauchamp and Childress 2013, 107)

Respect for autonomy is more than deference to the decision making of the research subject. It is an assurance of nonexploitation, an affirmation of human dignity, and a conferring of status on research subjects in a relationship often characterized by a real or perceived imbalance of status. Autonomy is also widely recognized by the public as an important element of research with human subjects. Extensive survey research confirms the importance of autonomy in individuals' opinions about the research use of their health information and biospecimens (Goldenberg et al. 2009; Kaufman et al. 2009; Westin 2007). A substantial majority of individuals object to the use of their health information and specimens in research without their consent. Most individuals say they would consent to have their health information and specimens used for research, including for unspecified future research (McQuillan and Porter 2011), but they want to be asked, even if the research is with de-identified materials (Hull et al. 2008; Rothstein 2010). These findings indicate widespread public support for broadly applicable informed consent rooted in autonomy.

STUDIES OF CONSENT BIAS

Sources of Consent Bias

Selection bias can occur if those included in a study are different from those who are not. If the association between the exposure (predictor of interest) and the outcome is different in the selected group than in the target population, selection bias occurs (Rothman, Greenland, and Lash 2008, 134–137). Bias in this context, and throughout this article,

refers to statistical bias—systematic differences in the estimate of interest from its true value in the population.

This article focuses on consent bias, a type of selection bias caused by certain participants refusing consent. For example, consider a study of oral cancer among smokeless tobacco users. If the study deliberately only seeks consent from males, the study may be prone to sample selection bias, because a small percentage of smokeless tobacco users are female. If the study deliberately only seeks consent from older adults who used smokeless tobacco as teenagers, the study may be prone to survivorship selection bias because some teenage users will have died before consent can be obtained (more males than females will have died due to shorter life expectancy among males). If the study asks for consent among all users, but only males consent (perhaps due to greater stigma among female users), then the study may be prone to consent bias. There is an underrepresentation of one group (females), but the question is whether that underrepresentation causes bias. There may be an underrepresentation of baseball fans, blue-eyed people, chocolate lovers, or any other individuals with an unassociated trait without any resulting bias to the study. Whether any bias is actually present depends on whether the association between smokeless tobacco and oral cancer is different between males and females, a fact that will be explored further in the next section.

Hernan, Hernandez-Diaz, and Robins (2004) showed that the same causal structure underlies all forms of selection bias. In particular, selection bias occurs if selection status (being included in the study) is associated with both the exposure (or cause of the exposure) and the outcome (or cause of the outcome). In the oral cancer example, underrepresentation of females would bias the result only if being female were also associated with the outcome (oral cancer).

To illustrate the extreme conditions necessary for large consent bias to occur, we consider two simple examples. In the first, researchers are only interested in estimating population prevalence, and thus there is only an outcome and not an exposure. In the second, researchers are interested in estimating an association between an exposure and an outcome.

For the first example, we consider a study intended to estimate the prevalence of taking a daily calcium/vitamin D supplement. The researchers plan to ascertain prevalence by looking at the medical records of patients in a large practice. If individuals who read about current medical studies are more likely to consent to their records being included in the study, then some consent bias may occur—but only if these individuals are also more likely to take the daily supplement. Consent bias occurs because the researchers can only estimate the proportion of people taking the supplement among those who consented, and those who take the supplement are more likely to consent than those who do not. If consenters are exactly like nonconsenters with respect to taking the supplement, then there is no bias, even if consent rates are low. The likely size of any bias and possible ways to reduce the bias are discussed further in the following section.

Table 1. No consent needed; True odds ratio (OR): 1.00

	Exercised	No exercise	Total
Donated blood	50	50	100
Did not donate	150	150	300
Total	200	200	400

In the second example, researchers plan a study to see whether individuals who donate blood are also more likely to exercise daily. Specifically, they will look for an association between donating blood in the previous year (the exposure) and reporting daily exercise (the outcome). For our example, assume there is no true association between blood donation and exercise. If no consent is required, the true odds ratio is 1.00,¹ reflecting no association (Table 1).

If consent is required, but it is associated with neither the exposure nor the outcome, there is no consent bias—the estimated odds ratio is still 1.00 (Table 2). If consent is required and it is associated with the exposure but not the outcome, there is still no consent bias. For example, even if consent rates among those who donated blood are 90% and consent rates among those who did not donate are only 50%, the estimated association between blood donation and exercise will still be unbiased because there is no association between consent and exercise (Table 3). Similarly, there is no consent bias if consent is associated with the outcome but not the exposure.

Consent bias only occurs if consent status is associated with both the exposure (blood donation) and the outcome (exercise). If those who donated blood are more likely to consent and those who exercise are more likely to consent, then three groups will be overrepresented compared to the total population: those who exercised and gave blood, those who gave blood but did not exercise, and those who did not give blood but did exercise. Only the group that did not give blood and did not exercise will be underrepresented compared to the total population. Therefore, it may falsely appear that there is a negative association between blood donation and daily exercise—blood donors are less likely to exercise—even when there is not an association in truth (Tables 4 and 5). It is important to note that two conditions are necessary for consent bias to be a potential problem: Consent status had to be associated with both the exposure (donating blood) and the outcome (exercising).

Although in the preceding examples we considered consent to be directly associated with the exposure (blood donation) and the outcome (exercise), we could have also considered more complicated pathways and reached the same

1. Throughout this article, we use odds ratios to summarize the association between two binary variables (e.g., blood donation and exercise). In this example, the odds ratio compares the odds ($p/(1-p)$) of exercising among blood donors to the odds of exercising among nondonors. An odds ratio of 1.00 indicates no association; an odds ratio less than 1 indicates smaller odds of exercising among blood donors; and an odds ratio greater than 1 indicates larger odds of exercising among blood donors.

Table 2. Consent rate 80% for everyone; Estimated OR: 1.00

	Exercised	No exercise	Total
Donated blood	40	40	80
Did not donate	120	120	240
Total	160	160	320

conclusion. For example, if older adults are more likely to donate blood and more likely to exercise daily, and older adults are more likely to consent, then consent is associated with both the exposure (blood donation) and the outcome (exercise) because consent is associated with age. However, the underlying principle remains the same: There needs to be an association between consent status and both the exposure and the outcome for consent bias to exist.

Both of these examples illustrate that in an extreme scenario consent bias could pose a threat to study validity. In practice, however, this bias is likely to be small, particularly if steps are taken at the design and analysis stages to minimize the possible bias. The next sections discuss these practical concerns more fully.

Magnitude of the Possible Bias

Quantifying the amount of bias possible due to various sources has been a subject of considerable interest in the biostatistics and epidemiology literature (see, e.g., VanderWeele and Arah 2011 and references therein). Greenland (2003) noted that selection bias due to consent is similar to classical confounding bias and therefore existing literature can be used to estimate the magnitude of the possible bias. The magnitude of the bias will be affected by the associations between consent status and the exposure of interest and consent status and the outcome of interest, as well as the probability of consent. Yanagawa (1984) derived mathematical bounds on the amount of possible bias due to a confounding factor in case control studies. For the bias to be anything other than small, the association between consent status and both the exposure and the outcome must be quite large—on the order of an odds ratio of 4 (or $\frac{1}{4}$) or more extreme. For this to be true in our example, the odds of exercising daily among those who consented would have to be at least 4 times that of those who did not consent, and similarly the odds of donating blood would have to be at

Table 3. Consent rate differs by donation status but not exercise (90% for donors, 50% for nondonors); Estimated OR: 1.00

	Exercised	No exercise	Total
Donated blood	45	45	90
Did not donate	75	75	150
Total	120	120	240

Table 4. Consent rate differs by donation status (90% for donors, 70% for nondonors) and Exercise Status (90% in Exercisers, 60% in Nonexercisers); Estimated OR: 0.56

	Exercised	No exercise	Total
Donated blood	45	45	90
Did not donate	135	75	210
Total	180	120	300

least 4 times that of those who did not consent. Such strong associations have simply not been seen in existing studies of possible consent bias (Kho et al. 2009). Similar studies done in other observational settings have reached similar conclusions about the magnitude of associations needed for substantial bias (see, e.g., Bross 1967; Lin, Psaty, and Kronmal 1998; Schlesselman 1978).

To see this effect, we return to the example of looking for an association between blood donation (the exposure) and exercise (the outcome). For our first hypothetical, we assume large associations between consent and the exposure (blood donation) and consent and the outcome (exercise). We assume among blood donors 90% consent, and among nondonors 70% consent, which means that blood donors have 3.86 times the odds of consenting as nondonors.² We assume an even larger association between consent and the outcome (exercise), with a 90% consent rate among those who exercise and a 60% rate among those who do not, which means that those who exercise have 6.0 times the odds of consenting as those who do not.³ Given these extremely strong associations, we do see substantial bias in the estimated odds ratio for the association of interest (blood donation with exercise)—among those who consent the odds ratio is 0.56, when in truth it should be 1.00 (Table 4).

However, given more moderate associations between consent and the exposure and consent and the outcome, the magnitude of the bias is much smaller. For example, assume that the odds of consent among those who donate blood are 1.89 times those who do not (85% among donors, 75% among nondonors) and that the odds of consenting among those who exercise are 1.79 times those who do not (82.5% and 72.5%, respectively). Under these assumptions, the estimated odds ratio for the association of interest among those who consent is 0.98, which is very nearly what it would be if everyone consented (Table 5).

Although consent bias is possible if consent is associated with both the exposure and the outcome, the magnitude of this bias is likely to be small in practice, given existing literature on the size of associations between consent and various demographic variables (Kho et al. 2009).

2. The odds of consent among blood donors are $0.9/(1 - 0.9) = 9$, and the odds of consent among nondonors are $0.7/(1 - 0.7) = 2.33$, so the odds ratio comparing the odds of consent of donors to nondonors is $9/2.33 = 3.86$.

3. Similarly, the odds of consent among those who exercise are $0.9/(1 - 0.9) = 9$, and the odds of consent among those who did not exercise are $0.6/(1 - 0.6) = 1.5$, so the odds ratio is $9/1.5 = 6.0$.

Table 5. Consent rate differs by donation status (85% for donors, 75% for nondonors) and exercise status (82.5% in exercisers, 72.5% in nonexercisers); Estimated OR: 0.98

	Exercised	No exercise	Total
Donated blood	45	40	85
Did not donate	120	105	125
Total	165	145	310

We note here that some previous work attempting to quantify the amount of consent bias has been in studies seeking retrospective consent. In these studies, the authors have lumped together both consent bias due to people refusing consent and other forms of selection bias. For example, an article frequently cited as an example of the perils of consent bias (Al-Shahi, Vousden, and Warlow 2005) found a difference in the association of interest when the study was restricted to those who provided retrospective consent. However, this study attempted to obtain consent from individuals after diagnoses of a cranial malformation and the investigators could not approach those who were (a) already deceased or (b) deemed unfit for consent by the treating physician due to anxiety about diagnosis or mental impairment. Unsurprisingly, these restrictions led to a very different study sample when restricted to those who gave consent. Nevertheless, calling all of the selection bias in this study "consent bias" is misleading, because not giving consent due to death or restriction by the treating physician is not the same as asking for, but being refused, consent from competent individuals.

Another study by Macleod and Watt (2008) showed a similar change in results related to breast cancer survival when individuals were contacted after the study for retrospective consent. Despite this survivorship bias, Macleod and Watt noted that most of the original findings (using both those who consented and those who did not) were consistent in the consent-only group. Only findings related to the absence of women with advanced or metastatic cancer at the time of diagnosis (who were too ill to consent or deceased) were different. Thus, even in a situation with very poorly timed consent—perhaps the worst possible scenario for "consent" bias—many of the associations were consistent even after restricting to the consent-only sample.

Finally, in large-database studies, the effects of interest are often small and the power to detect small effects is often large. Thus, large databases create two problems: (1) A small observed "effect" could be due to selection bias, including possibly consent bias; and (2) an effect due entirely to selection bias (no true association) could be statistically significant in a large enough sample. However, these concerns are not unique to consent bias. Large-database studies of small effects are subject to a multitude of possible concerns, including unmeasured confounding, ascertainment bias, measurement error, confounding by indication, and many others (Grimes and Schulz 2002). Although some concern due to consent bias may be warranted, it is just one of

many concerns in this context, and focusing only on consent bias is unlikely to improve the statistical validity of the study.

Factors Associated With Consent

Most existing studies on consent bias have focused on identifying factors associated with consent. As noted by Kho and colleagues (2009), there does not appear to be a consistent pattern of associations with willingness to consent, although few studies have been conducted in the United States and more research is warranted in this area. Age has been particularly perplexing. In some studies, younger age is associated with higher consent rates (Dunlop et al. 2007; Tu et al. 2004; Woolf et al. 2000), whereas in others older age is associated with higher consent rates (Beebe et al. 2007; Damery et al. 2011). Particularly interesting is the age–sex interaction noted by Dunn and colleagues (2004), which showed a strong age trend in men (older men are more likely to consent), but much more constant rates among different age groups for women. This trend is generally consistent with the differences in the studies previously noted and suggests that researchers should strongly consider looking for age–sex interactions in future studies on consent.

Other factors that studies have been found to be associated with higher consent rates include higher income (Damery et al. 2011; Dunlop et al. 2007), having the health condition/symptom being studied (Dunn et al. 2004), healthier lifestyle (Buckley et al. 2007), higher education (Beebe et al. 2007; Dunlop et al. 2007; Woolf et al. 2000), and white race (Armstrong et al. 2005; Damery et al. 2011; Tu et al. 2004; Woolf et al. 2000).

Willison and colleagues (2009) surveyed individuals about what types of consent should be required for different types of hypothetical studies. The study found no difference in consent preferences by disease, but did find differences based on the type of study, with highest trust (lowest consent requirements) for hospital/doctor-initiated quality improvement studies and lowest trust (strictest consent requirements) in for-profit studies or studies linking health information with work/education/income information. The authors argue that a variety of consent options should be presented to individuals selected for inclusion in future studies (some individuals may be willing to give general permission for some studies, whereas others may prefer to be contacted by individual studies). Their data also show that a sizeable proportion of individuals are “privacy sensitive” and thus they argue that it may be difficult to capture the opinions of these individuals even in studies on consent practice.

MEASURES TO MINIMIZE BIAS

Study Design

Although we have indicated that the magnitude of possible consent bias is likely to be small in practice, steps can be taken at the design, recruitment, and analysis stages to further minimize possible bias due to the need for informed

consent. Some authors (Al-Shahi, Vousden, and Warlow, 2005) have argued that one consequence of the need for consent is a loss of statistical power. The obvious solution to this potential problem is to plan studies larger than necessary to account for less than 100% consent rates. This approach is analogous to other epidemiological studies and clinical trials in which recruitment is not 100% (Rothman, Greenland, and Lash 2008).

As discussed earlier, a key component of any possible consent bias is a link between the exposure and outcome of interest and consent status. Without an association, consent bias cannot occur. Thus, in the design stage researchers should anticipate possible factors associated with consent and whether these factors may be associated with the variables of interest. To this end, further research into factors influencing individual decisions to consent for various types of studies (e.g., medical records, tissue samples, surveys, and clinical trials) would be helpful to future investigators.

For those studies in which investigators are concerned about strong associations between consent and their variables of interest, investigators might spend more time planning the recruitment/consent stage of the research. In addition to methods to improve recruitment, researchers should consider whether it would be possible to allow individuals to consent to minimal participation, even if they refuse full consent, so that the researchers can collect some basic information on even those who refuse consent. Minimal information on those who refuse full consent would allow researchers to use advanced statistical techniques to correct for possible bias. Methods for improved recruitment and statistical adjustment are discussed next.

Recruitment

One obvious protection against selection bias is to improve study participation rates. Participation in medical studies from historically underrepresented groups has been a subject of considerable research, particularly in the context of randomized clinical trials (Shavers, Lynch, and Burmeister 2002; Wendler 2011; Wendler et al. 2006; Williams and Corbie-Smith 2006). Much of this research is likely applicable in the setting of requesting consent for medical record research. For example, Dunlop and colleagues (2007) examined the impact of including a HIPAA (Health Insurance Portability and Accountability Act) form on factors affecting the decision of African Americans in Atlanta, GA, to participate in a hypothetical clinical trial. In a follow-up study, Dunlop and colleagues (2011) developed an 11-minute DVD that addressed frequently cited concerns and barriers to participation, notably including mistrust/fear of the research process. Participants randomized to watch the DVD before deciding about participation in the hypothetical study were much more likely to agree to participate than controls who did not watch the DVD. Such a simple, relatively easily implemented intervention would likely have a large effect on consent rates, particularly for medical record studies. In the presence of increased HIPAA requirements, research into improved consent rates would be broadly applicable and

particularly useful for studies concerned about the association between consent and variables of interest.

Large databases of tissue samples or large medical record databases underscore the need for organized consent processes. As discussed by Willison and colleagues (2009), under institutional review board (IRB) guidance, researchers could ask potential participants once about a variety of possible studies and participants could choose to release various amounts of their data. For example, one individual might choose to consent to all future studies using any part of that individual's medical record, another might choose to be contacted for each specific study, and a third might choose to withhold consent from all research studies.

An additional, obvious recruitment technique is simply to engage in better study practices. Several papers on consent bias and factors associated with consent admit to a lack of follow-up contact with those who do not respond to an initial mailing or telephone contact (Al-Shahi, Vousden, and Warlow 2005; Armstrong et al. 2005). Such administrative lack of follow-through may be acceptable in studies where consent status is not associated with the variables of interest, but it is inexcusable in settings where consent bias is of possible concern. Researcher inconvenience and poor study practices are not acceptable reasons to dispense with informed consent. High consent rates for the use of medical records are likely with sound research practices.

Data Analysis

Once the data have been obtained, various methods exist to estimate the sensitivity of any result to possible consent bias (Greenland 1996; Lin, Psaty and Kronmal 1998). The epidemiology and biostatistics literature contains several methods to estimate the robustness of a possible result to various unmeasured confounding variables (always an important issue in nonrandomized studies) and selection bias (including consent bias). Such methods cannot determine whether bias is present, but they can help estimate the magnitude of the associations that would be necessary to generate the observed effect or dampen a hypothesized effect that failed to be detected.

If some information is available on those who refused consent, other methods exist to help minimize consent bias. The most important of these is inverse probability weighting (Robins, Rotnitzky, and Zhao 1994), which uses demographic or other information available on all subjects (those who consented and those who did not) and determines the probability of an individual consenting to the study. The method then reweights individuals in the study based on the probability of consent; individuals who were likely to consent (and are thus overrepresented in the sample relative to the total population) are downweighted, and individuals who were unlikely to consent (and are thus underrepresented in the sample) are upweighted. For this method to be used, basic demographic information must be available for all subjects, including those who did not give consent to medical record review. Therefore, as noted in the study design section, researchers concerned about consent bias

should plan studies to ask for limited consent to basic information from individuals who refuse general consent to the study.

For a simple illustration of this technique, we again return to the example of a study trying to estimate the prevalence of individuals who take a daily vitamin D supplement. In our hypothetical study, we speculated that individuals who read current medical literature were more likely to consent and more likely to take the daily supplement. First, we note again that a difference in consent rates and supplement status by reading the medical literature is unlikely to cause a dramatic difference in the estimated prevalence, but we continue for the sake of illustrating inverse probability weighting. In this study, individuals would be contacted and asked both for consent to medical record review and for consent to a short questionnaire that asked about their familiarity with the medical literature (or a surrogate of this variable such as educational attainment). For the sake of this example, we assume that more individuals will consent to the short questionnaire than consent to full record review. Among those in the subset that consented to record review, investigators would know whether they took the daily supplement. Instead of simply calculating the prevalence of supplement use within the group that consented, investigators would upweight individuals who consented but did not read the medical literature and downweight those who consented and did read the medical literature, so that the proportion of those in the subsample who read the medical literature matches that in the original larger sample. The method of inverse probability weighting can handle situations much more complicated than the simple example just described, such as selection bias due to consent, lost to follow-up, and death (Haneuse et al. 2008).

