DISCUSSION PAPER

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

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

July 1, 2019

Introduction

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

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

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

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

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

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

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

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

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

The NAM Collaboration

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

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

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

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

Examples of IPD Sharing Initiatives

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

IPD Sharing Opportunities and Obstacles

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table 1 | Key Characteristics of IPD Sharing Case Examples

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

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

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

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

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

SOURCE: Developed by authors.

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

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

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

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

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

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

SOURCE: Developed by authors.

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

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

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

varied data-sharing platforms and repositories with different requirements.

Conclusion

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

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

References

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

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

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

coutcomes.114.001166.

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

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

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

DOI

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

Suggested Citation

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

Author Information

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

Acknowledgments

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

Conflict-of-Interest Disclosures

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

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

Disclaimer

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