

VIEWPOINT

Approaches and Costs for Sharing Clinical Research Data

Erin E. Wilhelm, MPH

Department of Pharmacology and Physiology, Program for Regulatory Science and Medicine, Georgetown University, Washington, DC.

Emily Oster, PhD

University of Chicago Booth School of Business, Chicago, Illinois.

Ira Shoulson, MD

Departments of Neurology, Pharmacology, and Human Science, Program for Regulatory Science and Medicine, Georgetown University.

The generation, dissemination, and sharing of research data are key ingredients in contributing to scientific progress and the public good. Data sharing has been encouraged to facilitate open science within the clinical research enterprise, improve the development of drugs and devices, and benefit public health.¹ But sharing data is complex. Investigators generally tend to guard research data to retain ownership and property rights, avoid competition, protect confidentiality and privacy, or avoid misuse by unqualified persons. Sharing research data also comes at a cost to the sharer. This Viewpoint examines some approaches and cost considerations involved in sharing participant-level clinical research data.

Models for data sharing vary based on accessibility of data to researchers who did not originally collect the data, the type and size of data collected, the sophistication of the database, and the design and capabilities of the user interface. As a minimalist approach, a data generator could upload deidentified participant-level data to an Internet-based storage

are entered promptly into a quality-controlled database, monitored to ensure ADNI standards, and, within days of deposit, made accessible via a password-protected website to individuals with research credentials who agree not to attempt to identify research participants.

Researchers who access ADNI data also agree to upload new analyses to the ADNI database and cite ADNI in the author line of related publications. More than 4000 investigators have accessed ADNI data, resulting in 411 publications to date (N. Buckholtz, PhD, R. Green, MD, MPH, A. Toga, PhD, J. Trojanowski, MD, PhD, M. Weiner, MD, written communication, August-October 2013). But such an elegant and comprehensive approach to sharing clinical research data comes with substantial costs. Based on interviews with key investigators who orchestrated ADNI (N. Buckholtz, PhD, R. Green, MD, MPH, A. Toga, PhD, J. Trojanowski, MD, PhD, M. Weiner, MD, written communication, August-October 2013),⁴ 4 major categories of costs were identified: (1) infrastructure and administration, (2) standardization, (3) human resources, and (4) opportunity costs.

Infrastructure costs—the costs required to maintain the physical or Internet-based data repository and other facilities—encompass storage, security, and quality-control procedures. Large databases and shared computing environments require firewalls, common “language” among data sets, compliance monitoring, and

ongoing maintenance. A large administration component involves the initial consent, monitoring, and often re-consent of human research participants to share additional or specific data or biospecimens, as well as legal costs to cover management of material transfer agreements, contracts, and risk mitigation. Accessibility to genomic family history data is another complicating and potentially costly obligation when providing adequate protection for research participants.

Regardless of the extent of sharing, standardization costs are associated with preparation of case report forms and a data dictionary, review of data for conformity, updating of data sets, notification of data users about changes, and ongoing quality control. Standardization costs for data-sharing models include the additional effort required to share, beyond what is required of any high-quality clinical research, because it takes considerably more effort to organize and make data understandable to others. In studies involving multiple research sites and investigators,

4 major categories of costs were identified: (1) infrastructure and administration, (2) standardization, (3) human resources, and (4) opportunity costs.

system and provide relatively open access to other researchers. This model typically provides limited or no support to potential users. More expansive models provide extensive and timely sharing of well-characterized data, often drawn from multiple sources, converted to a standardized format, and accompanied by a user-friendly interface and data dictionary. The National Database for Autism Research,² the Parkinson's Progression Markers Initiative,³ and the Alzheimer's Disease Neuroimaging Initiative (ADNI)⁴ are disease-oriented examples of the expansive model that promotes further research and data sharing.

Created in October 2004 and currently funded through 2017 by a combination of government, industry, and foundation sponsors, ADNI is perhaps the best known and most established of these initiatives. ADNI researchers contribute standardized data to a clinical, brain imaging, and biomarker database. At its creation, the ADNI model mandated prospective, timely, and wide sharing of data. After acquisition, clinical research data

Corresponding

Author: Ira Shoulson, MD, Department of Pharmacology and Physiology, Program for Regulatory Science and Medicine, Georgetown University, 2115 Wisconsin Ave NW, Ste 603, Washington, DC 20007 (is267@georgetown.edu).

standardization and quality-controlled data of uniform definition enable analyses of individual clinical trials as well as data from multiple studies.

The cost for human resources is the largest expense for data sharing, especially when sharing involves sophisticated technology that requires qualified and well-trained personnel. For prospective data sharing, site investigators and staff who enroll and evaluate research participants must be trained to record data according to agreed-on standards. Computer scientists, web engineers, system interface experts, programmers, and data management personnel build and maintain the database and ensure its ongoing functionality and accessibility. These experts also monitor for individuals who overtax the system with repeated queries and large data downloads.