In large databases, the method of inverse probability weighting is particularly promising as a way to ameliorate possible consent bias. This method would be possible if large biobanks or medical record repositories implemented tiered consent processes, such that individuals could consent to researchers having access to basic demographic information (age, sex, etc.) while withholding consent for more sensitive information (such as family income or disease history).

One potential objection to our analysis would be that even if inverse probability weighting and other statistical techniques are useful in hypothesis-testing research, they would not work for biobanks and other repositories that are hypothesis-generating because rare mutations or conditions would not be observed. The rejoinder is that the rare mutations issue is fundamentally a concern about sample size, as the probability of sampling people with rare mutations increases with larger sample size. Although dispensing with informed consent will increase sample size, it is not the only or preferred way to do so. Performing research without consent is a high price to pay in attempting to increase sample size and control for any slight consent bias.

The available data on nonparticipation rates (individuals who opt out or fail to opt in) in large biobanks indicate

that sample size and representativeness are not compromised by a consent mechanism. The opt-out rate for Vanderbilt's DNA Databank (BioVU) is 5% (Pulley et al. 2010); the opt-out rate for the Texas Newborn Bloodspot Repository is 3.5%⁴ (Roser 2009); and the non-opt-in rate for Duke University's Biobank is 2% (Jeffrey M. Stajich, Center for Human Genetics, Duke University, personal communication 2010). All of these nonparticipation rates are well below levels that would threaten the statistical validity of any research using health records and biological samples.

Quite apart from the matter of statistical validity of research there are legal (e.g., National Institutes of Health [NIH] requirements) and ethical reasons why any study sample ought to be representative of the population. These important issues are discussed next; however, concerns about consent bias in the literature overwhelmingly are expressed in terms of preserving the scientific integrity of the research. The preceding discussion has demonstrated that statistics-based arguments about consent bias have been overstated.

POLICY CONCERNs ABOUT CONSENT BIAS

Although the preceding sections of this article have demonstrated that the requirement of consent does not inevitably result in irremediable selection bias, concerns about consent bias increasingly are being raised in the literature (Al-Shahi, Vousden, and Warlow 2005; Hoffman and Podgurski 2012; Kho et al. 2009; Woolf et al. 2000). The Common Rule and other research regulations do not specifically address the issue of consent bias. The possibility of consent bias, however, is inherent in the concept of informed consent. Only compulsory enrollment can eliminate completely the possibility of any consent bias, but such an approach would violate the fundamental principle of ethical research that the informed consent of the research subject is absolutely essential.

In 2009, the Institute of Medicine (IOM) issued a report on privacy in health research that recommended eliminating informed consent for all research other than "interventional" or clinical research (Institute of Medicine 2009). The recommendation was based on the assumption that "informational" research, even with individual identifiers, does not present the same degree of risks to research subjects as interventional research. Therefore, it asserted that utilitarian concerns militate in favor of abandoning informed consent for research involving health information and biological specimens.

If society seeks to derive the benefits of medical research in the form of improved health and health care, information should be shared for the greater good, and governing regulations should support the use of such information, with appropriate oversight. (Institute of Medicine 2009, 35)

4. In the first 5 months of its opt-out strategy, May 27, 2009, to November 2, 2009, of 240,000 newborns in Texas, the parents of 6,900 opted out. As of June 1, 2012, Texas law mandates opt-in consent for the release of individually identifiable newborn screening blood spots for research (Texas H.B. 411 2012).

The IOM report is subject to numerous criticisms, including the following: (1) It underestimates the risks to individuals of "mere" informational research; (2) it fails to justify abandoning the bedrock ethical principle of informed consent for research; (3) it overvalues the interests of researchers; and (4) it overlooks the betrayal of trust of patients who, involuntarily, would become research subjects (Rothstein 2009).

The IOM report also asserted that

a universal requirement of informed consent can lead to invalid results because of significant differences between patients who do or do not grant consent, and to missed opportunities to advance medical science because it can be prohibitively costly and difficult to obtain consent for studies that require analysis of very large datasets. (Institute of Medicine 2009, 35)

The IOM report may have overestimated the burdens of obtaining consent for research involving large data sets, and it failed to make a compelling case that consent leads to irremediable harms associated with consent bias. The IOM report put great weight on surveys of researchers who, unsurprisingly, reported that they believe consent requirements burden and bias research; however, by relying on the opinions of self-interested researchers the authors of the IOM report have relied on responses tainted by self-serving bias.

One possible response to our criticism of consent bias is that because it is more difficult to recruit minority populations into research, these populations will be underweighted in studies, thereby resulting in a continuation or even an aggravation of health disparities. Although the weighting techniques discussed earlier are frequently used to address this problem, the argument requires a more fundamental response. The relationship between consent bias and health disparities is based on the historical abuses in research on vulnerable and minority populations that made these groups suspicious of researchers (Skloot 2010) and led to lower enrollment rates. Therefore, to increase enrollment and benefit these previously exploited populations, some scholars and members of the research community have asserted it is justifiable to dispense with the basic right of informed consent. We believe that eliminating consent simply compounds the indignity to members of these groups.

The consent bias rationale for limiting informed consent, in effect, endorses mandatory participation in research without any notice to the research subjects. Efforts to eliminate all informed consent for informational research represent paternalism and expediency masquerading as beneficence, utilitarianism, and scientific rigor. The consent bias argument assumes there is no other way to reduce the effects of selection bias, a claim refuted by the previous methodology sections.

Informed consent has been one of the most highly scrutinized elements of research ethics. Some provocative articles have suggested reconsidering informed consent. One line of argument asserts that everyone has a duty to serve as a research subject (Harris 2005; Rhodes 2008; Schaefer, Emanuel, and Wertheimer 2009) and that it is sometimes desirable and ethical to make research participation a

requirement for treatment (Orentlicher 2005). Nevertheless, such a universal moral obligation would benefit mostly “researchers, research institutions, public and private research funding agencies, and pharmaceutical companies” (Rennie 2011).

In 2011, the U.S. Department of Health and Human Services published an Advance Notice of Proposed Rulemaking (DHHS 2011) in which it proposed possible sweeping changes to the federal regulations on research with human subjects, including informed consent. Among the proposals is extending the requirement of informed consent to research with de-identified biospecimens, thereby reaffirming the significance of informed consent to research ethics and recognizing the link between informed consent and public support for large-scale health research. The issues of whether individuals ought to consent to research and possible amendments to the federal research regulations are beyond the scope of this article.

ACCEPTABLE LEVEL OF IMPRECISION

Despite the best efforts of many brilliant, hard-working, and dedicated investigators, biomedical research often is an imprecise enterprise (Ioannidis 2011). Besides its spectacular triumphs, the landscape of research is strewn with retracted papers, refuted theories, unreplicated findings, recalled or withdrawn drugs and medical devices, adverse events, and tragic deaths of research subjects. Occasionally, research has been compromised by conflicts of interest or scientific misconduct. Bias and confounding factors are seemingly everywhere. One recent review identified 235 types of bias in biomedical research, including ascertainment bias, confounding bias, cultural bias, observer bias, perceptual bias, publication bias, recall bias, reporting bias, response bias, and selection bias (Chavalarias and Ioannidis 2010).

In light of these limitations on biomedical research, it is incongruous for some researchers and commentators to attempt combating only one type of bias—consent bias—by urging revocation of the most fundamental right of research subjects. Although statistical methods and sound research practice can help ameliorate the magnitude and effects of selection bias, these measures are unlikely to remove all traces of bias. Researchers using human subjects will never have the same control as they do with *in vitro* studies or research with laboratory animals.

Scientific progress is a worthy goal, but it is not the only interest at stake in biomedical research. Among other interests, ensuring justice, respecting autonomy, and promoting human dignity are integral elements of ethical research with human subjects. New advances are not worth pursuing at any cost. The scientific community and science policymakers should recognize the concept of an acceptable level of imprecision. Even as ongoing efforts are needed to minimize imprecision, it should be acknowledged that perfection in research methodology has never been achieved in the past and cannot be achieved in the future because of the fundamental ethical requirements applicable to research with human subjects. Recent scandals involving the misuse of

health records or specimens in research—from the Havasupai Tribe to the Texas Newborn Screening Repository—were not caused by biased samples, but they were caused, at least in part, by inadequate consent.

To some readers, it might appear to be inconsistent to advocate for a relaxed standard of statistical rigor while at the same time opposing a relaxed standard for informed consent. The “imprecision” we are prepared to accept is the slight statistical remainder after using appropriate statistical methods to reduce the effects of selection bias; the imprecision is within statistically acceptable limits. Moreover, the acceptability of slight imprecision attributable to selection bias must be viewed in light of the numerous other technical and practical limitations that make the notion of obtaining bias-free data unrealistic. By contrast, we reject the claim that concerns about consent bias justify abandoning consent entirely for all noninterventional research. Although beyond the scope of this article, we are certainly receptive to modifications to the informed consent process to improve efficiency and reduce the burdens on researchers.

CONCLUSION

Virtually all research is biased to some extent. The requirement of informed consent, like other protections for human subjects, has a degree of inherent bias. The concept of consent bias is not a new insight; what is new is the claim that it constitutes a justification for dispensing with informed consent in one or more types of research. After reviewing the literature, we conclude that claims about the amount of consent bias are overstated, that the complained-of bias could be minimized through sound research methodology, and that any remaining bias would be within the bounds of generally accepted statistical measures. Our fundamental ethical observation is that a slight level of imprecision in research is an acceptable price to pay for conducting ethically appropriate research.

The argument that informed consent is incompatible with modern research represents an assault on the societal values on which biomedical research is based. The consent bias attack on traditional notions of research ethics is an explicit repudiation of Kant’s categorical imperative to treat human beings as ends and not as means.

As persons, research subjects possess an inviolability that rules out treating them as mere means to the ends of others, including others who may be suffering from a disease or in need of medical care. (Presidential Commission 2011, 70)

In practical terms, the assertion of consent bias is a self-fulfilling prophecy. Public trust in research affects participation rates, with greater effects in some population groups. Efforts to permit research without informed consent will decrease trust and thereby increase the likelihood of a smaller and less representative sample.

Informed consent, the first principle of ethical research with human subjects, was the world’s response to the appalling medical experiments in the Nazi death camps. The

primacy of informed consent was reaffirmed in the United States to deal with the outrageous mistreatment of vulnerable research subjects ranging from rural Alabama share-croppers (Jones 1993) to inmates in prisons and patients in mental hospitals (Beecher 1966). In light of this history, informed consent—in any context—should not be discarded or limited without a compelling justification.

Some commentators on consent bias who advocate easing consent requirements argue that codes of research ethics, from the Nuremberg Code to the Common Rule, were only designed to address the risk of physical injuries to human subjects. Thus, in their view, intangible infringements on autonomy, privacy, dignity, and similar interests are tangential and may be properly balanced against the societal interest in scientific discovery. No less an authority than Jay Katz flatly rejected this line of reasoning. “I want to emphasize, however, the centrality of dignity, not physical injuries, in any appraisal of the ethics of research” (Katz 1996, 545).

The realities of modern health care make it virtually certain that every individual will be required to seek care from a health care provider utilizing increasingly integrated and interoperable health records. Although most individuals will, and perhaps even more ought to, consent to the use of their health records in research, abolishing the requirement of consent for research is the equivalent of involuntarily removing the informational hospital gown of every patient.

Informed consent for research also should be considered in the larger context. The same technological pressures for unimpeded research—including large-scale, high-throughput analyses, computerized interrogation of huge databases, and health information technology linking health records and biospecimens—also underscore the need for maintaining the consent process. New technology has the potential to invade privacy on an unprecedented scale. Privacy and autonomy are under a relentless attack in modern life, including by visual surveillance in public places, global positioning system (GPS) tracking of individual movement, facial recognition software, data mining, personal computer utilization analysis, and synthesizing consumer credit and commercial transactions. Biomedical research, with its history of extraordinary abuses and its relationship to the fiduciary obligations of health professionals, should be positioned to resist this technological assault on privacy and autonomy rather than to join its vanguard.

Instead of abandoning the concept of informed consent for noninterventional studies, researchers and science policymakers ought to (1) redouble their efforts to overcome the historical and social factors that often lead to varied enrollment levels on the basis of race, ethnicity, age, education, income, health status, and other factors; (2) require application of sound research methodologies to minimize selection bias and other forms of bias at all stages of research, from study design to data analysis; and (3) recognize that a slight degree of imprecision is an inevitable and acceptable price to pay for conducting ethical research with human subjects. ■

REFERENCES

Al-Shahi, R., C. Vousden, and C. Warlow. 2005. Bias from requiring explicit consent from all participants in observational research: Prospective, population based study. *British Medical Journal* 331: 942. doi:10.1136/bmjj.38624.397569.68.

Armstrong, D., E. Kline-Rogers, M. J. Sandeep, et al. 2005. Potential impact of the HIPAA privacy rule on data collection in a registry of patients with acute coronary syndrome. *Archives of Internal Medicine* 165: 1125–1129.

Beauchamp, T. L., and J. F. Childress. 2013. *Principles of biomedical ethics*, 7th ed. New York, NY: Oxford University Press.

Beebe, T. J., N. J. Tally, M. Camilleri, et al. 2007. The HIPAA authorization form and effects on survey response rates, nonresponse bias, and data quality: A randomized community study. *Medical Care* 45(10): 959–965.

Beecher, H. K. 1966. Ethics and clinical research. *New England Journal of Medicine* 274: 1354–1360.

Bross, I. D. J. 1967. Pertinency of an extraneous variable. *Journal of Chronic Diseases* 20: 487–495.

Buckley, B., A. W. Murphy, M. Byrne, et al. 2007. Selection bias resulting from the requirement for prior consent in observational research: A community cohort of people with ischaemic heart disease. *Heart* 93: 1116–1120.

Chavalaris, D., and J. P. A. Ioannidis. 2010. Science mapping analysis characterizes 235 biases in biomedical research. *Journal of Clinical Epidemiology* 63(11): 1025–1025. doi:10.1016/j.jclinepi.2009.12.011

Damery, S., R. Ryan, R. J. McManus, et al. 2011. The effect of seeking consent on the representativeness of patient cohorts: Iron-deficiency anaemia and colorectal cancer. *Colorectal Disease* 13(11): e366–e373.

Department of Health and Human Services. 2011. Advance notice of proposed rulemaking, Human subjects research protections: Enhancing protections for research subjects and reducing burden, delay, and ambiguity for investigators. *Federal Register* 76: 44512–44531.

Dunlop, A. L., T. Graham, Z. Leroy, et al. 2007. The impact of HIPAA authorization on willingness to participate in clinical research. *Annals of Epidemiology* 17(11): 899–905.

Dunlop, A. L., Z. Leroy, K. M. Logue, et al. 2011. Preconsent education about research processes improved African Americans' willingness to participate in clinical research. *Journal of Clinical Epidemiology* 64(8): 872–877.

Dunn, K. M., K. Jordan, R. J. Lacey, et al. 2004. Patterns of consent in epidemiologic research: Evidence from over 25,000 responders. *American Journal of Epidemiology* 159: 1087–1094.

Emanuel, E. J., R. A. Crouch, J. D. Arras, et al., eds. 2003. *Ethical and regulatory aspects of clinical research*. Baltimore, MD: Johns Hopkins University Press.

Emanuel, E. J., and J. Menikoff. 2011. Reforming the regulations governing research with human subjects. *New England Journal of Medicine* 365: 114–1149.

Goldenberg, A. J., S. C. Hull, J. R. Botkin, et al. 2009. Pediatric biobanks: Approaching informed consent for continuing research after children grow up. *Journal of Pediatrics* 155: 578–583.

Greenland, S. 1996. Basic methods of sensitivity analysis of biases. *International Journal of Epidemiology* 25(6): 1107–1116.

Greenland, S. 2003. Quantifying biases in causal models: Classical confounding vs collider-stratification bias. *Epidemiology* 14(3): 300–306.

Grimes, D. A., and K. F. Schulz. 2002. Bias and causal associations in observational research. *Lancet* 359: 248–252.

Haneuse, S., J. Schildcrout, P. Crane, et al. 2008. Adjustment for selection bias in observational studies with application to the analysis of autopsy data. *Neuroepidemiology* 32: 229–239.

Harris, J. 2005. Scientific research is a moral duty. *Journal of Medical Ethics* 31(4): 242–248.

Hernan, M. A., S. Hernandez-Diaz, and J. M. Robins. 2004. A structural approach to selection bias. *Epidemiology* 15(5): 615–625.

Hoffman, S., and A. Podgurski. 2012. Balancing privacy, autonomy, and scientific needs in electronic health records research. *SMU Law Review* 65: 85–144.

Hull, S. C., R. R. Sharp, J. R. Botkin, et al. 2008. Patients' views on identifiability of samples and informed consent for genetic research. *American Journal of Bioethics* 8(10): 62–70.

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

Ioannidis, J. P. A. 2011. Why most published research findings are false. *PLoS Medicine* 2(8): e124. doi:10.1371/journal.pmed.0020124

Jonas, H. 1969. Philosophical reflections on experimenting with human subjects. *Daedalus* 98: 219–247.

Jones, J. H. 1993. *Bad blood: The Tuskegee syphilis experiment*. New York, NY: Free Press.

Katz, J. 1996. Statement by Committee Member Jay Katz. In *Final report of the Advisory Committee on Human Radiation Experiments*, ed., 545. Advisory Committee on Human Radiation Experiments. New York, NY: Oxford University Press.

Kaufman, D. J., J. Murphy-Bollinger, J. Scott, et al. 2009. Public opinion about the importance of privacy in biobank research. *American Journal of Human Genetics* 85: 643–654.

Kho, M. E., M. Duffett, D. J. Willison, et al. 2009. Written informed consent and selection bias in observational studies using medical records: Systematic review. *British Medical Journal* 338: b866. doi: 10.1136/bmj.b866

Lin, D. Y., B. M. Psaty, and R. A. Kronmal. 1998. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics* 54(3): 948–963.

Lo, B., and M. Barnes. 2011. Protecting research participants while reducing regulatory burdens. *Journal of the American Medical Association* 306: 2260–2261.

Macleod, U., and G. C. M. Watt. 2008. The impact of consent on observational research: A comparison of outcomes from consenters and non consenters to an observational study. *BMC Medical Research Methodology* 8:15.

MacQuillan, G. M., and K. S. Porter. 2011. Consent for future genetic research: The NHANES experience in 2007–2008. *IRB: Ethics in Human Research* 33(1): 9–14.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. *Belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Available at: http://www.fda.gov/ohrms/dockets/ac/05/briefing/2005-4178b_09_02_Belmont%20Report.pdf (accessed November 29, 2011).

Nuremberg Code. 1947. Available at: <http://ohsr.od.nih.gov/guidelines/nuremberg.html> (accessed November 22, 2011).

Orentlicher, D. 2005. Making research a requirement of treatment: Why we should sometimes let doctors pressure patients to participate in research. *Hastings Center Report* 35(5): 20–28.

Presidential Commission for the Study of Bioethical Issues. 2011. *Moral science: Protecting participants in human subjects research*. Washington, DC: Presidential Commission. Available at: www.bioethics.gov (accessed January 9, 2012).

Pulley, J., E. Clayton, G. R. Bernard, et al. 2010. Principles of human subjects protections applied in an opt-out deidentified biobank. *Clinical and Translational Science* 3(1): 42–48.

