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

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

DOI: 10.1056/NEJMp1608351

Data Sharing at a Crossroads

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

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

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

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

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

From the start, we believed that proposals for research conducted with patient-level data should be reviewed for scientific merit as a condition of access and that such review could be conducted objectively only by a panel independent of study sponsors. We appointed the original independent review panel (IRP). Sponsors checked research proposals to ensure that they were complete and met conditions for data access, then sent them to the IRP. Sponsors communicated the outcomes to researchers. To strengthen independence, in 2015 the Wellcome Trust began managing proposal review, interacting with sponsors, the IRP, and researchers. The Trust has appointed a new IRP with no sponsor involvement, which has been operating since December 2015 (its members are listed, along with their charter, at https://clinicalstudydatarequest.com).

The original panel's default approach was to approve proposals and permit access unless there was a compelling reason not to do so.[2] More than 200 research proposals have been submitted to date. Of those that have met requirements and not been withdrawn, less than 10% have been rejected or have resulted in the researchers being advised to resubmit the proposal. There are currently nearly 100 ongoing research projects that are using requested data. To date, however, only four analyses have been published using these data. In the first year, the majority of proposed research aimed to ask

new questions using data from multiple studies rather than reanalyzing single studies.[2] This finding is consistent with a review of proposals including those submitted to two other data-sharing systems, the Yale University Open Data Access (YODA) project and the Supporting Open Access Research (SOAR) initiative.[3]

Should more research have been conducted, and more articles been published, with the available data? Greater awareness of data availability and expertise in using data from clinical trials in pooled and meta-analyses are likely to increase use. Since the research is conducted and published independently of the study sponsors, we don't know why there are so few publications. The yield may reflect the time required for preparing and submitting publications, but also perhaps difficulties in conducting analyses of data from trials that used different structures and standards for data and metadata. Monitoring of research outputs can demonstrate the benefits of data sharing but also inform adjustment of systems, processes, and mechanisms and encourage consistency of data standards.

Several factors were important in developing the system. GSK encouraged other sponsors to join and benefit from the infrastructure. An industry commitment to data sharing was important,[4] as was sponsors' ability to list studies for data sharing according to individual company policies. Sponsors have collaborated, respecting different perspectives in finding solutions.

Not surprisingly, such collaboration can be challenging. For example, some sponsors wanted the option to deny requests in cases of potential conflict of interest or competitive risk. Others didn't require this safety net, which could be seen as compromising independence. The Wellcome Trust agreed to accommodate this approach to enable more sponsors to share data through a single system, which benefits researchers by increasing the amount of data available from as many sponsors as possible. The safety net has not been used to date, and as sponsors gain experience, they may also gain confidence that it isn't needed.

Our vision of a simple single system may be challenging to achieve. But there's a risk that if myriad systems emerge, the benefits will be limited by the complexity of obtaining data.

Our vision of a simple single system may be challenging to achieve. Some industry sponsors such as Johnson & Johnson and Pfizer have chosen to set up their own systems, the Duke Clinical Research Institute has announced that its patient data related to cardiovascular disease will be made available through SOAR (http://soar.dcri.org), and there are other disease-based systems such as the Virtual International Stroke Trials Archive (http://vistacollaboration.org) and Project Data Sphere (https://projectdatasphere.org).

Though these commitments and the proposal of the International Committee of Medical Journal Editors to make data sharing a condition of publication[5] are welcome, there's a risk that if myriad systems emerge, the benefits will be limited by the complexity of obtaining data. The Multi-Regional Clinical Trials (MRCT) Center of Brigham and Women's Hospital and Harvard University, with whom we consulted, is leading an initiative to create a single portal through which data from multiple sponsors and systems can be shared. It is critical that this portal be focused and fit for its purpose and that it work in concert with existing systems.

One possible approach is for everyone to move to a single system whereby sponsors or investigators send study details, data, or both to an independent custodian who manages scientific review, privacy, and other aspects. This approach would require sponsors to give up their own effective systems but would realize economies of scale, helping to address cost barriers. Alternatively, the provision of data-sharing services for some sponsors could be combined with a federated model offering a central portal linking to other systems.

This portal would have to be more than a directory of systems. For receiving and reviewing data requests, it makes sense to have a common proposal form and data-sharing agreement, along with a mutual-recognition approach, whereby approval by an IRP linked to one system is considered sufficient by sponsors who

routinely rely on a different IRP. For accessing data, it may be difficult or impossible for a single project to use data from diverse secure systems. One possible solution is for the researcher (rather than the sponsor) to provide the secure environment into which anonymized data from different sponsors can be securely downloaded. That approach would reduce sponsors' costs and enable researchers to use software other than that provided in secure access systems. Since it could increase privacy risks and misuse of data, however, it might have to be combined with a researcher-accreditation system. As part of maintaining accreditation for data use, the researcher's secure system and the research conducted could be subject to independent audit.

We believe the clinical research enterprise needs to come together to build on what exists and create a simple one-stop shop for clinical trial data sharing. If we get this system right, it could provide a basis for sharing other types of data, such as preclinical data and real-world epidemiologic data. If we allow inevitable differences in systems or processes to produce a fragmented, uncoordinated approach, we will miss the opportunity to realize great value for patients.

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

From the Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC (F.R.); the Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA (P.N.); and Global Medical, GSK, Middlesex, United Kingdom (A.F.).

1. Nisen P, Rockhold F. Access to patient-level data from GlaxoSmithKline clinical trials. N Engl J Med 2013;369:475-8.
2. Strom BL, Buyse M, Hughes J, Knoppers BM. Data sharing, year 1 — access to data from industry-sponsored clinical trials. N Engl J Med 2014;371:2052-4.
3. Navar AM, Pencina MJ, Rymer JA, Louzao DM, Peterson ED. Use of open access platforms for clinical trial data. JAMA 2016;315:1283-4.
4. Bierer BE, Li R, Barnes M, Sim I. A global, neutral platform for sharing trial data. N Engl J Med 2016;374:2411-3.
5. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data — a proposal from the International Committee of Medical Journal Editors. N Engl J Med 2016;374:384-6.

DOI: 10.1056/NEJMp1608086