Opportunity costs constitute the time and effort lost to data-sharing activities that could have been spent conducting new research, generating new data and analyses, and other pursuits. The investigators who lead ADNI also 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 (N. Buckholtz, PhD, R. Green, MD, MPH, A. Toga, PhD, J. Trojanowski, MD, PhD, M. Weiner, MD, written communication, August-October 2013). At present, there are no prospectively derived metrics by which to reliably estimate costs for each category. However, opportunity costs are theoretically offset by the published analyses conducted by users of the database and the accrued knowledge that ensues.

Each data-sharing project has unique goals, partners, and characteristics; however, all data sharing carries some costs, and the more comprehensive and extensive the data sharing, the greater these costs. The minimalist approach, whereby nonstandardized data from a variety of sources are made available through a file-storage website, shares data at relatively low cost but has limited utility for re-

searchers who access the data. The value is greater when the data are of high quality, uniform, timely, and readily accessible.

The cost categories described are borne by those researchers who originally collected the data, with little if any cost to data users. Therefore, cost recovery in data sharing is needed and justifiable, especially because the current funding milieu provides limited support for data sharing.⁵ A simple mechanism for partial cost recovery is a subscription fee for data; this fee should be set to capture any costs experienced on the margin by adding users and also to cover some of the fixed costs of data generation. Profit sharing in any innovations that result from data usage would be a second cost-recovery mechanism, one that is more uncertain but potentially more lucrative.

A widespread belief is that data-sharing costs are not sufficiently supported in budgets by research sponsors. Although there is high demand from sponsors, journals, and the public for data generators to share, few research grants support data-sharing activities, such as those described herein.⁶ Meanwhile, there are few benchmarks by which to ascertain the costs of data sharing and as yet no prospectively derived metrics by which to reliably estimate the categorical costs. A recognition and systematic assessment of data-sharing costs would help sponsors, government and private alike, in planning for infrastructure and administration, standardization, human resources, and opportunity costs of data sharing. It may be useful to create a publicly accessible online database that catalogs data-sharing initiatives, and their associated costs, to inform data generators and sponsors who may then plan appropriate budgets and funding targets, respectively.

The increasing commitments by companies and investigators to more widely share individual participant-level clinical research data expand the opportunity to examine costs associated with a variety of sharing models. In the long run, data sharing is only valuable to the extent that data are organized, accessible, and readily usable by others. Understanding and planning for the costs at the outset of research can help realize the full potential of data sharing.

ARTICLE INFORMATION

Published Online: February 20, 2014.
doi:10.1001/jama.2014.850.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Shoulson reported serving as a consultant regarding experimental therapeutics for Alkermes, AstraZeneca, Auspex Pharmaceuticals, AZTherapies, Caravel Group, Chelsea Therapeutics, Edison Pharmaceuticals, Impax, Ipsen, Isis, Johnson & Johnson, Knopp Biosciences, Link Medicine Lundbeck, Genzyme, Innovation Consulting Group, Medtronic, Neuroglobe Ltd, Omeros Corporation, Orphazyme, Pick Research Solutions, Prana Biotechnology, Salamandra LLC, Seneb Biosciences, Shire HRT, Sofinnova Venture Partners, Velocity Pharmaceuticals Voyager, and Third Rock Ventures; serving as a consultant for the University of California at Irvine, University of Rochester, Columbia University, the Parkinson's Disease Foundation, and the Micheal J. Fox Foundation;

serving on the oversight committee for National Institute of Neurological Disorders and Stroke (NeuroNext); and serving on a Randomized Clinical Trial Steering Committee for Partners (Massachusetts General Hospital) Care. Dr Shoulson is also associate editor, *JAMA Neurology*. No other authors reported disclosures.

Funding/Support: This research was supported by the Georgetown University Center of Excellence in Regulatory Science and Innovation, a cooperative agreement (FD U01004319) funded by the US Food and Drug Administration (FDA).

Role of the Sponsors: The FDA had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; the preparation or approval of the manuscript; or the decision to submit the manuscript for publication. The FDA reviewed the manuscript without comment prior to submission.

Disclaimer: The opinions expressed in this Viewpoint are those of the authors and do not represent the position or opinion of the FDA.

REFERENCES

1. Ross JS, Krumholz HM. Ushering in a new era of open science through data sharing: the wall must come down. *JAMA*. 2013;309(13):1355-1356.
2. National Database for Autism Research (NDAR). NDAR Data Distribution. NDAR website. <http://ndar.nih.gov>. Accessed January 28, 2014.
3. Parkinson's Progression Markers Initiative (PPMI) website. <http://www.ppmi-info.org>. Accessed January 28, 2014.
4. Alzheimer's Disease Neuroimaging Initiative (ADNI) website. <http://www.adni-info.org>. Accessed January 28, 2014.
5. Spertus JA. The double-edged sword of open access to research data. *Circ Cardiovasc Qual Outcomes*. 2012;5(2):143-144.
6. Berman F, Cerf V. Science priorities: who will pay for public access to research data? *Science*. 2013;341(6146):616-617.