Rennie, S. 2011. Viewing research participation as a moral obligation: In whose interest? *Hastings Center Report* 41(2): 40–47.

Rhodes, R. 2008. In defense of the moral duty to participate in biomedical research. *American Journal of Bioethics* 8(10): 37–38.

Robins, J. M., A. Rotnitzky, and L. P. Zhao. 1994. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 89: 846–866.

Roser, M. A. 2010. State agrees to destroy more than 5 million stored blood samples from newborns. *Austin American-Statesman*, December 23. Available at: www.statesman.com/news/news/state-regional/state-agrees-to-destroy-more-than-5-million-stored/nRZK9 (accessed November 18, 2012).

Rothman, K. J., S. Greenland, and T. L. Lash. 2008. *Modern epidemiology*, 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins.

Rothstein, M. A. 2010. Is deidentification sufficient to protect health privacy in research? *American Journal of Bioethics* 10(9): 3–11.

Rothstein, M. A. 2009. Improve privacy in research by eliminating informed consent? IOM report misses the mark. *Journal of Law, Medicine and Ethics* 37: 507–512.

Schaefer, G. O., E. J. Emanuel, and A. Wertheimer. 2009. The obligation to participate in biomedical research. *Journal of the American Medical Association* 302: 67–72.

Schlesselman, J. J. 1978. Assessing effects of confounding variables. *American Journal of Epidemiology* 108: 3–8.

Shavers, V. L., C. F. Lynch, and L. F. Burmeister. 2002. Racial differences in factors that influence the willingness to participate in medical research. *Annals of Epidemiology* 12: 248–256.

Skloot, R. 2010. *The immortal life of Henrietta Lacks*. New York, NY: Crown.

Tu, J. V., D. J. Willison, F. L. Silver, et al. 2004. Impracticability of informed consent in the registry of the Canadian Stroke Network. *New England Journal of Medicine* 350(14): 1414–1421.

VanderWeele, T. J., and O. A. Arah. 2011. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology* 22(1): 42–51.

Wendler, D. 2011. How to enroll participants in research ethically. *Journal of the American Medical Association*. 305: 1587–1588.

Wendler, D., R. Kington, J. Madans, et al. 2006. Are racial and ethnic minorities less willing to participate in health research? *PLoS Medicine* 3(2): e19. doi:10.1371/journal.pmed.0030019

Westin, A. F. 2007. *IOM project survey findings on health research and privacy*. Available at: www.iom.edu/~/media/Files/Activity%20Files/Research/HIPAAandResearch/AlanWestinIOMsrvyRept.ashx (accessed November 29, 2011).

Williams, I. C., and G. Corbie-Smith. 2006. Investigator beliefs and reported success in recruiting minority participants. *Contemporary Clinical Trials* 26(6): 580–586.

Willison, D. J., V. Steeves, C. Charles, et al. 2009. Consent for use of personal information for health research: Do people with potentially stigmatizing health conditions and the general public differ in their opinions? *BMC Medical Ethics* 10: 10.

Woolf, S. H., S. F. Rothenmich, R. E. Johnson, et al. 2000. Selection bias from requiring patients to give consent to examine data for health services research. *Archives of Family Medicine* 10: 1111–1118.

World Medical Association. 2004. *Declaration of Helsinki: Ethical principles for medical research involving human subjects*. Available at: <http://ohsr.od.nih.gov/guidelines/helsinki.html> (accessed November 22, 2011).

Yanagawa T. 1984. Case-control studies: Assessing the effect of a confounding factor. *Biometrika* 71(1): 191–194.

THE REGULATION OF RESEARCH INVOLVING HUMAN PARTICIPANTS

Kristen B. Rosati*

Coppersmith Brockelman, PLC

October 2020

The past few years have seen significant developments in federal regulation of human subjects research at the statutory, regulatory, and policy levels. This White Paper covers Institutional Review Board (IRB) review, informed consent, and the use and disclosure of patient information in human subjects research, including regulatory requirements in the Common Rule ([45 C.F.R. Part 46](#)), National Institutes of Health (NIH) policies, the Food and Drug Administration (FDA) regulations for the conduct of clinical trials, the Health Insurance Portability and Accountability Act and its implementing regulations (collectively, HIPAA), and the federal Confidentiality of Substance Use Disorder Patient Records regulations ([42 C.F.R. Part 2](#)).

I. THE COMMON RULE

On January 19, 2017, the Office for Human Research Protections (OHRP) and other federal agencies that have adopted the Common Rule¹ published final regulations amending those rules.² The amended Common Rule was generally effective on January 19, 2019 (with the exception of the rule mandating the use of a single IRB for multi-site studies, which is effective on January 20, 2020).

A. APPLICABILITY OF THE COMMON RULE

The Common Rule applies to human subjects research that is conducted or supported by a federal department or agency that has adopted the Common Rule.³ The Common Rule has been adopted by 16 federal agencies, including HHS.⁴ Notably, the Common Rule has not been adopted by the FDA, which has its own human research protection regulations, which are very similar to the Common Rule (see [Section III](#)).

“Research” is defined in the Common Rule as “a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable

* Kristen Rosati is a partner at Coppersmith Brockelman, PLC, and is considered one of the nation’s leading “Big Data” and HIPAA compliance attorneys. She also has deep experience in data sharing for research and clinical integration initiatives, clinical research compliance, biobanking and genomic privacy, and health information exchange. Kristen is a Past President of the American Health Law Association (AHLA), the nation’s largest educational organization devoted to legal issues in the healthcare field. You can reach Kristen at krosati@cblawyers.com; 602-391-4997.

¹ The “Common Rule” regulations are regulations common across many federal agencies that conduct human subjects research. The Department of Health and Human Services regulations are found at [45 C.F.R. Part 46](#).

² Federal Policy For The Protection of Human Subjects, [82 Fed. Reg. 7149](#) (Jan. 19, 2017).

³ [45 C.F.R. § 46.101](#); [§ 46.103](#). The Office for Human Research Protections (OHRP) has issued “[Human Subjects Regulations Decision Charts](#)” that are helpful in understanding the analysis of what activities are subject to IRB review.

⁴ See [45 C.F.R. Part 46](#) (HHS implementation of the Common Rule).

knowledge.”⁵ Under the amended Common Rule the definition of “research” was revised to expressly exclude a number of activities, including:

- Certain scholarly and journalistic activities;⁶
- Public health surveillance;⁷
- Criminal justice and investigations;⁸ and
- Homeland security.⁹

Whether or not an activity is “research,” the Common Rule does not apply if the research does not involve “human subjects.” An organization does not conduct “human subjects” research under the amended Common Rule as long as both of the following conditions are met: (1) the data or biospecimens were not collected for currently proposed research; and (2) the investigator cannot readily ascertain the identity of the subjects.¹⁰ (The amended Common Rule now explicitly includes identifiable biospecimens in the definition of “human subjects.”¹¹ It does not, however, expand the definition of “human subject” to include de-identified biospecimens.¹²)

The amended Common Rule expressly defines “identifiable biospecimen” and “identifiable private information” as:

(5) Identifiable private information is private information for which the identity of the subject is or may readily be ascertained by the investigator or associated with the information.¹³

(6) An identifiable biospecimen is a biospecimen for which the identity of the subject is or may readily be ascertained by the investigator or associated with the biospecimen.¹⁴

⁵ [45 C.F.R. § 46.102](#); see also [82 Fed. Reg. at 7260](#).

⁶ *Id.* [§ 46.102\(l\)\(1\)](#) (excluding scholarly and journalistic activities (e.g., oral history, journalism, biography, literary criticism, legal research, and historical scholarship), including the collection and use of information, that focus directly on the specific individuals about whom the information is collected).

⁷ [45 C.F.R. § 46.102\(l\)\(2\)](#) (excluding public health surveillance activities, including the collection and testing of information or biospecimens, conducted, supported, requested, ordered, required, or authorized by a public health authority. Such activities are limited to those necessary to allow a public health authority to identify, monitor, assess, or investigate potential public health signals, onsets of disease outbreaks, or conditions of public health importance, including trends, signals, risk factors, patterns in diseases, or increases in injuries from using consumer products. Such activities include those associated with providing timely situational awareness and priority setting during the course of an event or crisis that threatens public health, including natural or man-made disasters).

⁸ *Id.* [§ 46.102\(l\)\(3\)](#) (excluding collection and analysis of information, biospecimens, or records by or for a criminal justice agency for activities authorized by law or court order solely for criminal justice or criminal investigative purposes).

⁹ *Id.* [§ 46.102\(l\)\(4\)](#) (excluding authorized operational activities (as determined by each agency) in support of intelligence, homeland security, defense, or other national security missions).

¹⁰ *Id.* [§ 46.102\(f\); Id. § 46.102\(e\)\(1\); see also OHRP, GUIDANCE ON RESEARCH INVOLVING CODED PRIVATE INFORMATION OR BIOLOGICAL SPECIMENS \(Oct. 16, 2008\).](#)

¹¹ [45 C.F.R. § 46.102\(e\)\(1\)](#).

¹² See [COUNCIL ON GOVERNMENTAL RELATIONS, COMMON RULE OVERVIEW 1 \(Feb. 1, 2017\)](#).

¹³ [45 C.F.R. § 46.102\(e\)\(5\)](#).

¹⁴ *Id.*

Thus, the amended Common Rule generally applies to federally funded research that involves the use of information or biospecimens collected for particular research or from which an investigator can readily ascertain the identity of the subject.

However, the amended Common Rule requires that the federal departments and agencies that implement it re-examine the meaning of identifiable private information and biospecimens, with the help of appropriate experts, within one year of publication of the final rule (*i.e.*, January 19, 2018) and every four years thereafter.¹⁵ The concern is that identities that cannot be readily ascertained today may, with the advent of new technology, be readily ascertainable tomorrow.¹⁶ “The ultimate goal is to implement the Common Rule in a way that is aligned with the evolving understanding of the concept of identifiability while protecting subjects and encouraging and facilitating valuable research.”¹⁷

Through this process, and if appropriate and permitted by law, “such Federal departments and agencies may alter the interpretation of these terms, including through the use of guidance.”¹⁸ Additionally, if this process leads to a determination that new technologies or techniques, when applied to information or biospecimens previously considered non-identifiable, could enable investigators to identify the subjects, then those technologies or techniques will be placed on a list of technologies that can produce identifiable information or biospecimens.¹⁹ The effect of being placed on the list may mean that such technologies and techniques cannot be used, unless the subject has consented or the use is otherwise permissible under the amended Common Rule.²⁰ Importantly, the preamble to the amended Common Rule makes two important clarifications:

- “[A]n investigator who possesses information or biospecimens to which such a technology or technique might be applied is not to be considered in possession of identifiable private information or identifiable biospecimens merely as a result of such a circumstance: that would only be true were the investigator to actually apply the technology or technique to generate identifiable private information or identifiable biospecimens.”²¹
- The public will have the right to notice and the opportunity to comment before any technology or technique is placed on such a list.²²

According to the preamble of the amended Common Rule, the expectation is that whole genome sequencing will be one of the first technologies evaluated for placement on this list.²³

¹⁵ [45 C.F.R. § 46.102\(e\)\(7\)](#).

¹⁶ [82 Fed. Reg. at 7169](#).

¹⁷ *Id.*

¹⁸ [45 C.F.R. § 46.102\(e\)\(7\)\(i\)](#).

¹⁹ *Id.* [§ 46.102\(e\)\(7\)\(ii\); 82 Fed. Reg. at 7169](#).

²⁰ [82 Fed. Reg. at 7169](#).

²¹ *Id.*

²² [45 C.F.R. § 46.102\(e\)\(7\)\(ii\)](#).

²³ [82 Fed. Reg. at 7169](#).

It is unclear how the amended Common Rule's new framework for assessing "identifiability" will interact with HIPAA and HIPAA's de-identification standard. The existing guidance suggests that the current Common Rule's de-identification standard is less rigorous than HIPAA's. For instance, OHRP has stated that under the current Common Rule the code used to mask the patient identity can be derived from identifying information linked to or related to the individual, so long as investigators cannot readily ascertain the identity of the individuals to whom the coded private information or specimen pertains.²⁴ HIPAA does not permit this. However, under the amended Common Rule, data and biospecimens deemed de-identified under HIPAA may eventually be considered identifiable under the amended Common Rule, which may be the case with whole genome sequencing.

B. EXEMPT RESEARCH

The amended Common Rule modified some of the existing exemption categories²⁵ and adds new categories.²⁶ The most significant new exemption categories under the amended Common Rule are those applicable to secondary research regulated by HIPAA or secondary research conducted pursuant to a patient's "broad consent" after limited IRB review. "Secondary" research is research using data or biospecimens collected for purposes other than the specific research being conducted, including collection for clinical purposes or for repositories intended for future research.²⁷

1. Exemption for HIPAA-Regulated Research

The amended Common Rule exempts "[s]econdary research uses of identifiable private information or identifiable biospecimens, if . . . [t]he research involves only information collection and analysis involving the investigator's use of identifiable health information when that use is regulated under [45 C.F.R. parts 160](#) and [164, subparts A](#) and [E](#), for the purposes of 'health care operations' or 'research' as those terms are defined at [45 C.F.R. 164.501](#) or for 'public health activities and purposes' as described under [45 C.F.R. 164.512\(b\)](#)."²⁸ In other words, if the secondary research conducted is regulated by HIPAA, the research is exempt from application of the Common Rule.

²⁴ [OHRP, GUIDANCE ON RESEARCH INVOLVING CODED PRIVATE INFORMATION OR BIOLOGICAL SPECIMENS \(Oct. 16, 2008\)](#).

²⁵ For example, the amended Common Rule exempts survey and interview research if "[t]he information obtained is recorded by the investigator in such a manner that the identity of the human subjects can readily be ascertained, directly or through identifiers linked to the subjects, and an IRB conducts a limited IRB review to make the determination required by [45 C.F.R. § 46.111\(a\)\(7\)](#)." [45 C.F.R. § 46.104\(d\)\(2\)\(iii\)](#); see also [82 Fed. Reg. at 7243-44](#) (expansion of exemption categories).

²⁶ For example, there is a new exemption for benign behavioral interventions involving adults. [45 C.F.R. § 46.104\(d\)\(3\)](#). This exemption is subject to limited IRB review, which is discussed in greater detail below in connection with the broad consent exemption. SACHRP has also issued recommendations to HHS on the benign behavioral intervention exemption, which can be found at: <https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-c-august-2-2017/index.html>.

²⁷ [82 Fed. Reg. at 7191](#).

²⁸ [45 C.F.R. § 46.104\(d\)\(4\)\(iii\)](#).

HIPAA, like the Common Rule, broadly defines research as “a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.”²⁹ However, HIPAA only protects protected health information (PHI), not the actual biospecimens themselves. OHRP added this exemption category to the amended Common Rule on the ground that “HIPAA protections are adequate for this type of research, and that it is unduly burdensome and confusing to require applying the protections of both HIPAA and an additional set of protections.”³⁰

The HIPAA exemption applies where a covered entity follows the HIPAA Research Rules.³¹ As explained in [Section IV](#) below, HIPAA does not require IRB review if one of the following HIPAA research rules is satisfied: (1) patient authorization will be obtained (although the covered entity’s institutional policies may still require IRB review); (2) a Limited Data Set (with Data Use Agreement) is used; (3) only de-identified data is used; (4) the activity is to prepare for research; (5) the activity is to recruit patients; (6) the research involves only information of decedents and the covered entity obtains the required representations; (7) the disclosure is required by law; or (8) the research is “grandfathered” under HIPAA. However, if the covered entity seeks waiver of HIPAA authorization, IRB involvement still will be required, even if the research is exempt from the Common Rule under this exception.³²

2. Exemption for Storage and Use of Data and Biospecimens Collected with Broad Consent after Limited IRB Review

Both the storage of identifiable information and biospecimens for potential secondary research, and the actual use for secondary research, is exempt from the Common Rule if certain

²⁹ *Id.* [§ 164.501](#) (definition of “research”). SACHRP opines that the reason “health care operations” and “public health activities and purposes” are included is because: “even though the text of the definition of ‘research’ is the same in both the HIPAA Privacy Rule and the Common Rule, an activity can be considered ‘research’ under the Common Rule but not subject to research requirements under HIPAA when the primary purpose of the activity is health care operations but a secondary purpose is research. Although this ‘primary purpose’ distinction is not, in contrast to the definition of ‘health care operations,’ an express component of the HIPAA Privacy Rule provision on public health activities and purposes, public health activities appear to have been added to the HIPAA Exemption for the same reason: when the activities permitted under [45 C.F.R. § 164.512\(b\)](#) may be considered ‘research’ under the Common Rule (e.g., collecting adverse event information on an FDA-regulated product and using it to study the efficacy or safety of the product) but are treated as public health activities under HIPAA, then those activities should also be able to receive the benefit of the HIPAA Exemption.” SACHRP, “Recommendations on the Interpretation and Application of § 104(d)(4) the ‘HIPAA Exemption’” (Dec. 12, 2017), available at <https://www.hhs.gov/ohrp/sachrp-committee/recommendations/attachment-b-december-12-2017/index.html>.

³⁰ [82 Fed. Reg. at 7194](#).

³¹ See [SACHRP, Recommendations on the Interpretation and Application of § 104\(d\)\(4\) the “HIPAA Exemption” \(Dec. 12, 2017\)](#).

³² *See id.* (noting that “for the situations in which a research project would need to rely upon a HIPAA waiver or alteration of authorization (as opposed to an express HIPAA authorization) to qualify for the HIPAA Exemption, IRBs or Privacy Boards (depending on the institution) will play an important role in examining and determining that the uses of the identifiable information contemplated by the secondary research project will involve no more than a minimal risk to the privacy of individuals and that the research could not practicably be conducted without the waiver or alteration, criteria that are required to be satisfied under the HIPAA Privacy Rule. Therefore, in the role prescribed for IRBs or Privacy Boards under [45 C.F.R. § 164.512\(i\)](#)—a more limited role than convened IRB review—IRBs and Privacy Boards will serve as crucial gatekeepers of the HIPAA Exemption.”).

requirements are met.³³ The storage of such information and biospecimens is exempt if an IRB conducts a limited IRB review and makes the determination that “broad consent” will be obtained and documented, and that there are provisions to protect privacy and confidentiality in the event of a change in the way the information or biospecimens are stored.³⁴ Secondary research may then be conducted with the stored information or biospecimens if “broad consent” was obtained and documented, an IRB conducts a limited IRB review and makes the determination that the research is within the scope of the “broad consent” granted by the subjects, and that the researchers will not return individual research results to subjects as part of the study plan.³⁵ (It is permissible under this exemption to return individual research results when required by law regardless of whether or not such return is described in the study plan.³⁶) The requirements of “broad consent” and the barriers to implementation are discussed in greater detail below. Likewise, the difference between traditional full IRB review and limited IRB review are also separately discussed below.

3. Other Exemptions for Secondary Research

The amended Common Rule also provides the following exemptions for secondary research involving the use of identifiable information and biospecimens:

- *Exemption for publicly available information and biospecimens:* Secondary research using identifiable information and biospecimens is exempt from the Common Rule if the identifiable information or biospecimens are publicly available.³⁷ For example, it would apply to research uses of a public library’s archives or from a commercial entity that provides information or biospecimens to the public upon request or on a fee or subscription basis.³⁸ This exemption takes the place of the current Common Rule’s exemption for research involving existing, publicly available data.
- *De-identified, non-linked information:* Secondary research using identifiable information or biospecimens is also exempt from the Common Rule if it is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained directly or through identifiers linked to the subjects, the investigator does not contact the subjects, and the investigator will not re-identify subjects.³⁹ This exemption takes the place of the current Common Rule’s exemption for research involving data in which subjects cannot be identified, but broadens it to cover research with information or biospecimens from which identifiers have been removed, even if the original collection of information or biospecimens occurs in the future.⁴⁰

³³ [45 C.F.R. § 46.104\(d\)\(7\)-\(8\)](#).

³⁴ [45 C.F.R. § 46.104\(d\)\(7\)](#) *Id.* [§ 46.111\(a\)\(8\)](#).

³⁵ *Id.* [§ 46.104\(d\)\(8\)](#) (criteria for exemption).

³⁶ [82 Fed. Reg. at 7199](#).

³⁷ [45 C.F.R. § 46.104\(d\)\(4\)\(i\)](#).

³⁸ [82 Fed. Reg. at 7194](#).

³⁹ [45 C.F.R. § 46.104\(d\)\(4\)\(ii\)](#).

⁴⁰ See [82 Fed. Reg. at 7194](#).

- *Federal department or agency research.* There is also an exemption for secondary research conducted by or on behalf of a federal department or agency using government-generated or government-collected nonresearch information, provided that certain conditions are met.⁴¹

C. TRADITIONAL VERSUS LIMITED IRB REVIEW

Like the past Common Rule, full IRB review under the amended Common Rule requires that the IRB make the following determinations:⁴²

- (1) Risks to subjects are minimized:
 - (i) By using procedures that are consistent with sound research design and that do not unnecessarily expose subjects to risk; and
 - (ii) Whenever appropriate, by using procedures already being performed on the subjects for diagnostic or treatment purposes.
- (2) Risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result. In evaluating risks and benefits, the IRB should consider only those risks and benefits that may result from the research (as distinguished from risks and benefits of therapies subjects would receive even if not participating in the research). The IRB should not consider possible long-range effects of applying knowledge gained in the research (for example, the possible effects of the research on public policy) as among those research risks that fall within the purview of its responsibility.
- (3) Selection of subjects is equitable. In making this assessment the IRB should take into account the purposes of the research and the setting in which the research will be conducted. The IRB should be particularly cognizant of the special problems of research that involves a category of subjects who are vulnerable to coercion or undue influence, such as children, prisoners, individuals with impaired decision-making capacity, or economically or educationally disadvantaged persons.⁴³

⁴¹ [45 C.F.R. § 46.104\(d\)\(4\)\(iv\).](#)

⁴² *Id.* [§ 46.111](#); *id.* [§ 46.111\(a\)](#). The amended Common Rule does revise two of the existing criteria: (1) the considerations regarding vulnerable populations has been revised to better reflect the populations that should be given special consideration; and (2) it requires the HHS Secretary to issue guidance to assist IRBs in assessing what provisions are adequate to protect the privacy of subjects and to maintain the confidentiality of information. [82 Fed. Reg. at 7207](#). The elements quoted in this paper reflect the changes made by the amended Common Rule.

⁴³ This element is slightly different under the current Common Rule. The current Common Rule includes “pregnant” women in the list and refers to “mentally disabled persons,” as opposed to those with impaired decision-making capacity. [45 C.F.R. § 46.111\(a\)\(3\)](#); *see also* [82 Fed. Reg. at 7204](#) (“[T]he final rule no longer includes pregnant women or ‘handicapped’ or physically disabled individuals as examples of populations that are potentially

- (4) Informed consent will be sought from each prospective subject or the subject's legally authorized representative, in accordance with, and to the extent required by the Common Rule.
- (5) Informed consent will be appropriately documented or appropriately waived in accordance with the Common Rule's requirements.
- (6) When appropriate, the research plan makes adequate provision for monitoring the data collected to ensure the safety of subjects.
- (7) When appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data.⁴⁴

The amended Common Rule introduced the concept of “limited IRB review” for certain kinds of exempt research. Specifically, “limited IRB review” is required for the following exempt categories:

- Research involving educational tests, surveys, interviews or observation of public behavior if the information obtained is recorded by the investigator in such a manner that the identity of human subjects can be readily ascertained, directly or through identifiers linked to the subjects.⁴⁵
- Research involving benign behavioral interventions involving adults if the information obtained is recorded by the investigator in such a manner that the identity of human subjects can be readily ascertained, directly or through identifiers linked to the subjects.⁴⁶
- The storage and maintenance of identifiable data and biospecimens for secondary research for which broad consent is required.⁴⁷
- Secondary research with identifiable data and biospecimens for which broad consent is required.⁴⁸

vulnerable to coercion or undue influence. Adopting a suggestion from public comment and SACHRP, the final rule uses the term ‘individuals with impaired decision-making ability’ to replace the term ‘mentally disabled persons.’”).

⁴⁴ The amended Common Rule further requires that the Secretary of HHS will: “after consultation with the Office of Management and Budget’s privacy office and other Federal departments and agencies that have adopted this policy, issue guidance to assist IRBs in assessing what provisions are adequate to protect the privacy of subjects and to maintain the confidentiality of data.” [45 C.F.R. § 46.111\(a\)\(7\)\(i\)](#).

⁴⁵ *Id.* [§ 46.104\(d\)\(2\)](#).

⁴⁶ *Id.* [§ 46.104\(d\)\(3\)](#).

⁴⁷ *Id.* [§ 46.104\(d\)\(7\)](#).

⁴⁸ *Id.* [§ 46.104\(d\)\(8\)](#).

Continuing IRB review is not required for research approved under limited IRB review, unless an IRB determines otherwise.⁴⁹

What a “limited IRB review” consists of depends on the exemption category at issue. For research involving educational tests or benign behavioral interventions involving adults, the IRB is required to determine only if there are “adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data.”⁵⁰ The preamble to the amended Common Rule offers the following framework for anticipated subregulatory guidance on how to evaluate whether research satisfies this requirement:

- The extent to which identifiable private information is or has been de-identified and the risk that such de-identified information can be re-identified;
- The use of the information;
- The extent to which the information will be shared or transferred to a third party or otherwise disclosed or released;
- The likely retention period or life of the information;
- The security controls that are in place to protect the confidentiality and integrity of the information; and
- The potential risk of harm to individuals should the information be lost, stolen, compromised, or otherwise used in a way contrary to the contours of the research under the exemption.⁵¹

For secondary research involving the storage and maintenance of identifiable information or biospecimens, an IRB must determine that:

- Broad consent for storage, maintenance, and secondary research use is obtained in accordance with the Common Rule’s broad consent is requirements;
- Broad consent is appropriately documented or waiver of the documentation requirement is appropriate in accordance with the Common Rule’s requirements; and
- If there is a change made for research purposes in the way the identifiable private information or identifiable biospecimens are stored or maintained, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data.⁵²

⁴⁹ *Id.* § 46.109(f)(1).

⁵⁰ [45 C.F.R. § 46.104\(d\)\(2\)\(iii\), \(3\)\(i\)\(C\); id. § 46.111\(a\)\(7\).](#)

⁵¹ [82 Fed. Reg. at 7207.](#)

⁵² [45 C.F.R. § 46.111\(a\)\(8\).](#)

With respect to the last requirement, the preamble to the amended Common Rule offers the following examples:

Examples of changed aspects of storage or maintenance for research purposes that would require the IRB to find, before those changes go into effect, whether there are adequate provisions to protect the privacy of subjects and maintain the confidentiality of data include the following: If information or biospecimens are moved from one electronic or physical storage location to another due to considerations related to research plans; if information or biospecimens will be stored for longer than they otherwise would have been for the original purpose; if information or biospecimens are placed in a research registry or repository created to serve as a resource for investigators; or investigators are given electronic or physical access to the information or biospecimens. The relevant changes do not necessarily involve moving information or biospecimens from one location to another. Rather, the relevant changes include any change for research purposes that introduces or alters risks to the privacy or security of the stored information or biospecimens, including giving access to or transferring information or biospecimens for research purposes to someone who otherwise would not have access.⁵³

For the exemption for secondary research applies if:

- Broad consent for storage, maintenance, and secondary research use was obtained in accordance with the Common Rule's broad consent requirements;
- Broad consent is appropriately documented (or waiver of the documentation requirement as appropriate in accordance with the Common Rule's requirements);
- An IRB conducts a limited review and determines that there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data, and that the research is within the scope of the broad consent provided; and
- The investigator does not include returning individual research results to subjects as part of the study plan.⁵⁴

The amended Common Rule also permits use of an expedited review process for studies approved under limited IRB review.⁵⁵

D. PREPARATORY TO RESEARCH ACTIVITIES

The amended Common Rule now permits an IRB to approve a research proposal in which the investigator will obtain information or biospecimens for “preparatory to research

⁵³ [82 Fed. Reg. at 7198](#).

⁵⁴ [45 C.F.R. § 46.104\(d\)\(8\)\(iii\); id. § 46.111\(a\)\(8\); id. § 46.104\(d\)\(8\)\(iv\)](#).

⁵⁵ [Id. § 46.110\(b\)\(iii\)](#).

activities” (*i.e.*, for purposes of screening, recruiting, or determining the eligibility of prospective subjects) without informed consent if the investigator either obtains the information: (1) through oral or written communication with the prospective subject or his/her legally authorized representative; or (2) by accessing records or stored identifiable information.⁵⁶ The OHRP describes this in the preamble to the amended Common Rule as an exception to the informed consent requirement, not a waiver.⁵⁷ Although “preparatory to research activities” are not “exempt” under the amended Common Rule, the amended Common Rule is now more closely aligned with HIPAA, which also permits such activities without obtaining patient authorization if certain conditions are met. The amended Common Rule is a change from the current rule, which requires an IRB to *waive* informed consent before investigators may use such information for “preparatory-to-research activities.”⁵⁸ However, the amended Common Rule still requires IRB review and approval for preparatory to research activities, which is not required under the HIPAA Research Rules.

E. INFORMED CONSENT

Unless activities are not “research,” do not involve “human subjects,” or are otherwise “exempt” from the Common Rule, a research subject’s informed consent is required under the amended Common Rule.⁵⁹

Informed consent to participate in research is a *process*, not just a free-standing document. An investigator must have a participant sign an informed consent document that meets the regulatory requirements. In addition, an investigator may seek informed consent only in circumstances that provide the participant or the participant’s representative sufficient opportunity to consider whether to participate in the study and that minimize the possibility of coercion or undue influence.⁶⁰

The basic requirements for informed consent include the following:

- A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures which are experimental;
- A description of any reasonably foreseeable risks or discomforts to the subject;
- A description of any benefits to the subject or to others which may reasonably be expected from the research;

⁵⁶ *Id.* [§ 46.116\(g\)](#).

⁵⁷ [82 Fed. Reg. at 7227](#).

⁵⁸ *See id.*

⁵⁹ [45 C.F.R. § 46.116](#).

⁶⁰ *Id.*

- A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject;
- A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained;
- For research involving more than minimal risk, an explanation as to whether any compensation and an explanation as to whether any medical treatments are available if injury occurs and, if so, what they consist of, or where further information may be obtained;
- An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights, and whom to contact in the event of a research-related injury to the subject; and
- A statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.⁶¹

This information must be given in a language understandable to the subject or representative and not include any exculpatory language through which the subject or representative waives (or appears to waive) any legal rights.⁶²

The amended Common Rule requires a “re-ordering” of the information in an informed consent in the following ways:

- The consent must begin with a concise, focused and plain presentation of the key information that is most likely to assist the subject in understanding the reasons why one might or might not want to participate in the research;⁶³
- It must, as a whole, present information with sufficient detail and be organized in a way that facilitates understanding.⁶⁴ A listing of isolated facts is not sufficient;⁶⁵ and

⁶¹ *Id.* [§ 46.116\(a\)](#). Additional elements that might be required, if appropriate, include: “(1) A statement that the particular treatment or procedure may involve risks to the subject (or to the embryo or fetus, if the subject is or may become pregnant) which are currently unforeseeable; (2) Anticipated circumstances under which the subject's participation may be terminated by the investigator without regard to the subject's consent; (3) Any additional costs to the subject that may result from participation in the research; (4) The consequences of a subject's decision to withdraw from the research and procedures for orderly termination of participation by the subject; (5) A statement that significant new findings developed during the course of the research which may relate to the subject's willingness to continue participation will be provided to the subject; and (6) The approximate number of subjects involved in the study.” *Id.* [§ 46.116\(b\)](#).

⁶² [45 C.F.R. § 46.116\(a\)](#).

⁶³ [45 C.F.R. § 46.116\(a\)\(5\)\(i\)](#).

⁶⁴ *Id.* [§ 46.116\(a\)\(5\)\(ii\)](#).

⁶⁵ *Id.* [§ 46.116\(a\)\(5\)\(ii\)](#).

- The subject (or legally authorized representative) must receive the information that a reasonable person would want to have in order to make an informed decision about whether to participate, and an opportunity to discuss that information.⁶⁶

Additionally, the amended Common Rule adds the following requirements:

- If the research involves the collection of identifiable private information or identifiable biospecimens, the consent form must include a statement that either: (i) identifiers might be removed and that, after such removal, the information or biospecimens could be used for future research studies or distributed to another investigator for future research studies without additional informed consent; or (ii) that the subject's information or biospecimens, even if identifiers are removed, will not be used or distributed for future research studies.⁶⁷
- If the research involves biospecimens, a statement that the subject's biospecimens (even if identifiers are removed) may be used for commercial profit and whether the subject will or will not share in this commercial profit.⁶⁸
- If appropriate, a statement regarding whether clinically relevant research results, including individual research results, will be disclosed to subjects, and if so, under what conditions.⁶⁹
- If the research involves biospecimens, a statement regarding whether the research involves whole genome sequencing (*i.e.*, sequencing of a human germline or somatic specimen with the intent to generate the genome or exome sequence of that specimen).⁷⁰

Under the amended Common Rule, a copy of the informed consent form approved by the IRB for a “clinical trial” must be posted on a federal website that will be established.⁷¹ A clinical trial is a subset of research governed by the amended Common Rule; it is “a research study in which one or more human subjects are prospectively assigned to one or more interventions⁷² (which may include placebo or other control) to evaluate the effects of the interventions on biomedical or behavioral health-related outcomes.”⁷³ The federal agency supporting the research may allow or require the redaction of information that should not be made publicly available,

⁶⁶ *Id.* [§ 46.116\(a\)\(4\)](#).

⁶⁷ *Id.* [§ 46.116\(b\)\(9\)](#).

⁶⁸ *Id.* [§ 46.116\(c\)\(7\)](#).

⁶⁹ *Id.* [§ 46.116\(c\)\(8\)](#).

⁷⁰ [45 C.F.R. § 46.116\(c\)\(9\)](#).

⁷¹ *Id.* [§ 46.116\(h\)\(1\)](#).

⁷² An intervention “includes both physical procedures by which information or biospecimens are gathered (e.g., venipuncture) and manipulations of the subject or the subject's environment that are performed for research purposes.” *Id.* [§ 46.102\(e\)\(2\)](#).

⁷³ *Id.* [§ 46.102\(b\)](#).

such as confidential commercial information.⁷⁴ The form must be posted after the clinical trial is closed to recruitment, but no later than 60 days after the last study visit by any subject, as required by the protocol.⁷⁵ The preamble to the amended Common Rule notes that, “in rare instances, it could be the case that the federal department or agency would determine that the very existence of a particular clinical trial should not be publicly disclosed, in which case no posting related to such a trial would be required.”⁷⁶ The preamble suggests that HHS is considering using ClinicalTrials.gov as the website for posting the informed consent forms.⁷⁷

F. BROAD CONSENT

The amended Common Rule introduced the new concept of “broad consent” for the storage and secondary research use of identifiable information or biospecimens. The elements of broad consent incorporate some of the basic and additional elements of full informed consent,⁷⁸ and also require the following additional elements:

- A general description of the types of research that may be conducted with the identifiable information.⁷⁹ The description must include enough information that a reasonable person would expect the broad consent to permit the future research conducted.⁸⁰

⁷⁴ *Id.* [§ 46.116\(h\)\(2\)](#).

⁷⁵ *Id.* [§ 46.116\(h\)\(3\)](#).

⁷⁶ [82 Fed. Reg. at 7229](#).

⁷⁷ *Id.*

⁷⁸ *Id.* [§ 46.116\(d\)\(1\)](#) (“. . . [T]he following shall be provided to each subject or the subject’s legally authorized representative: The information required in paragraphs (b)(2), (b)(3), (b)(5), and (b)(8) and, when appropriate, (c)(7) and (9) of this section.”). Specifically, the broad consent must include the following basic elements and additional elements of full informed consent (if appropriate):

- A description of any reasonably foreseeable risks or discomforts to the subject. [45 C.F.R. § 46.116\(b\)\(2\)](#);
- A description of any benefits to the subject or to others that may reasonably be expected from the research [45 C.F.R. § 46.116\(b\)\(3\)](#);
- A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained [45 C.F.R. § 46.116\(b\)\(5\)](#);
- A statement that participation is voluntary, refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled [§ 46.116(b)(8)];
- [If appropriate a] statement regarding whether clinically relevant research results, including individual research results, will be disclosed to subjects, and if so, under what conditions [§ 46.116(c)(8)]; and
- For research involving biospecimens, whether the research will (if known) or might include whole genome sequencing (i.e., sequencing of a human germline or somatic specimen with the intent to generate the genome or exome sequence of that specimen) [§ 46.116(c)(9)].

⁷⁹ [45 C.F.R. § 46.116\(d\)\(2\)](#).

⁸⁰ *Id.*

- A description of the identifiable information that might be used in research, whether sharing might occur, and the types of institutions or researchers that might conduct the future research.⁸¹
- A description of the period of time that the identifiable information may be stored and maintained and/or used for research purposes.⁸² The period of time for either could be indefinite.⁸³
- Unless the subject or legally authorized representative will be provided details about specific research studies, a statement that they will not be informed of such details, including the purposes of the future research.⁸⁴ The statement must also inform the subject or legally authorized representative that they might have chosen not to consent to some of those specific research studies.⁸⁵
- Unless it is known that clinically relevant research results, including individual research results, will be disclosed to the subject in all circumstances, a statement that such results may not be disclosed to the subject.⁸⁶
- An explanation of whom to contact for answers to questions about the subject's rights, storage, and use of the subject's identifiable information or biospecimens, and whom to contact in the event of a research-related harm.⁸⁷

Under the amended Common Rule, researchers are not required to use broad consent. As the Secretary's Advisory Committee on Human Research protections (SACHRP) explained in recommendations to HHS, the amended Common Rule "allows for broad consent to be obtained as an alternative to traditional informed consent for the non-exempt storage, maintenance, and secondary research use of identifiable private information or identifiable biospecimens (collected for either research studies other than the proposed research or nonresearch purposes)."⁸⁸ A researcher, therefore, can take the traditional route of obtaining full IRB-approved informed consent for a specific research study, or seek a waiver from the IRB of informed consent to store and use identifiable information for research.⁸⁹

⁸¹ *Id.* [§ 46.116\(d\)\(3\)](#).

⁸² *Id.* [§ 46.116\(d\)\(4\)](#).

⁸³ *Id.*

⁸⁴ *Id.* [§ 46.116\(d\)\(5\)](#).

⁸⁵ *Id.*

⁸⁶ *Id.* [§ 46.116\(d\)\(6\)](#).

⁸⁷ *Id.* [§ 46.116\(d\)\(7\)](#).

⁸⁸ [SACHRP, ATTACHMENT C – RECOMMENDATIONS FOR BROAD CONSENT GUIDANCE \(Aug. 2, 2017\)](#).

⁸⁹ [45 C.F.R. § 46.116\(d\); see also SACHRP, ATTACHMENT C – RECOMMENDATIONS FOR BROAD CONSENT GUIDANCE \(Aug. 2, 2017\)](#) ("Under the Final Rule, therefore, researchers may opt to conduct secondary research on identifiable private information or identifiable biospecimens through one of the following mechanisms: (i) study specific consent and full or expedited IRB review, (ii) broad consent and limited IRB review (under the exemptions that rely on broad consent), (iii) waiver of consent and full or expedited IRB review, (iv) other exemptions, or (v) de-identification to remove the research activity from the scope of the Common Rule, which would not require consent or IRB review.").

There are significant implementation barriers to broad consent. As explained by SACHRP:

To implement fully a broad consent program, health care institutions would be required to install a system to track biospecimens and data for which individuals provide their broad consent, as well as the terms of the broad consent to determine which future research uses remain within scope. Notably, if an individual is offered to provide broad consent but refuses, the limitation only proscribes secondary research uses of the identifiable materials – meaning that researchers could simply choose to de-identify the subject’s data and biospecimens to conduct further research with them. A subject’s refusal to give broad consent also does not prevent the unconsented uses of their identifiable data and biospecimens for purposes that are not considered “research” under the revised Common Rule. For these reasons, if a person who is offered a broad consent refuses to give that consent, health care institutions have three basic options. First, if allowed by other law, they may simply destroy that person’s identifiable information and biospecimens. Second, they may de-identify the person’s information and biospecimens and use them for future research without restraint. Third, they may decide to retain the identifiable information and biospecimens, but allow their future use only for non-research purposes, such as quality improvement. In this third option, however, the institution must track that person’s information and biospecimens to ensure they are not used for future research purposes.

Extensive and seamless IT system capacity will be necessary for any institution or health system to implement fully a broad consent tracking system, as both broad consents as well as refusals to consent (unless the materials are destroyed) must be tracked over the lifetimes of persons who give broad consent and persons who refuse to give such consent. Due to these systems requirements for electronic tracking processes, SACHRP expects that, practically speaking, institutions or systems without interconnected, interfacing and fully interoperable medical records systems will not be able to implement and benefit from the broad consent regimen established in the Final Rule. A “confederated,” non-IT-unified health system will simply not be able, without significant error, to track these consents and refusals to consent. These logistical barriers will greatly limit the utility of the broad consent option.⁹⁰

SACHRP suggests that the practical utility of broad consent will be limited to an identified biorepository or databank study, or primary research studies in which the researchers seek to use an “add-on” or integrated broad consent to facilitate future research uses.⁹¹

Like SACHRP, we caution against utilizing the broad consent option. In addition to the implementation issues described above, the most troubling aspect of utilizing broad consent is that if an individual has “declined” to give broad consent, then the amended Common Rule

⁹⁰ [SACHRP, ATTACHMENT C – RECOMMENDATIONS FOR BROAD CONSENT GUIDANCE \(Aug. 2, 2017\).](#)

⁹¹ *Id.*

prohibits an IRB from waiving consent “for the storage, maintenance, or secondary research use of the identifiable private information or identifiable biospecimens . . .”⁹²

However, the regulations do not define what it means to “decline” broad consent. For example, if a broad consent form is presented in a packet of papers that patients sign on admission, and the broad consent form is not returned, did the patient “decline” to sign the broad consent? Or is a declination of consent a more affirmative rejection of the concept, such as a “yes” or “no” choice, where “no” is checked? SACHRP has recommended that HHS interpret “refusal to consent” to include only a person’s express declination to give broad consent, as demonstrated by an individual’s unambiguous written or oral communication to that effect;⁹³ however, HHS has not yet adopted that recommendation.

Moreover, the amended Common Rule does not identify the parties bound by the patient’s refusal to give broad consent. As pointed out by SACHRP in its recent recommendations, “[w]ithout more limitation or explanation . . . such a prohibition could apply to all U.S. IRBs, investigators, and institutions with respect to that individual’s identifiable biospecimens and identifiable private information.”⁹⁴ That interpretation would require downstream tracking of the provenance of information and biospecimens used for research.

There is also the issue of withdraw of broad consent once given. The commentary to the amended Common Rule suggests that if an individual gives broad consent but later withdraws it, investigators may continue to use that person’s collected and stored data and biospecimens so long as they have been subsequently stripped of identifiers so as to not be subject to the Common Rule.⁹⁵ However, as SACHRP has pointed out: “That same guidance also indicates, however, that if an investigator promises – most likely in the informed consent form – that withdrawal of a broad consent would mean that there would be no future research use of that person’s information and biospecimens, then the promise should be honored, and future use of that information and those biospecimens even in de-identified form would not be allowed.”⁹⁶

All of this complexity creates the practical difficulty of tracking whether broad consent has been declined, withdrawn, and deciding what to do with the information or biospecimens on withdrawal (*e.g.*, destroying the information or biospecimens, removing them from the repository, flagging them in the repository to prevent research use, or de-identifying the information or biospecimens), and effectively communicating limitations to personnel involved in research. If researchers choose to utilize the broad consent option, researchers should use a separate broad consent form, and should not include the consent in the clinical admissions form or in an informed consent document for the specific research study. This is consistent with SACHRP recommendations regarding the use of a single form.⁹⁷

⁹² [45 C.F.R. § 46.116\(f\)\(1\)](#).

⁹³ [SACHRP, ATTACHMENT C – RECOMMENDATIONS FOR BROAD CONSENT GUIDANCE \(Aug. 2, 2017\)](#).

⁹⁴ *Id.*

⁹⁵ [82 Fed. Reg. at 7217](#).

⁹⁶ [SACHRP, ATTACHMENT C – RECOMMENDATIONS FOR BROAD CONSENT GUIDANCE \(Aug. 2, 2017\)](#).

⁹⁷ *Id.*

G. WAIVER OR ALTERATION OF THE INFORMED CONSENT REQUIREMENT

Under the amended Common Rule, an IRB may waive or alter the informed consent requirement if it finds that: (1) the research involves no more than minimal risk to the subjects; (2) the waiver or alteration will not adversely affect the rights and welfare of the subjects; (3) the research could not practicably be carried out without the waiver or alteration; (4) whenever appropriate, the subjects will be provided with additional pertinent information after participation; and (5) if the research involves using identifiable information or identifiable biospecimens, the research could not practicably be carried out without using the information in an identifiable format.⁹⁸ This criterion was modeled on the comparable element of waiver under HIPAA.⁹⁹

Additionally, if alteration of consent is sought, the amended Common Rule will not allow certain core elements of informed consent to be changed or omitted.¹⁰⁰ Specifically, the amended Common Rule does not permit omission or alteration of the following requirements:

- That the researcher obtain legally effective informed consent before involving a human subject in research;
- That the research subject (or legally authorized representation) have sufficient opportunity to discuss and consider whether or not to participate in the research and that the circumstances minimize the possibility of coercion or undue influence;
- That the information shall be given in a language understandable to the subject (or legally authorized representative);
- That the subject (or legally authorized representative) be provided with the information that a reasonable person would want to have in order to make an informed decision about whether to participate, and an opportunity to discuss that information;
- That the informed consent begin with a concise and focused presentation of the key information that is most likely to assist a subject (or legally authorized representative) in understanding the reasons why one might or might not want to participate in the research. It must be organized and presented in a way that facilitates comprehension;
- That the informed consent as a whole must present information in sufficient detail relating to the research, and must be organized and presented in a way that does not merely provide lists of isolated facts, but rather facilitates the subject's (or legally

⁹⁸ [45 C.F.R. § 46.116\(f\)\(3\)\(iii\)](#).

⁹⁹ [82 Fed. Reg. at 7224](#).

¹⁰⁰ [45 C.F.R. § 46.116\(f\)\(2\)](#) (“. . . An IRB may not omit or alter any of the requirements described in paragraph (a) of this section.”).

authorized representative's) understanding of the reasons why one might or might not want to participate; and

- That the informed consent not include any exculpatory language through which the subject (or legally authorized representative) is made to waive or appear to waive any legal rights, or release the investigator, the sponsor, the institution, or its agents from liability for negligence.¹⁰¹

The amended Common Rule will also not permit the omission or alteration of any of the elements specifically required for broad consent, if broad consent is used.¹⁰² As discussed in greater detail above, the amended Common Rule also will prohibit waiver of informed consent for individuals who previously were asked to sign a “broad consent” to store or use the information for future research, but declined.¹⁰³

H. CENTRAL IRB REVIEW

The amended Common Rule also mandates single IRB review for cooperative research, subject to some limited exceptions.¹⁰⁴ Cooperative research projects are projects that involve more than one research site responsible for safeguarding the rights and welfare of subjects.¹⁰⁵

II. NATIONAL INSTITUTES OF HEALTH POLICIES

A. NIH CERTIFICATES OF CONFIDENTIALITY

On September 7, 2017, NIH announced updates to its policy for issuing Certificates of Confidentiality to implement Section 2012 of the 21st Century Cures Act (the “Cures Act”).¹⁰⁶ For decades, HHS and later the NIH have issued Certificates to protect individuals participating in biomedical, behavioral, clinical, and other NIH-funded research by enabling investigators to withhold from all persons not connected with the research the names or other identifying characteristics of such individuals, including in response to legal demands, such as a subpoena.¹⁰⁷ This new policy broadens its applicability and increases privacy protections for research participants. It went into effect on October 1, 2017, and is included in the NIH Grants

¹⁰¹ [45 C.F.R. § 46.116\(a\)](#).

¹⁰² *Id.* [§ 46.116\(f\)\(2\)](#).

¹⁰³ *Id.* [§ 46.116\(f\)\(1\)-\(2\)](#).

¹⁰⁴ *Id.* [§ 46.114](#). The limited exceptions to this requirement include the following: “(i) Cooperative research for which more than single IRB review is required by law (including tribal law passed by the official governing body of an American Indian or Alaska Native tribe); or (ii) Research for which any Federal department or agency supporting or conducting the research determines and documents that the use of a single IRB is not appropriate for the particular context.” *Id.* [§ 46.114\(b\)\(2\)](#). Cooperative research that is not subject to the single IRB requirement “may enter into a joint review arrangement, rely on the review of another IRB, or make similar arrangements for avoiding duplication of effort.” *Id.* [§ 46.114\(c\)](#).

¹⁰⁵ *Id.* [§ 46.114\(a\)](#).

¹⁰⁶ [NIH, NOTICE OF CHANGES TO NIH POLICY FOR ISSUING CERTIFICATES OF CONFIDENTIALITY \(Not. No. NOT-OD-17-109\) \(Sept. 7, 2017\)](#).

¹⁰⁷ [42 U.S.C. § 241\(d\)](#) (effective Dec. 22, 2006 to Dec. 12, 2016).

Policy Statement as a standard term and condition of award for new and non-competing awards issued on or after that date.¹⁰⁸

1. Applicability and Scope

The new policy applies to “all biomedical, behavioral, clinical, or other research funded wholly or in part by the NIH, whether supported through grants, cooperative agreements, contracts, other transaction awards, or conducted by the NIH Intramural Research Program, that collects or uses identifiable, sensitive information.”¹⁰⁹ The NIH defines “identifiable, sensitive information” (also called “Covered Information” by the policy) as:

Information about an individual that is gathered or used during the course of biomedical, behavioral, clinical, or other research, where the following may occur:

- An individual is identified; or
- For which there is *at least a very small risk*, that some combination of the information, a request for the information, and other available data sources could be used to deduce the identity of an individual.¹¹⁰

The NIH gives the following examples of Covered Information: “name, address, social security or other identifying number; and fingerprints, voiceprints, photographs, genetic information, tissue samples, or data fields that when used in combination with other information may lead to identification of an individual.”¹¹¹ The NIH does not generally consider summary research results, such as genomic summary results or summary results of clinical trials, to be identifiable, sensitive information.¹¹² Rather, the NIH considers summary results to be about groups of individuals and thinks that they pose less than a very small risk that individuals could be re-identified, even when used in conjunction with other available data sources.¹¹³

The latter part of the Covered Information definition, which includes information where there is a very small risk of identification, is arguably broader than HIPAA’s definition of

¹⁰⁸ [NIH, NOTICE OF CHANGES TO NIH POLICY FOR ISSUING CERTIFICATES OF CONFIDENTIALITY \(Not. No. NOT-OD-17-109\) \(Sept. 7, 2017\)](#); [NIH OFFICE OF EXTRAMURAL RESEARCH, NIH GRANTS POLICY STATEMENT § 4.1.4.1 \(rev. Dec. 2019\)](#).

¹⁰⁹ [NIH, NOTICE OF CHANGES TO NIH POLICY FOR ISSUING CERTIFICATES OF CONFIDENTIALITY \(Not. No. NOT-OD-17-109\) \(Sept. 7, 2017\)](#).

¹¹⁰ *Id.* (emphasis added); *see also* [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS A. GENERAL INFORMATION ABOUT CERTIFICATES](#) (rev. April 28, 2020) (“2. What information is protected by a Certificate? Certificates protect “covered information.” Covered information includes names or any information, documents, or biospecimens containing identifiable, sensitive information related to a research participant. In addition, if there is at least a very small risk that information, documents, or biospecimens can be combined with other available data sources to determine the identity of an individual, then they are also protected by the Certificate.”).

¹¹¹ [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS A. GENERAL INFORMATION ABOUT CERTIFICATES](#) (“6. What is meant by identifiable, sensitive information?”).

¹¹² *Id.* (“10. Are summary results of research prohibited from disclosure by Certificates?”).

¹¹³ *Id.*

“individually identifiable health information,”¹¹⁴ the Common Rule’s “readily ascertainable” standard,¹¹⁵ and may apply even if the research is exempt from the Common Rule. That is, the NIH explains that this policy applies to following types of research in which Covered Information is collected or used:

- Human subjects research as defined in the [Common Rule], *including exempt research* except for human subjects research that is determined to be exempt from all or some of the requirements of [Common Rule] if the information obtained is recorded in such a manner that human subjects cannot be identified or the identity of the human subjects cannot readily be ascertained, directly or through identifiers linked to the subjects;
- Research involving the collection or use of biospecimens that are identifiable to an individual or for which there is at least a very small risk that some combination of the biospecimen, a request for the biospecimen, and other available data sources could be used to deduce the identity of an individual;
- Research that involves the generation of individual level, human genomic data from biospecimens, or the use of such data, regardless of whether the data is recorded in such a manner that human subjects can be identified or the identity of the human subjects can readily be ascertained as defined in the [Common Rule]; or
- Any other research that involves information about an individual for which there is at least a very small risk, as determined by current scientific practices or statistical methods, that some combination of the information, a request for the information, and other available data sources could be used to deduce the identity of an individual, as defined in [\[42 U.S.C. § 241\(d\)\(4\)\]](#).¹¹⁶

¹¹⁴ [45 C.F.R. § 160.103](#) (“*Individually identifiable health information* is information that is a subset of health information, including demographic information collected from an individual, and: (1) Is created or received by a health care provider, health plan, employer, or health care clearinghouse; and (2) Relates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual; and (i) That identifies the individual; or (ii) With respect to which there is a reasonable basis to believe the information can be used to identify the individual.”).

¹¹⁵ *Id.* [§ 46.102\(f\)](#) (“*Human subject* means a living individual about whom an investigator (whether professional or student) conducting research obtains (1) Data through intervention or interaction with the individual, or (2) Identifiable private information. . . . Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects.”).

¹¹⁶ [NIH, NOTICE OF CHANGES TO NIH POLICY FOR ISSUING CERTIFICATES OF CONFIDENTIALITY \(Not. No. NOT-OD-17-109\) \(Sept. 7, 2017\).](#)

The burden is on investigators and institutions to determine whether the new policy applies.¹¹⁷ The NIH offers the following decision-making tool to determine applicability:

To determine if this Policy applies to research conducted or supported by NIH, investigators will need to ask, and answer the following question:

- Is the activity biomedical, behavioral, clinical, or other research?

If the answer to this question is no, then the activity is not issued a Certificate. If the answer is yes, then investigators will need to answer the following questions:

- Does the research involve Human Subjects as defined by [45 C.F.R. Part 46](#)?
- Are you collecting or using biospecimens that are identifiable to an individual as part of the research?
- If collecting or using biospecimens as part of the research, is there a small risk that some combination of the biospecimen, a request for the biospecimen, and other available data sources could be used to deduce the identity of an individual?
- Does the research involve the generation of individual level, human genomic data?

If the answer to any one of these questions is yes, then this Policy will apply.¹¹⁸

Effective October 1, 2017, the NIH automatically “issues” Certificates for research subject to the new policy,¹¹⁹ but does not issue a paper or digital document.¹²⁰ Rather, “[d]ocumentation of NIH funding or support, the NIH CoC Policy (NOT-OD-17-109), the NIH Grants Policy Statement (See 4.1.4.1) subsection 301(d) of the Public Health Service Act, and any additional future guidance issued by NIH, will serve as documentation of the issuance of a Certificate for a specific study.”¹²¹ In short, compliance with the new policy is now a term and condition of NIH funding. Under the prior process, researchers submitted a separate application

¹¹⁷ [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS B. CERTIFICATES FOR NIH-FUNDED RESEARCH](#) (“3. Will NIH indicate which specific NIH-funded studies are issued a Certificate? In general, no. It is the responsibility of recipients and their investigators to determine if their research is collecting or using covered information.”).

¹¹⁸ [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS B. CERTIFICATES FOR NIH-FUNDED RESEARCH](#) (“2. What NIH-funded research is issued a Certificate?”).

¹¹⁹ [NIH, NOTICE OF CHANGES TO NIH POLICY FOR ISSUING CERTIFICATES OF CONFIDENTIALITY \(Not. No. NOT-OD-17-109\) \(Sept. 7, 2017\)](#).

¹²⁰ [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS B. CERTIFICATES FOR NIH-FUNDED RESEARCH](#) (“4. Will NIH provide a paper or digital document to indicate an NIH-funded study is protected by a Certificate?”).

¹²¹ *Id.*

from the grant application and received a physical or digital certificate.¹²² The NIH has abandoned this process for research subject to the new policy because the Act requires that NIH “minimize the burden to researchers, streamline the process, and reduce the time it takes to comply with the requirements associated with applying for a Certificate.”¹²³

For non-federally funded research or research funded by non-HHS federal agencies (that is, research not automatically subject to the new policy), researchers may still request a Certificate by submitting an online request to the NIH.¹²⁴

Importantly, for NIH-funded research that was initiated or ongoing as of December 13, 2016, but before the October 1, 2017 effective date, researchers and institutions will need to determine whether the new policy applies to these research studies.¹²⁵ Because the new policy broadened the meaning of sensitive, identifiable information, there may be NIH-funded research initiated or ongoing as of December 13, 2016 without a Certificate that might now be subject to the new policy. Additionally, the increased privacy protections discussed in greater detail below apply to all research previously issued a Certificate, regardless of the funding source or if the Certificate has expired.¹²⁶

2. Increased Privacy Protections: Disclosure Restrictions and Safeguards

The new policy imposes strict restrictions on disclosures. Recipients of Covered Information must not:

- Disclose or provide, in any Federal, State, or local civil, criminal, administrative, legislative, or other proceeding, the name of such individual or any such information, document, or biospecimen that contains identifiable, sensitive information about the individual and that was created or compiled for purposes of the research, unless such disclosure or use is made with the consent of the individual to whom the information, document, or biospecimen pertains; or
- Disclose or provide to any other person not connected with the research the name of such an individual or any information, document, or biospecimen that contains

¹²² [NIH, CERTIFICATES OF CONFIDENTIALITY: BACKGROUND INFORMATION](#).

¹²³ [NIH, NOTICE OF CHANGES TO NIH POLICY FOR ISSUING CERTIFICATES OF CONFIDENTIALITY \(Not. No. NOT-OD-17-109\)](#) (Sept. 7, 2017); *see also* [42 U.S.C. § 241\(d\)\(1\)\(G\)](#) (“The Secretary shall take steps to minimize the burden to researchers, streamline the process, and reduce the time it takes to comply with the requirements of this subsection.”).

¹²⁴ [NIH, NOTICE OF CHANGES TO NIH POLICY FOR ISSUING CERTIFICATES OF CONFIDENTIALITY \(Not. No. NOT-OD-17-109\)](#) (Sept. 7, 2017); *see also* [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS: D. CERTIFICATES FOR NON-FEDERALLY FUNDED RESEARCH](#); *id.*, [C. CERTIFICATES FOR NON-NIH, FEDERALLY FUNDED RESEARCH](#).

¹²⁵ [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS B. CERTIFICATES FOR NIH-FUNDED RESEARCH](#) (“5. Does the Policy issue certificates to NIH-funded research that began after the enactment of section 2012 of the 21st Century Cures Act on December 13, 2016? What about research that was ongoing at the time of the law’s enactment?”).

¹²⁶ *Id.* (“6. Does the policy issue Certificates to studies covered by a Certificate issued prior to October 1, 2017?”); *id.*, [E. EXISTING CERTIFICATES OF CONFIDENTIALITY](#) (“3. I have an expired Certificate for research that has ended. Does the 21st Century Cures Act impact me?”).

identifiable, sensitive information about such an individual and that was created or compiled for purposes of the research.¹²⁷

Disclosure is permitted only when:

- Required by Federal, State, or local laws (e.g., as required by the Federal Food, Drug, and Cosmetic Act, or state laws requiring the reporting of communicable diseases to State and local health departments), excluding instances of disclosure in any Federal, State, or local civil, criminal, administrative, legislative, or other proceeding;
- Necessary for the medical treatment of the individual to whom the information, document, or biospecimen pertains **and** made with the consent of such individual;
- Made with the consent of the individual to whom the information, document, or biospecimen pertains; or
- Made for the purposes of other scientific research that is in compliance with applicable Federal regulations governing the protection of human subjects in research.¹²⁸

Additionally, recipients must establish and maintain internal controls (e.g., policies and procedures) that provide “reasonable assurance” of compliance with applicable laws and the terms and conditions of the award, including requiring that a non-funded investigator institution that receives Covered Information protected by a Certificate understands that they are also subject to the requirements of subsection 301(d) of the Public Health Service Act ([42 U.S.C. § 241](#)).¹²⁹ The internal control requirements are described in greater detail at [45 C.F.R. § 75.303](#) and Chapter 8.3 of the NIH Grants Policy Statement.

These enhanced privacy protections are permanent and continue in perpetuity, even after a Certificate expires or NIH funding ends.¹³⁰ They also apply to all research previously issued a Certificate, regardless of the funding source or if the Certificate has expired.¹³¹ However, new

¹²⁷ [NIH, NOTICE OF CHANGES TO NIH POLICY FOR ISSUING CERTIFICATES OF CONFIDENTIALITY \(Not. No. NOT-OD-17-109\) \(Sept. 7, 2017\)](#).

¹²⁸ *Id.* (emphasis added). The list of permissible disclosures mirrors the Act. *Compare with* [42 U.S.C. § 241\(d\)\(1\)\(C\)](#) (“(C) The disclosure prohibition in subparagraph (B) shall not apply to disclosure or use that is--(i) required by Federal, State, or local laws, excluding instances described in subparagraph (D); (ii) necessary for the medical treatment of the individual to whom the information, document, or biospecimen pertains and made with the consent of such individual; (iii) made with the consent of the individual to whom the information, document, or biospecimen pertains; or (iv) made for the purposes of other scientific research that is in compliance with applicable Federal regulations governing the protection of human subjects in research.”) (emphasis added).

¹²⁹ *Id.*; [NIH, WHAT IS A CERTIFICATE OF CONFIDENTIALITY?](#)

¹³⁰ [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS B. CERTIFICATES FOR NIH-FUNDED RESEARCH](#) (“7. Does the protection afforded by a Certificate change if the Certificate was issued prior to the date of enactment of section 2012 of the 21st Century Cures Act (Pub.L. 114 – 255) and has expired?”).

¹³¹ *Id.* (“5. Does the Policy issue certificates to NIH-funded research that began after the enactment of section 2012 of the 21st Century Cures Act on December 13, 2016? What about research that was ongoing at the time of the law’s

data collected *after* a Certificate expires or NIH funding ends may not be protected because the Certificate only protects Covered Information collected or used during the period in which the research is funded by NIH.¹³²

3. Informed Consent Requirements

Finally, NIH also expects investigators to inform subjects of the protections and limitations provided by the Certificate, at least for studies in which informed consent is sought.¹³³ The NIH offers the following model language for describing the Certificate protections:

This research is covered by a Certificate of Confidentiality from the National Institutes of Health. This means that the researchers cannot release or use information, documents, or samples that may identify you in any action or suit unless you say it is okay. They also cannot provide them as evidence unless you have agreed. This protection includes federal, state, or local civil, criminal, administrative, legislative, or other proceedings. An example would be a court subpoena.

There are some important things that you need to know. The Certificate DOES NOT stop reporting that federal, state or local laws require. Some examples are laws that require reporting of child or elder abuse, some communicable diseases, and threats to harm yourself or others. The Certificate CANNOT BE USED to stop a sponsoring United States federal or state government agency from checking records or evaluating programs. The Certificate DOES NOT stop disclosures required by the federal Food and Drug Administration (FDA). The Certificate also DOES NOT prevent your information from being used for other research if allowed by federal regulations.

Researchers may release information about you when you say it is okay. For example, you may give them permission to release information to insurers, medical providers or any other persons not connected with the research. The Certificate of Confidentiality does not stop you from willingly releasing information about your involvement in this research. It also does not prevent you from having access to your own information.¹³⁴

enactment?”); *id.* (“6. Does the policy issue Certificates to studies covered by a Certificate issued prior to October 1, 2017?”); *id.*, [E. EXISTING CERTIFICATES OF CONFIDENTIALITY](#) (“3. I have an expired Certificate for research that has ended. Does the 21st Century Cures Act impact me?”).

¹³² [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS: A. GENERAL INFORMATION ON CERTIFICATES](#), (“4. How long does a Certificate’s protection last?”); *id.*, [B. CERTIFICATES FOR NIH-FUNDED RESEARCH](#) (“8. Does a Certificate protect all identifiable, sensitive information collected or used during the duration of a research project?”).

¹³³ [NIH, NOTICE OF CHANGES TO NIH POLICY FOR ISSUING CERTIFICATES OF CONFIDENTIALITY \(Not. No. NOT-OD-17-109\) \(Sept. 7, 2017\)](#); [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS: A. GENERAL INFORMATION ON CERTIFICATES](#) (“8. What is the researcher’s responsibility to inform participants of a Certificate?”).

¹³⁴ [NIH, EXAMPLE INFORMED CONSENT LANGUAGE](#).

The NIH does not expect researchers who recruited subjects prior to the policy's effective date to re-consent or otherwise notify subjects of the expanded protections.¹³⁵ However, the applicable IRB may decide it is appropriate to notify subjects of the change to the Confidentiality protections.

B. NIH SINGLE IRB POLICY

Before the Common Rule was amended to implement a single IRB policy, the NIH issued a policy requiring multi-site research protocols funded by the NIH to use a single IRB for all research sites in the United States.¹³⁶ The NIH policy was effective September 25, 2017.¹³⁷

C. NIH GENOMIC DATA SHARING POLICY

The Genomic Data Sharing Policy of the NIH (the "GDS Policy"), requires informed consent for use of de-identified biospecimens for certain genetic research, even if the biospecimens were initially collected for non-research purposes, such as clinical treatment.¹³⁸ The GDS Policy applies to all "NIH-funded research that generates large-scale human or non-human genomic data as well as the use of these data for subsequent research. Large-scale data include genome-wide association studies, single nucleotide polymorphisms (SNP) arrays, and genome sequence, transcriptomic, metagenomic, epigenomic, and gene expression data."¹³⁹ Moreover, the GDS Policy applies only to:

- Competing grant applications that were submitted to NIH on or after January 25, 2015;
- Proposals for contracts that were submitted to NIH on or after January 25, 2015; and
- NIH intramural research projects generating genomic data on or after January 25, 2015.¹⁴⁰

The Notice of the Implementation of the GDS Policy provides further clarification of the application of the policy to research conducted pursuant to contracts or awards before the effective date:

¹³⁵ [NIH, FAQS: CERTIFICATES OF CONFIDENTIALITY, FAQS B. GENERAL INFORMATION ON CERTIFICATES](#) ("18. If part of my cohort was recruited prior to issuance of the Certificate, but they are no longer actively participating in the study, do I need to re-consent them?").

¹³⁶ See [NIH, FINAL NIH POLICY ON THE USE OF A SINGLE INSTITUTIONAL REVIEW BOARD FOR MULTI-SITE RESEARCH, \(JUNE 21, 2016\) \(NIH PUB. NO. NOT-OD-16-094\)](#). The only exceptions are where review by the proposed single IRB would be prohibited by a federal, tribal, or state law, regulation, or policy, or if there is a compelling justification for the exception.

¹³⁷ *Id.*

¹³⁸ [79 Fed. Reg. at 51345, 51347](#) (Aug. 28, 2014).

¹³⁹ *Id.* at [51350](#).

¹⁴⁰ *Id.* at [51350](#).

Although the GDS Policy does not apply to research submitted prior to the Policy's effective date, NIH, nonetheless, strongly encourages investigators to comply with the expectations outlined in the Policy. Investigators should provide an updated genomic data sharing plan to the funding IC in the submission of the research performance progress report. For studies involving human participants that were initiated before the Policy's effective date and used consents that do not meet the expectations of the GDS Policy, investigators are expected to plan to transition to a consent for future research uses and broad sharing, if possible, particularly for new or additional collections of specimens. There will be reasonable accommodation, determined on a case-by-case basis by the funding IC, for long-term projects ongoing at the time of the Policy's effective date to come into alignment with NIH's expectations for consent and data sharing. The goal is to bring these projects into alignment, to the extent possible, in a reasonable timeframe.¹⁴¹

III. FDA REGULATIONS

The FDA human subject research protection regulations have not changed. However, as noted below, the FDA recently issued new guidance permitting waiver of informed consent consistent with the amended Common Rule.

A. APPLICABILITY

The FDA regulations related to protection of human subjects apply to a “clinical investigation,”¹⁴² which is “any experiment that involves a test article and one or more human subjects,” where the sponsor is required to submit data to the FDA for a product or research approval.¹⁴³ A “human subject,” for purposes of the FDA regulations, is “an individual who is or becomes a participant in research, either as a recipient of the test article or as a control. A subject may be either a healthy human or a patient.”¹⁴⁴ A “test article” is “any drug (including a biological product for human use), medical device for human use, human food additive, color additive, electronic product, or any other article.”¹⁴⁵ Additionally, the FDA regulations on Investigational Device Exemptions (which apply to the development and marketing of IVDs) apply the FDA regulations to the use of human specimens for testing, even if those specimens are de-identified.¹⁴⁶ The use of data only is not a clinical investigation subject to the FDA regulations.

¹⁴¹ NIH, [IMPLEMENTATION OF THE NIH GENOMIC DATA SHARING POLICY FOR NIH GRANT APPLICATIONS AND AWARDS \(Notice No. NOT-OD-14-111\) \(Aug. 27, 2014\)](#).

¹⁴² [21 C.F.R. § 50.1](#).

¹⁴³ *Id.* [§ 50.3\(c\)](#).

¹⁴⁴ *Id.* [§ 50.3\(g\)](#).

¹⁴⁵ *Id.* [§ 50.3\(j\)](#).

¹⁴⁶ See [21 C.F.R. § 812.3\(p\)](#) (defining “subject” as “a human who participates in an investigation, either as an individual on whom or on whose specimen an investigational device is used or as a control. A subject may be in normal health or may have a medical condition or disease.”).

B. INFORMED CONSENT

In general, FDA regulations require informed consent to use human specimens for clinical investigations.¹⁴⁷ The FDA's required elements for informed consent are similar (but not identical) to the current Common Rule's requirements (but have not been amended to reflect the new amended Common Rule informed consent elements). Specifically, the FDA requires the following elements for informed consent:

- (1) A statement that the study involves research, an explanation of the purposes of the research and the expected duration of the subject's participation, a description of the procedures to be followed, and identification of any procedures which are experimental.
- (2) A description of any reasonably foreseeable risks or discomforts to the subject.
- (3) A description of any benefits to the subject or to others which may reasonably be expected from the research.
- (4) A disclosure of appropriate alternative procedures or courses of treatment, if any, that might be advantageous to the subject.
- (5) A statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained and that notes the possibility that the Food and Drug Administration may inspect the records.
- (6) For research involving more than minimal risk, an explanation as to whether any compensation and an explanation as to whether any medical treatments are available if injury occurs and, if so, what they consist of, or where further information may be obtained.
- (7) An explanation of whom to contact for answers to pertinent questions about the research and research subjects' rights, and whom to contact in the event of a research-related injury to the subject.
- (8) A statement that participation is voluntary, that refusal to participate will involve no penalty or loss of benefits to which the subject is otherwise entitled, and that the subject may discontinue participation at any time without penalty or loss of benefits to which the subject is otherwise entitled.¹⁴⁸

Additional elements that might be required when appropriate can be found at [21 C.F.R. § 20.25\(b\)](#). The informed consent must also be given in a language that is understandable to the subject or his/her representative and cannot include any exculpatory language that is made to

¹⁴⁷ *Id.* [§ 50.20](#).

¹⁴⁸ [21 C.F.R. § 50.25](#).

waive or appear to waiver the subject's legal rights or release the sponsor, investigator or institution from liability for negligence.¹⁴⁹

The FDA's informed consent requirements apply even if biospecimens are completely de-identified, because it is still "human subject" research under the FDA regulations. (This is different under the Common Rule, where the use of de-identified biospecimens is either not "human subject" research or is exempt.¹⁵⁰) However, the FDA has issued two guidance documents, in which the FDA has stated its intent to allow waiver of informed consent, which are described below.

C. IVD DEVICE STUDIES

The FDA issued a guidance document in 2009, in which it permits the use of unconsented biospecimens for *in vitro* diagnostic (IVD) device studies under certain circumstances and agrees not to enforce its informed consent regulations.¹⁵¹ Under this guidance, to use de-identified biospecimens for IVD studies:

- The IVD device study must meet the IDE exemption criteria at [21 C.F.R. § 812.2\(c\)\(3\)](#);
- The study uses leftover specimens that had been collected for clinical purposes or had been collected for a repository (or for other unrelated research);
- The specimens are not individually identifiable to the investigator, the sponsor or any individual associated with the study. Specimens may be coded if no one associated with the study holds the link.
- Individuals caring for patients are different from and do not share information with individuals associated with study;
- The supplier of specimens has a policy that prevents release of identifiable information; and
- The study protocol is reviewed by an IRB.¹⁵²

¹⁴⁹ *Id.* [§ 50.20](#).

¹⁵⁰ See [OHRP, CODED PRIVATE INFORMATION OR SPECIMENS USE IN RESEARCH GUIDANCE \(Oct. 16, 2018\)](#), (discussing distinction between research not involving human subjects versus exempt human subjects research under the current Common Rule).

¹⁵¹ See [FDA, GUIDANCE ON INFORMED CONSENT FOR IN VITRO DIAGNOSTIC DEVICE STUDIES USING LEFTOVER HUMAN SPECIMENS THAT ARE NOT INDIVIDUALLY IDENTIFIABLE: GUIDANCE FOR SPONSORS, INSTITUTIONAL REVIEW BOARDS, AND FOOD AND DRUG ADMINISTRATION STAFF \(Apr. 25, 2006\)](#).

¹⁵² *Id.* at 4-5.

D. WAIVER OR ALTERATION OF INFORMED CONSENT

More recently, the FDA has issued a guidance document permitting IRB waiver of informed consent when the proposed clinical testing poses no more than minimal risk to the human subject, and includes appropriate safeguards to protect the rights, safety, and welfare of the human subject.¹⁵³ Under this guidance, an IRB may waive or alter some or all of the elements of informed consent when the IRB finds and documents that:

- The clinical investigation involves no more than minimal risk (as defined in [21 C.F.R. 50.3\(k\)](#) or [56.102\(i\)](#)) to the subjects;
- The waiver or alteration will not adversely affect the rights and welfare of the subjects;
- The clinical investigation could not practicably be carried out without the waiver or alteration; and
- Whenever appropriate, the subjects will be provided with additional pertinent information after participation.¹⁵⁴

These requirements mirror the waiver requirements under the current Common Rule. The FDA guidance does not adopt the amendment to the Common Rule's waiver requirements that requires the IRB to find that, if the research involves using identifiable information or identifiable biospecimens, the research could not practicably be carried out without using the information in an identifiable format.¹⁵⁵ The FDA, however, has indicated that as it revises its regulations, to the extent appropriate and permissible, it will consider including this new criterion in any provision.¹⁵⁶

E. EMERGENCIES

Where it is not possible to obtain informed consent in an emergency, an investigational drug or device may be used if a research investigator and a second physician not involved in the clinical study certify in writing that all of the following requirements are met:

- (1) The participant is confronted by a life-threatening situation necessitating the use of the investigational drug or device;

¹⁵³ [FDA, IRB WAIVER OR ALTERATION OF INFORMED CONSENT FOR CLINICAL INVESTIGATIONS INVOLVING NO MORE THAN MINIMAL RISK TO HUMAN SUBJECTS: GUIDANCE FOR SPONSORS, INVESTIGATORS AND IRBs 1](#) (July 2017); see also [21st Century Cures Act § 3024](#), Pub. L. 114-255, 130 Stat. 1033 (Dec. 13, 2016).

¹⁵⁴ *Id.*

¹⁵⁵ See [82 Fed. Reg. at 7265-67](#); [45 C.F.R. § 46.116\(f\)\(3\)\(ii\)](#).

¹⁵⁶ [FDA, IRB WAIVER OR ALTERATION OF INFORMED CONSENT FOR CLINICAL INVESTIGATIONS INVOLVING NO MORE THAN MINIMAL RISK TO HUMAN SUBJECTS: GUIDANCE FOR SPONSORS, INVESTIGATORS AND IRBs 3](#) n.6 (July 2017).

- (2) They cannot obtain informed consent from the participant because of the participant's inability to communicate or give legally effective consent;
- (3) There is not sufficient time to seek consent from the participant's legally authorized representative; and
- (4) There is no alternative therapy that provides an equal or greater likelihood of saving the life of the individual.¹⁵⁷

This certification may be performed after use of the drug or device only if immediate use of the drug or device is required to preserve the participant's life. This certification must be submitted to the IRB within five working days after use of the drug or device.¹⁵⁸

The regulations also permit an IRB to waive informed consent, with the concurrence of a physician who is a member of or consultant to the IRB and who is not participating in the clinical study, for research in emergency services.¹⁵⁹ The IRB must approve the research activity and the waiver of informed consent and complete detailed documentation, not discussed here.¹⁶⁰ Procedures must be in place to inform the patient or the patient's representative, as soon as is feasible, of the research and of the patient's right to withdraw from the research.

IV. HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA)

A. THE HIPAA RESEARCH RULES

Under the HIPAA Privacy Rule,¹⁶¹ covered entities may use or disclose protected health information (PHI) only as expressly permitted by the regulations. Covered entities may use PHI internally for research or disclose PHI externally to third parties for research if the requirements of at least **one** of the following rules are met:¹⁶²

- The research subject or the subject's personal representative has signed a written HIPAA authorization (or an informed consent document that integrates all HIPAA authorization requirements);¹⁶³
- An IRB has waived the requirement for authorization;¹⁶⁴

¹⁵⁷ [21 C.F.R. § 50.23](#).

¹⁵⁸ *Id.*

¹⁵⁹ *Id.* [§ 50.24](#); *see id.* [56.109](#).

¹⁶⁰ *Id.* [§ 50.24](#).

¹⁶¹ *See* [45 C.F.R. Part 160](#) and [Part 164, Subpart E](#).

¹⁶² [45 C.F.R. § 164.512\(i\)](#) (general rules for use and disclosure of patient information for research). Other HIPAA rules are cited as applicable.

¹⁶³ [45 C.F.R. § 164.508](#).

¹⁶⁴ *Id.* [§ 164.512\(i\)\(i\)](#).

- The research involves only de-identified data;¹⁶⁵
- The research uses or discloses a Limited Data Set (“LDS”) and the covered entity has a Data Use Agreement (“DUA”) in place with the recipient of the LDS;¹⁶⁶
- The activities are just to prepare for research and required representations are obtained from the researchers;¹⁶⁷
- The use or disclosure is for patient recruitment purposes (because it is “treatment” or “health care operations”);¹⁶⁸
- The research involves only the information of decedents and required representations are obtained from the researchers;¹⁶⁹
- The disclosure is required by law;¹⁷⁰ or
- The research is “grandfathered” under the HIPAA rules.¹⁷¹

This White Paper discusses the first seven HIPAA Research Rules in greater detail below. We do not cover the last two because they are rarely used.

1. HIPAA Authorization

A HIPAA authorization may be obtained for the use of disclosure of PHI for research. A valid HIPAA authorization must include a number of items:

- A specific and meaningful description of the PHI to be used or disclosed in the research (such as the patient’s medical records or other more limited portions of the record, such as laboratory results);
- The name or specific identification of the persons or class of persons authorized to make the disclosure (such as the hospital);
- The name or specific identification of the persons or class of persons who will have access to the PHI;

¹⁶⁵ *Id.* [§ 164.514\(a\)-\(b\)](#).

¹⁶⁶ *Id.* [§ 164.514\(e\)\(1\)](#).

¹⁶⁷ *Id.* [§ 164.512\(i\)\(1\)\(iii\)](#).

¹⁶⁸ *Id.* [§ 164.506](#) (treatment or health care operations).

¹⁶⁹ *Id.* [§ 164.512\(i\)](#).

¹⁷⁰ *Id.* [§ 164.512\(a\)](#).

¹⁷¹ *Id.* [§ 164.532\(c\)](#). Research is “grandfathered” under HIPAA if the patient signed an informed consent before April 14, 2003 (and the informed consent has not been modified since that date) or if the IRB waived informed consent before April 14, 2003.

- A description of the research;
- An expiration date or event (such as the end of the study), or a statement that the authorization has no expiration;
- A statement of the patient's right to revoke the authorization in writing and a description of how to do so;
- A statement that the patient may not revoke the authorization as to information already disclosed for the research where the information is necessary to maintain the integrity of the study data, or a description of other exceptions where the patient may not revoke the authorization;
- A statement that the entity disclosing the PHI may not condition treatment, payment, enrollment or eligibility for benefits on the patient signing the authorization. If the individual will not be allowed to participate in a clinical trial without signing the authorization, the authorization must include a statement to that effect;
- A statement that the information disclosed for the research may be subject to redisclosure by the recipient and no longer be protected by the federal privacy rule;¹⁷²
- If the patient will not be given access to medical records during the study, a statement that the patient agrees to the denial of access when consenting to participate in the study, and that the right of access to the records will be reinstated upon completion of the study;
- The patient's signature and the date of signature; and
- If the authorization is executed by a personal representative of the patient (the patient's health care decision maker), a description of that person's authority to act for the patient.¹⁷³

An authorization may also seek permission to use or disclose PHI for future research, as long as the authorization adequately describes the future research purposes "such that it would be reasonable for the individual to expect that his or her protected health information could be used or disclosed for such future research."¹⁷⁴ The OCR expressly provided covered entities with

¹⁷² A reference that the recipient's use of PHI is governed by the informed consent is permissible.

¹⁷³ [45 C.F.R. § 164.508](#).

¹⁷⁴ See [78 Fed. Reg. 5566, at 5612-13](#) (Jan. 25, 2013). ("In order to satisfy the requirement that an authorization include a description of each purpose of the requested use or disclosure, an authorization for uses and disclosures of protected health information for future research purposes must adequately describe such purposes such that it would be reasonable for the individual to expect that his or her protected health information could be used or disclosed for such future research. ... We also agree with commenters that this approach best harmonizes with practice under the Common Rule regarding informed consent for future research, and allows covered entities, researchers and Institutional Review Boards to have flexibility in determining what adequately describes a future research purpose depending on the circumstances. We have consulted with Office for Human Research Protections (OHRP) and the

substantial flexibility in determining appropriate language to accomplish this. (This changed the OCR's pre-2013 interpretation that a HIPAA authorization could not seek permission to use or disclose PHI for future unspecified research, which conflicted with the Common Rule.¹⁷⁵) Recently released OCR guidance on research required by the 21st Century Cures Act of 2016 confirms this and clarifies that:

The Privacy Rule does not require that a research authorization describe each specific future study if the particular studies to be conducted are not yet determined. Instead, to satisfy the requirement that an authorization include a description of each purpose of the requested use or disclosure, an authorization for use and disclosure of PHI for future research purposes must adequately describe future purposes such that it would be reasonable for the individual to expect that his or her PHI could be used or disclosed for such future research. For example, the description could include specific statements with respect to whether sensitive research, such as genetic or mental health research, or [sic] is contemplated. However, the Privacy Rule does not prescribe a fixed level of detail about the future research or identify particular types of PHI as "sensitive." In short, the Privacy Rule gives covered entities and researchers (who may or may not be covered by HIPAA) the flexibility to describe the future research and the health information to be used or disclosed for the future research, so long as such description reasonably puts the individual on notice that his or her protected health information could be used or disclosed for the future research.¹⁷⁶

However, HIPAA's requirements get more complicated if an authorization for the use or disclosure of PHI for research study that involves treatment is combined with an authorization for future research. While a covered entity may require an individual to sign an authorization to disclose PHI for a research study that involves treatment (such as a clinical trial), a covered entity may not condition treatment on signing an authorization for future research.¹⁷⁷ Where the authorization combines an authorization for a clinical trial and an authorization for future research (a "compound" HIPAA authorization), the patient must opt in to the future research.

The OCR has also released additional guidance on the circumstances under which it might be appropriate to provide an individual with reminders of his or her right to revoke an authorization, as well as the appropriate mechanisms by which an individual may revoke an

FDA on this approach to ensure consistency and harmonization with the HHS and FDA human subjects protections regulations, where appropriate.").

¹⁷⁵ See [NIH, RESEARCH REPOSITORIES, DATABASES, AND THE HIPAA PRIVACY RULE](#) (July 2004) (NIH Pub. No. 04-5489); [NIH, INSTITUTIONAL REVIEW BOARDS AND THE HIPAA PRIVACY RULE](#), at 11-12 (Aug. 15, 2003) (NIH Pub. No. 03-5428) (previous OCR interpretation); see also [21 C.F.R. § 50.25](#); [45 C.F.R. § 46.116](#) (Common Rule and FDA regulations informed consent requirements).

¹⁷⁶ [OCR, GUIDANCE ON HIPAA AND INDIVIDUAL AUTHORIZATION OF USES AND DISCLOSURES OF PROTECTED HEALTH INFORMATION FOR RESEARCH](#) (June 2018).

¹⁷⁷ [OCR, RESEARCH](#) ("If research-related treatment is conditioned on the provision of one of the authorizations, such as in the context of a clinical trial, then the compound authorization must clearly differentiate between the conditioned and unconditioned components and provide the individual with an opportunity to opt in to the unconditioned research activity.").

authorization for future research purposes.¹⁷⁸ The OCR concluded that “[t]he Privacy Rule does not require a covered entity to provide periodic reminders about an individual’s right to revoke an authorization;” however, “a covered entity is free to provide reminders to individuals of their right to revoke a research authorization. For example, a covered entity might choose to ask, while obtaining an individual’s authorization, whether the individual would like to receive reminder(s) in the future about the right to revoke authorization. Or, a covered entity might remind a minor participant who reaches the age of majority of their right to revoke a HIPAA authorization originally signed by the minor’s personal representative (usually a parent or guardian).”¹⁷⁹ However, reminders of the right to revoke an authorization are not required.

The mechanism by which an individual may revoke an authorization for future research must be stated on the authorization form or by cross-reference the covered entity’s Notice of Privacy Practices (NPP), if the NPP has a clear description of the revocation process.¹⁸⁰ In either case, the process:

should not be unduly burdensome to the individual such that it would create a barrier to or unreasonably delay the individual’s exercising the right to revoke the authorization. For example, a covered entity cannot require all individuals to use a portal to submit a revocation if one or more individuals may not have easy access to the internet. In addition, if a covered entity provides a standard form for individuals to request revocation, the form should be readily available and accessible to the individual.¹⁸¹

The OCR further explained that a covered entity does not have “knowledge” of a written revocation unless it receives that written revocation from the individual or is informed orally of the revocation by the individual or the researcher.¹⁸² Because a covered entity has knowledge of a revocation when it receives verbal notice, a covered entity should have a process of follow-up with the individual to get the revocation in writing.

¹⁷⁸ [OCR, GUIDANCE ON HIPAA AND INDIVIDUAL AUTHORIZATION OF USES AND DISCLOSURES OF PROTECTED HEALTH INFORMATION FOR RESEARCH \(June 2018\)](#).

¹⁷⁹ *Id.*

¹⁸⁰ *Id.*

¹⁸¹ *Id.*

¹⁸² *Id.* (“To illustrate these points, consider a situation in which a person other than a covered entity making the disclosure obtains an individual’s authorization, which it then presents to such covered entity, thus allowing the covered entity to disclose PHI. If the individual revokes the authorization by writing to that non-disclosing person who obtained the authorization, and neither the individual nor the other person informs the disclosing covered entity of the revocation, that covered entity will not ‘know’ that the authorization has been revoked. For example, a non-HIPAA covered researcher studying cardiac health might obtain an individual’s authorization for “all providers who have seen the individual in the past year” to disclose PHI related to the individual’s heart condition. Later, the individual may decide to revoke the authorization by writing to the researcher requesting such revocation. The Privacy Rule does not require the researcher in this example to inform all covered entities to whom it has presented the authorization that the authorization has been revoked, so one or more disclosing providers may not “know.” At the same time, however, if the individual tells a covered entity that the individual has revoked the authorization in writing to the researcher, the covered entity ‘knows’ of the revocation and must consider the authorization defective (i.e., invalid) under § [164.508\(b\)\(2\)](#).”

2. IRB Waiver of Authorization

If a covered entity wishes to disclose PHI without obtaining patient consent, one option is to ask an IRB to waive authorization. To have the IRB grant this request, a researcher must demonstrate three things:

- The use or disclosure of the patients' identifiable information involves no more than minimal risk to their privacy, based on: (a) an adequate plan to protect information identifying the patients from improper use and disclosure; (b) an adequate plan to destroy information identifying the patients at the earliest opportunity consistent with conduct of the research (unless there is a health or research justification for retention or if retention is required by law); and (c) adequate written assurances that the information identifying the patients will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the study, or for other research permitted by the rules;
- The research could not practicably be conducted without the waiver or alteration of authorization; and
- The research could not practicably be conducted without access to and use of information identifying the patients.¹⁸³

If the researchers can get HIPAA authorization from patients for some purposes but not others, the researchers can ask the IRB for partial waiver or alteration of the authorization. For example, researchers could ask the IRB to waive authorization for the initial review of PHI in medical records to determine which patients may be appropriate participants for a clinical trial, but seek authorization to enroll those patients in the clinical trial.

3. De-Identified PHI

Covered entities may create de-identified data sets that are not subject to HIPAA restrictions on use and disclosure. The HIPAA Privacy Rule protects “individually identifiable health information.” Individually identifiable health information is “health information, including demographic information collected from an individual” that identifies an individual or where “there is a reasonable basis to believe the information can be used to identify the individual.”¹⁸⁴ PHI is a subset of individually identifiable health information that excludes certain health information held by employers and educational institutions.”¹⁸⁵

¹⁸³ 45 C.F.R. § 164.512(i)(2)(ii).

¹⁸⁴ 45 C.F.R. § 160.103 (defining “individually identifiable information” as “information that is a subset of health information, including demographic information collected from an individual, and: (1) Is created or received by a health care provider, health plan, employer, or health care clearinghouse; and (2) Relates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual; and (i) That identifies the individual; or(ii) With respect to which there is a reasonable basis to believe the information can be used to identify the individual.”

¹⁸⁵ *Id.* § 160.103 (defining “protected health information”).

HIPAA provides two ways in which PHI may be “de-identified” so that it is no longer protected by the Privacy Rule.¹⁸⁶

a. Removal of Identifiers

A covered entity may follow the “safe harbor” method of de-identification and remove or code all of the HIPAA “identifiers” in the information. These identifiers include all of the following data about individuals and their family members, household members, or employers:

- Name;
- Street address, city, county, precinct, or zip code (unless only the first three digits of the zip code are used and the area has more than 20,000 residents);
- All elements of dates (except year) directly related to an individual;
- Age over 89 (unless aggregated into a single category of age 90 and older);
- Telephone numbers;
- Fax numbers;
- Email addresses;
- Social security numbers;
- Medical record numbers;
- Health plan beneficiary numbers;
- Account numbers;
- Certificate/license numbers;
- Vehicle identifiers, serial numbers, and license plate numbers;
- Device identifiers and serial numbers;
- Web Universal Resource Locators (URLs) and Internet Protocol (IP) addresses;
- Biometric identifiers, such as fingerprints;
- Full-face photographs and any comparable images; or
- Any other unique identifying number, characteristic, or code.¹⁸⁷

If a covered entity has actual knowledge that, even with these identifiers removed the remaining information could be used alone or in combination with other information to identify the individual, then the information must be treated as PHI.¹⁸⁸

Information may be de-identified under the safe harbor method by coding (rather than removing) the identifiers. Codes may not be derived from any information about the individual, such as the individual’s social security number, medical record number or name (such as initials), and may not be capable of being translated to identify the individual.¹⁸⁹ For example, a

¹⁸⁶ [45 C.F.R. § 164.514\(a\)-\(b\).](#)

¹⁸⁷ *Id.* [§ 164.512\(b\)\(2\)\(i\).](#)

¹⁸⁸ *Id.* [§ 164.514\(b\)\(2\).](#)

¹⁸⁹ See [OCR, GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT \(HIPAA\) PRIVACY RULE \(Nov. 26, 2012\).](#)

valid code may not be derived from the individual's social security number, medical record number or name (such as initials), and may not be capable of being translated to identify the individual.

b. Statistical Expert

The second method of de-identification is to have a qualified statistical expert determine that the risk is very small that the information could be used alone, or in combination with other available information, to identify the patient.¹⁹⁰ The statistical expert must be a person with knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information non-individually identifiable, and must document the methods and results of the analysis that justifies the conclusion of very small risk.¹⁹¹ For this analysis, whether or not there are "identifiers" in the information is not necessarily relevant. For example, a statistical expert could conclude that there is a very small risk of identification if certain dates of service are present in the information.

c. Other Considerations

The process of de-identifying PHI is treated as a covered entity "health care operation," which may be done without the individual's authorization.¹⁹² If the covered entity uses a third party to de-identify the information, the covered entity must first have a BAA in place with that third party.¹⁹³ When a business associate de-identifies PHI on behalf of a covered entity, that process is a "health care operations" function of the covered entity, *whether or not the covered entity participates in the financial benefit of using the de-identified data*. The HIPAA Privacy Rule specifically says that a covered entity may disclose PHI to a business associate for purposes of de-identification "whether or not the de-identified information is to be used by the covered entity."¹⁹⁴ Moreover, the definition of health care operations does not carry any requirement that the covered entity receive financial or other benefit from the activity.¹⁹⁵ After the de-identification process, the business associate may not retain the fully identifiable information for

¹⁹⁰ [45 C.F.R. § 164.514\(b\)](#) ("(1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable: (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and (ii) Documents the methods and results of the analysis that justify such determination").

¹⁹¹ *Id.* [§ 164.514\(b\)](#).

¹⁹² *See id.* [§ 164.501](#), defining health care operations as "any of the following activities of the covered entity to the extent that the activities are related to covered functions: *** (6) Business management and general administrative activities of the entity, including, but not limited to:*** (v) Consistent with the applicable requirements of [§164.514](#), creating de-identified health information or a limited data set, and fundraising for the benefit of the covered entity." *See also id.* [§ 164.506](#) (use or disclosure of PHI for health care operations).

¹⁹³ *Id.* [§ 164.502\(e\)](#); [45 C.F.R. § 164.504\(e\)](#).

¹⁹⁴ *Id.* [§ 164.502\(d\)\(1\)](#). *See also* [NIH, CLINICAL RESEARCH AND THE HIPAA PRIVACY RULE \(NIH Feb. 2004\)](#) (concluding that a covered entity may disclose its PHI to a third party researcher, for the researcher to de-identify that information to support the researcher's research (not the covered entity's research)).

¹⁹⁵ *See id.* [§ 164.501](#) (defining "health care operations").

research without following one of the other HIPAA rules for use or disclosure of PHI for research. OCR issued a guidance on de-identification on November 26, 2012,¹⁹⁶

4. Limited Data Sets and Data Use Agreements

Alternatively, a covered entity could release a Limited Data Set (LDS) to researchers pursuant to a Data Use Agreement (DUA). A LDS is partially de-identified patient information. A LDS excludes all of the direct HIPAA identifiers listed above, except that a Limited Data Set may include: (1) geographic designations above the street level or PO Box; (2) dates directly related to a patient, such as dates of service, birth date, admission date, discharge date, or date of death; or (3) any other unique identifying number, characteristic, or code that is not expressly listed as a direct identifier.¹⁹⁷

The research personnel who access, review, collect, or receive a LDS must sign a DUA in which they agree to protect the confidentiality of the information. This requirement applies to internal personnel, as well to outside researchers.¹⁹⁸ A DUA must do all of the following:

- (A) Establish the permitted uses and disclosures of such information by the limited data set recipient [the purpose of which must be limited to research, public health activities or health care operations]. The data use agreement may not authorize the limited data set recipient to use or further disclose the information in a manner that would violate the requirements of this subpart, if done by the covered entity;
- (B) Establish who is permitted to use or receive the limited data set; and
- (C) Provide that the limited data set recipient will:
 - (1) Not use or further disclose the information other than as permitted by the data use agreement or as otherwise required by law;
 - (2) Use appropriate safeguards to prevent use or disclosure of the information other than as provided for by the data use agreement;
 - (3) Report to the covered entity any use or disclosure of the information not provided for by its data use agreement of which it becomes aware;

¹⁹⁶ [OCR, GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT \(HIPAA\) PRIVACY RULE \(Nov. 26, 2012\)](#).

¹⁹⁷ [45 C.F.R. § 164.514\(e\)\(1\)](#).

¹⁹⁸ The OCR explained: "In the case of a covered entity that wants to create and use a limited data set for its own research purposes, the requirements of the data use agreement could be met by having affected workforce members sign an agreement with the covered entity, comparable to confidentiality agreements that employees handling sensitive information frequently sign." [67 Fed. Reg. 53181, 53236](#) (Aug. 14, 2002).

- (4) Ensure that any agents, including a subcontractor, to whom it provides the limited data set agrees to the same restrictions and conditions that apply to the limited data set recipient with respect to such information; and
- (5) Not identify the information or contact the individuals represented in the information.¹⁹⁹

5. Activities to Prepare for Research

If researchers merely want to review PHI to prepare for research, a covered entity may permit researchers to do if they provide the covered entity with the following representations in writing:

- The PHI is sought solely to prepare for research;
- The PHI is necessary to prepare for research; and
- No information identifying individuals will be removed from the premises in the course of the review.²⁰⁰

Activities to prepare for research include activities such as preparing a research protocol or developing a research hypothesis, identifying prospective research participants, or screening patient records to identify whether there are a sufficient number of patients at a facility to function as a site for a clinical trial.²⁰¹ Contacting patients to solicit participation in a clinical trial is not an activity to prepare for research.²⁰² This is a recruitment activity, which is discussed separately below.

If researchers will need to remove the information from the covered entity's premises to review it, that will not meet the requirements of the representations above. In that circumstance, the researchers should seek to satisfy a different HIPAA option, such as asking the IRB to waive HIPAA authorization. In its guidance document, entitled "Health Services Research and the HIPAA Privacy Rule," the OCR provided more details on when remote access to a server containing PHI is removing the PHI from the premises. The OCR explained:

Remote access connectivity (i.e., out-of-office computer access achieved through secure connections with access permission and authentication) involves a transmission of electronic PHI, which is not necessarily a removal of PHI under the Privacy Rule. However, although the access to PHI through a remote access connection is not itself a removal of PHI, the printing, copying, saving, or

¹⁹⁹ 45 C.F.R. § 164.514(e)(4).

²⁰⁰ *Id.* § 164.512(i)(1)(ii).

²⁰¹ See [NIH, CLINICAL RESEARCH AND THE HIPAA PRIVACY RULE](#), at 11 (Feb. 2004) (NIH Pub. No. 04-5495).

²⁰² *Id.*

electronically faxing of such PHI would be considered to be a removal of PHI from a covered entity.

The Privacy Rule permits a covered entity to rely on representations from persons requesting PHI if such reliance is reasonable under the circumstances. In the case of a request by a researcher to access PHI remotely, this means that, among other things, the risk of removal, as described above, should be assessed in order to determine whether it is reasonable to rely on the researcher's representation that the PHI will not be removed from the covered entity. The covered entity should determine whether its reliance is reasonable based on the circumstances of the particular case.

For example, a covered entity may conclude that it can reasonably rely on representations from researchers who are its employees or contractors because their activity is manageable through the covered entity's employment and related policies establishing sanctions for the misuse of PHI. On the other hand, where the researcher has no connection to the covered entity, the covered entity may conclude that it cannot reasonably rely on the researcher's representations that PHI will not be removed from the covered entity, unless the researcher's activity is managed in some other way.

Covered entities that permit their workforce or other researchers to access PHI via a remote access connection must also comply with ... the Security Rule's requirements for appropriate safeguards to protect the organization's electronic PHI. Specifically, the standards for access control ([45 C.F.R. § 164.312\(a\)](#)), integrity ([45 C.F.R. § 164.312\(c\)\(1\)](#)), and transmission security ([45 C.F.R. § 164.312\(e\)\(1\)](#)) require covered entities to implement policies and procedures to protect the integrity of, and guard against the unauthorized access to, electronic PHI. The standard for transmission security ([§ 164.312\(e\)](#)) also includes addressable specifications for integrity controls and encryption. This means that the covered entity must assess its use of open networks, identify the available and appropriate means to protect electronic PHI as it is transmitted, select a solution, and document the decision.²⁰³

6. Patient Recruitment

HIPAA also permits the use or disclosure of PHI for patient recruitment.²⁰⁴ A health care provider may contact the provider's own patients to determine if the patients are interested in participating in a clinical trial. If the provider or the provider's employees contact the provider's own patients, that use of PHI is either for "treatment" (if the clinical trial involves treatment) or "health care operations" purposes, both of which are permitted without patient authorization

²⁰³ [NIH, HEALTH SERVICES RESEARCH AND THE HIPAA PRIVACY RULE](#), at 14-15 (May 2005) (NIH Pub. No. 05-5308).

²⁰⁴ [NIH, CLINICAL RESEARCH AND THE HIPAA PRIVACY RULE](#), at 11 (Feb. 2004) (NIH Pub. No. 04-5495).

under HIPAA.²⁰⁵ The health care provider also may use a non-employed third party (including the researcher) to contact patients for recruitment purposes, but the provider first would have to obtain a business associate agreement with the third party to do so.²⁰⁶ Of course, a researcher also could request an IRB to waive HIPAA authorization for the initial contact to recruit a patient, even if authorization will be sought from the patient for enrollment in the study. Contacting patients for recruitment is not a “preparatory to research” activity.²⁰⁷

7. Decedent Information

Where the research involves only the information of deceased individuals, researchers may access this information if they provide the covered entity with the following representations in writing:

- The use or disclosure of information is sought solely for the research on the information of decedents;
- The information is necessary for the research; and
- The researcher will provide documentation of the death of the research participants upon request.²⁰⁸

B. MINIMUM NECESSARY STANDARD

Unless a covered entity is using or disclosing PHI for research pursuant to an individual's written authorization, the covered entity must limit its uses and disclosures of PHI for research purposes to the minimum necessary amount of information required for the particular purpose.²⁰⁹ To comply with the minimum necessary standard, research personnel may only request, use and disclose PHI needed for a particular research project. To implement this, covered entities should consider limiting direct access to PHI to those who are involved in the research project.

C. PSYCHOTHERAPY NOTES

The HIPAA Privacy Rule also has strict requirements on the use and disclosure of “psychotherapy notes,” defined as “notes recorded (in any medium) by a health care provider who is a mental health professional documenting or analyzing the contents of conversation during a private counseling session or a group, joint, or family counseling session and that are separated from the rest of the individual's medical record. Psychotherapy notes do not include medication prescription and monitoring, counseling session start and stop times, the modalities

²⁰⁵ 45 C.F.R. §§ 164.501, 164.506.

²⁰⁶ *Id.* §§ 164.502(e), 164.504(e).

²⁰⁷ See [NIH, CLINICAL RESEARCH AND THE HIPAA PRIVACY RULE](#), at 11 (Feb. 2004) (NIH Pub. No. 04-5495) (“Under the ‘preparatory to research’ provision, covered entities may use or disclose PHI to researchers to aid in study recruitment. The covered entity may allow a researcher, either within or outside the covered entity, to identify, but not contact, potential study participants under the ‘preparatory to research’ provision.”).

²⁰⁸ 45 C.F.R. § 164.512(i)(1)(iii).

²⁰⁹ See *id.* § 164.502(b).

and frequencies of treatment furnished, results of clinical tests, and any summary of the following items: Diagnosis, functional status, the treatment plan, symptoms, prognosis, and progress to date.”²¹⁰

Because psychotherapy notes, by definition, are not included in the medical record, this issue does not often present itself in clinical research. However, in order to use psychotherapy notes for research purposes, the individual’s written authorization must be obtained.²¹¹ This authorization cannot be combined with a general HIPAA authorization for the use and disclosure of other health information or with the informed consent document for the specific research study.²¹²

V. SUBSTANCE USE DISORDER INFORMATION

Federal law—[42 U.S.C. § 290dd-2](#) and [42 C.F.R. Part 2](#) (collectively, Part 2)—provides heightened privacy protection for substance use disorder (SUD) information that originates from SUD treatment providers—called Part 2 programs. This Section covers when SUD information protected by Part 2 (“Part 2 Information”) may be used for research purposes.

A. DISCLOSING SUD INFORMATION FOR RESEARCH PURPOSES

Part 2 Information may be released for research purposes under the following three circumstances:

- *Patient consent.* Part 2 Information may be used and disclosed for research purposes pursuant to a patient’s Part 2-compliant consent.²¹³
- *Researchers subject to HIPAA, the Common Rule and/or FDA Regulations.* Part 2 Information may be disclosed to a researcher who is subject to HIPAA, the Common Rule, FDA regulations regarding the protection of human subjects, and/or any combination of all of the above, so long as the disclosing entity’s director, managing director, or other individual authorized to act as the chief executive officer (or his or her designee) determines that such a recipient researcher meets at least one of the following requirements:
 - If subject to HIPAA (*i.e.*, a covered entity or business associate), the researcher has obtained and documented a HIPAA authorization from the patient, or a waiver or alteration of the HIPAA authorization requirement, consistent with the HIPAA Privacy Rule ([45 C.F.R. §§ 164.508](#) or [164.512\(i\)](#), as applicable);
 - If subject to the Common Rule, the researcher has provided documentation either that the researcher is in compliance with the Common Rule requirements,

²¹⁰ *Id.* [§ 164.501](#).

²¹¹ *Id.* [§ 164.508\(a\)\(2\)](#).

²¹² [45 C.F.R. § 164.508\(b\)\(3\)\(ii\)](#).

²¹³ [42 C.F.R. § 2.31](#).

including the requirements related to informed consent or a waiver of consent ([45 C.F.R. §§ 46.111](#) and [46.116](#)) or that the research qualifies for exemption from the Common Rule; and/or

- If subject to the FDA regulations, the researcher has provided documentation that the research is in compliance with the FDA requirements, including the requirements related to informed consent or an exception to, or waiver of, consent.²¹⁴
- *HIPAA Research Rules at [45 C.F.R. § 164.512\(i\)](#).* Part 2 Programs and other lawful holders of Part 2 Information that are subject to HIPAA (*i.e.*, covered entities or business associates) may disclose Part 2 Information for research purposes so long as the disclosure is made in accordance with the HIPAA Privacy Rule requirements at [45 C.F.R. § 164.512\(i\)](#).²¹⁵

B. A RESEARCHER'S PART 2 OBLIGATIONS AND DATA LINKAGES

A researcher who *receives* Part 2 Information for research purposes is subject to Part 2's stringent disclosure restriction with respect to the Part 2 Information.²¹⁶ Even if the researcher receives the Part 2 Information under one of the abovementioned circumstances without the patient's Part 2-compliant consent, the researcher is subject to all of the following requirements:

- The researcher is fully bound by Part 2 with respect to the Part 2 Information and must resist in judicial proceedings any efforts to obtain access to patient records except as permitted by Part 2;
- The researcher must not redisclose Part 2 Information except back to the individual or entity from whom the Part 2 Information was obtained or as permitted under the data linkage provisions at [42 C.F.R. § 2.52\(c\)](#) (see [Section V.B](#));
- The researcher may include Part 2 Information in research reports only in aggregate form in which patient identifying information has been rendered non-identifiable such that the information cannot be re-identified and serve as an unauthorized means to identify a patient, directly or indirectly, as having or having had a SUD;
- The researcher must maintain and destroy patient identifying information in accordance with the security policies and procedures established under [42 C.F.R. § 2.16](#); and
- The researcher must retain records in compliance with applicable federal, state, and local record retention laws.²¹⁷

²¹⁴ *Id.* [§ 2.52\(a\)\(1\)](#).

²¹⁵ [42 C.F.R. § 2.52\(a\)\(2\)](#).

²¹⁶ See *id.* [§ 2.12\(d\)\(2\)\(C\)](#) (consent disclosures for research purposes), [2.52\(b\)](#) (research exception).

²¹⁷ *Id.* [§ 2.52\(b\)](#).

Researchers who have met these requirements may also request linkages (that is, disclose Part 2 Information needed for the linkages) to data sets from both federal and non-federal data repositories.²¹⁸ However, such data linkages are subject to a number of conditions. Specifically, the researcher must:

- Have the request reviewed and approved by an IRB registered with OHRP to ensure patient privacy is considered and the need for identifiable data is justified;
- Upon request, provide evidence of the IRB approval of the research project that contains the data linkage component; and
- Ensure that patient identifying information obtained is not provided to law enforcement agencies or officials.²¹⁹

A data repository, which is fully bound by Part 2 upon receipt of Part 2 Information, also has the following obligations with respect to the Part 2 Information received to do the data linking:

- After providing the researcher with the linked data, the data repository must destroy or delete the Part 2 Information received for the linkages from its records, including sanitizing any associated hard copy or electronic media, to render the patient identifying information non-retrievable in a manner consistent with the policies and procedures established under [42 C.F.R. § 2.16](#); and
- Ensure that the Part 2 Information obtained in accordance with [42 C.F.R. § 2.52\(a\)](#) is not provided to law enforcement agencies or officials.²²⁰

C. PART 2 COURT ORDERS AND RESEARCHERS

Part 2 imposes strict requirements on when a court may order the disclosure of Part 2 Information.²²¹ But even a Part 2 court order cannot be used to authorize researchers, who have received Part 2 Information without the patient's Part 2-compliant consent for research purposes, to disclose that information or use it to conduct any criminal investigation or prosecution of a patient. However, a Part 2 court order issued under [42 C.F.R. § 2.66](#) may authorize the disclosure or use of such information to investigate or prosecute the researcher who is holding the Part 2 Information.²²²

²¹⁸ *Id.* [§ 2.52\(c\)](#); see also [82 Fed. Reg. 6052, 6101](#) (Jan. 18. 2017).

²¹⁹ *Id.* [§ 2.52\(c\)\(1\)](#).

²²⁰ [42 C.F.R. § 2.52\(c\)\(2\)](#).

²²¹ [42 C.F.R. Part 2, Subpart E](#).

²²² *Id.* [§ 2.62](#).

D. PART 2'S RELATIONSHIP TO FEDERAL STATUTES PROTECTING RESEARCH SUBJECTS AGAINST COMPULSORY DISCLOSURE OF THEIR IDENTITY

Certain other federal regulations and administrative actions—such as Section 502(c) of the Controlled Substances Act ([21 U.S.C. 872\(c\)](#)) and the implementing regulations at [21 C.F.R. part 1316](#)), and section 301(d) of the Public Health Service Act ([42 U.S.C. 241\(d\)](#)) and the implementing regulations at [42 C.F.R. part 2a](#))—may confer research privileges on HHS and the Attorney General to authorize researchers conducting certain types of research to withhold from all persons not connected with the research the names and other identifying information concerning research subjects.²²³ Part 2 does not affect these research privileges statutes.²²⁴ That is, a patient's Part 2-compliant consent or application of a Part 2 research exception to the consent requirement cannot be used to circumvent or abrogate these research privileges, if applicable.

²²³ *Id.* [§ 2.21\(a\)](#).

²²⁴ *Id.* [§ 2.21\(b\)](#).



Critics say new NIH policy on scientific data sharing falls short

By [Ed Silverman](#)

November 3, 2020



A building on the National Institutes of Health campus in Bethesda, Maryland. *Lydia Polimeni/National Institutes of Health*

After a five-year effort, the National Institute of Health late last week released its final policy for managing and sharing publicly funded scientific data. But

some critics say the approach falls short, because the language may still make it possible for researchers to withhold their data.

The [new policy](#)⁶, which replaces one issued in 2003, reflects an ongoing push for transparency by academics and scientists who maintain that, without access to scientific data, research cannot be easily replicated. For this reason, they have argued a lack of information inhibits greater scientific understanding that can adversely affect research decisions and, eventually, treatments and health care costs.

The policy, which goes into effect in January 2023, requires researchers to submit a plan to the NIH for preserving and sharing data. The information would be part of a budget justification for funds and technical evaluation for contracts. The agency, meanwhile, has the right to request additional or specific information concerning data sharing plans outlined by researchers.

The move was greeted with a mix of praise and concern, though. While the new policy marks a sign of progress after years of agitating for open science principles, some academics maintained the NIH needs to do more to ensure that scientific data is, in fact, properly shared and in a timely manner.

“We have turned the corner on the explicit expectations,” said Harlan Krumholz, a Yale University professor of medicine who also heads the Yale Open Data Access Project, which is designed to increase access to clinical research data and works with such companies as Johnson & Johnson (JNJ) and Medtronic (MDT). “Now we need to fill in the when and how, and what happens if you don’t.”

Indeed, a key concern expressed by academics is that the policy recommends — but does not require — a timeframe for sharing data. As a result, a researcher may comply with NIH requirements to submit a plan with details for managing and sharing data, but may not feel compelled to actually share any data if there is no required timetable for doing so.

“In the absence of such an explicit mandate, researchers could comply with a data sharing plan but could still withhold their data,” said Rebecca Li, executive director of Vivli, a nonprofit that runs a global data-sharing platform. “Overall, [there are] no surprises here and a bit of a disappointment that bolder policy action was not taken by the NIH in the final version.”

“I do believe that this NIH policy moves data sharing forward incrementally, so in a sense we all ‘win’ as a society if, based on this policy, more data is shared, reused, and knowledge is gained from that data. However, the win could have been so much greater if rather than a limited policy, they had chosen the bolder policy move.”

The story of mRNA: How a once-dismissed idea became a leading technology in the Covid vaccine race¹⁰

Ultimately, the outcome will reflect how the NIH chooses to implement the policy and how academic medical centers choose to support and provide incentives to their researchers, according to Deborah Zarin, a former director of ClinicalTrials.gov, wrote us. Zarin is now program director at the Multi-Regional Clinical Trials Center, run by Brigham and Women’s Hospital and Harvard University.

“It will be many years before we know if the policy results in the deposition of usable data in a generally accessible registry in a timely manner,” she wrote us.

We should note the new policy is separate from a disclosure policy for publicly funded [clinical trial data](#)¹¹ the NIH put into effect three years ago. This policy also followed [sustained](#) controversy over a lack of accessible trial data, because the agency had failed to force clinical trial sponsors to register their studies.

Earlier this year, a federal judge [ruled](#)¹² trial sponsors had to submit a decade’s worth of missing trial data.

Doctors are accepting as much money from industry as they were when OpenPayments launched¹³

For her part, Zarin also argued that there should a requirement to list data sharing plans for NIH-funded studies in ClinicalTrials.gov.

“This is essential to provide public accountability for the data sharing plan, to comply with the International Committee of Medical Journal Editors policy on this issue; to facilitate the process of identifying relevant data by those who may be seeking data from prior studies; and to ensure that those who use the data do so with adequate recognition of the context,” she wrote.

“For clinical trials, timely reporting of summary results is foundational. It is why the trial was conducted. So even though sharing of the participant level data is important, it should not divert attention from ensuring that the summary results are made available in a timely manner. I hope that NIH demonstrates its commitment to these principles by rigorously implementing and enforcing their trial reporting policies.”

About the Author



Ed Silverman³

Pharmalot Columnist, Senior Writer

ed.silverman@statnews.com¹⁴

Links

1. <https://www.statnews.com/category/the-regulars/pharmalot/>
2. <https://www.statnews.com/stat-plus/latest/>
3. <https://www.statnews.com/staff/ed-silverman/>

4. <https://twitter.com/Pharmalot>
5. <https://www.parsintl.com/publication/stat/>
6. <https://public-inspection.federalregister.gov/2020-23674.pdf>
7. <https://www.statnews.com/signup/>
8. <https://www.statnews.com/privacy/>
9. <https://www.statnews.com/most-popular/>
10. <https://www.statnews.com/2020/11/10/the-story-of-mrna-how-a-once-dismissed-idea-became-a-leading-technology-in-the-covid-vaccine-race/>
11. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-16-149.html>
12. <https://www.statnews.com/2020/08/04/nih-warns-missing-clinical-trial-data/>
13. <https://www.statnews.com/pharmalot/2020/11/03/physician-conflicts-open-payments-pharma/>
14. <https://www.statnews.com/pharmalot/2020/11/03/nih-scientific-data-transparency-policy/>
mailto:ed.silverman@statnews.com
15. <https://www.statnews.com/topic/government-agencies/>
16. <https://www.statnews.com/topic/stat-plus/>