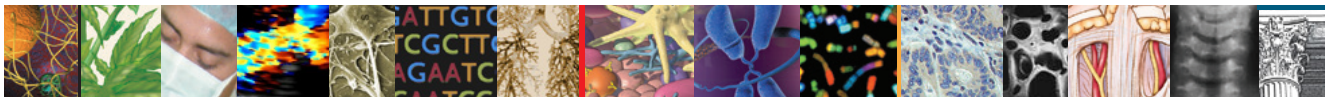

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

A Global, Neutral Platform for Sharing Trial Data

Barbara E. Bierer, M.D., Rebecca Li, Ph.D., Mark Barnes, J.D., LL.M., and Ida Sim, M.D., Ph.D.

Sharing clinical trial data is critical in order to inform clinical and regulatory decision making and honor trial participants who put themselves at risk to advance science. A recent Institute of Medicine (IOM) report argues that availability of deidentified (anonymized) patient-level data from clinical trials can permit verification of original results, enhancing public trust and accountability; facilitate other critical research (e.g., evaluation of adverse event rates according to compound class or subpopulation or identification of surrogate end points); and avert duplicate trials, shielding participants from unnecessary risk.[1] If such goals are to be achieved, patient-level data must be readily findable and available for aggregation and analysis across multiple sources to enable the widest range of secondary research uses.

Recently, several data generators, including pharmaceutical companies and academic consortia, have shared patient-level data and pioneered transparency efforts by establishing electronic portals through which clinical trial data may be requested, shared, and analyzed.[1] However, data largely remain siloed according to sponsor, funder, or disease condition and are often hosted on separate platforms and servers. Most data generators — ranging from multinational pharmaceutical and device companies to small biotechnology firms to academic investigators — have no formal mechanism or easily accessed electronic platform for sharing their data. Established clinical trial registries, such as ClinicalTrials.gov and the World Health Organization International Clinical Trials Registry Platform, include only protocol summaries and summary results.

Furthermore, no organized mechanism exists for negotiating and enforcing the data-use agreements (DUAs) that protect trial participants from reidentification and ensure that data requesters make the results of their own analyses publicly available in turn. The operating rules of a unified platform would also need to accommodate legitimate proprietary interests of data generators, by, for example, allowing some defined interval before sharing begins, during which generators could perform planned secondary analyses to support regulatory applications or to publish additional findings. Sophisticated governance mechanisms and well-defined operating principles and standards would be required.

Despite emerging requirements that clinical trial data be made available, as yet there has been no organized effort to coordinate existing platforms and servers and provide a basic platform to enable most data generators to share their trial data simply, efficiently,

and appropriately and to enable discovery of patient-level data wherever they reside. Now such an entity is being created, sponsored by the Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard University (MRCT Center). We are working with institutional and individual partners to design a data-sharing platform that we call "Vivli" — meant to recall the Greek "vivlithíki," or library, and the Latin "viv," or life. Vivli will link existing data-sharing platforms and communities, while hosting data from investigators who aspire to share data but lack the resources to do so. To establish a platform expeditiously, we plan to launch Vivli in partnership with existing data-sharing systems.

A recent proposal by the International Committee of Medical Journal Editors will require data generators to state their data-sharing plans at trial registration and share the patient-level data underlying published results.[2] To allow implementation of this requirement, Vivli will offer a global, neutral platform to host data securely, with search and data-request services. By coordinating, integrating, and partnering with experience-tested data platforms, Vivli will allow for combining of data sets from multiple clinical trials now held in different electronic locations, thus providing a path to broad-based, international data sharing.

To identify patient-level data relevant to a given research question, requesters will have to search studies on the basis of such criteria as clinical and laboratory eligibility standards, prior treatments, and temporal restrictions. Vivli aims to provide more granular and accurate searching than existing platforms by curating detailed data from current repositories and protocol documents into accurate, structured, computable metadata. To give data requesters a comprehensive view of available patient-level data, metadata will be curated for all studies in which patient-level data are or will be available on Vivli or elsewhere. Curation will initially be manual but we anticipate that it will be progressively supplemented by natural language processing and other machine-based methods.

Hosted patient-level data must be sufficiently anonymized to protect participants' privacy. Risk-based international anonymization standards for deidentification will be used, and before receiving data, requesters will be bound by legally enforceable agreements prohibiting reidentification of trial participants and disclosure of data to others. Vivli will assist data generators by providing, or connecting to, anonymization services and will monitor, by statistical sampling, the sufficiency of data anonymization.

We think that most data requesters will wish to aggregate multiple data sets for analysis. Standardizing the formats of patient-level data facilitates aggregation and anonymization. Data that are anonymized but not mapped to a data standard require substantial curation effort, and curating data that require both anonymization and standardization is even more labor-intensive. In deciding how to deploy limited data-curation resources, Vivli will give priority to data sets that have been specifically requested by researchers. Over time, curation costs will decline if data generators adopt and use common standards from the outset.

Vivli will aim for flexibility in the types of data captured and will eventually develop the capacity to host non–clinical trial data, including public health, epidemiologic, and surveillance data, and genomic and other "omic" data. For data that cannot now be sufficiently anonymized (e.g., genomic data, some imaging data, and data on patients with rare diseases), special provisions for controlled access and monitoring will be needed.

Data sharing will have to be tracked, incentives provided, and data sharing rewarded. All patient-level data sets will have unique identifiers (e.g., digital object identifiers) linking to related data sets, to their generators (e.g., through Open Researcher and Contributor ID [ORCID] identifiers), and to publications citing them, so that data generators may be credited.

At least initially, not all data generators will choose to upload their data onto a common platform, even if they allow them to be discoverable through Vivli's search engine. Thus, Vivli must provide a secure computation environment to enable aggregation of both locally and remotely hosted data sets and to support the use of resident and imported analytic tools.

A central objective of a global data-sharing platform is wide cooperation and collaboration, through inclusion of trials funded and conducted by academia, government, industry, and nongovernmental organizations. Such collaboration requires shared governance with wide representation from diverse communities and countries, as well as from trial

participants. Data generators now using other data-sharing platforms will participate in governance in anticipation of Vivli's bridging to, or potentially assimilating, such platforms.

Data anonymization and platform-security controls are required to protect study participants' privacy. Data requesters will execute a DUA as a condition of access, including commitments to publish results of secondary analyses, to not reidentify participants, and to not share data beyond those specified in the DUA. Some data generators may impose additional constraints on requesters. Vivli will monitor compliance and enforce DUA duties. To assist trial participants, informed-consent templates will be developed explaining in lay terms the differences among identifiable, coded or pseudonymized, and deidentified or anonymized data.

Legitimate interests of data generators, funders, sponsors, and data requesters merit a variety of data-request mechanisms. Data sets available for download will be accessible under a minimal-terms DUA. Data sets accessible only by request will require adjudication by independent review panels; several of these already exist (e.g., Wellcome Trust and Yale University Open Data Access Project) and can be used. In special cases, data generators will review their own received requests; although the IOM report endorses independent review, we believe that allowing self-review at the outset will ease the transition to a data-sharing culture.

Broad sharing of patient-level data should proceed only when data generators and users accept fundamental principles protecting all relevant interests and safeguarding study participants' privacy. Vivli will be one such mechanism, but we are sure there will be others. In any case, we think that the characteristics noted above will provide value to the community. Data platforms that use standardized approaches can achieve efficiencies through shared services (e.g., metadata curation) and, by building trust and sharing governance, can move the clinical research enterprise toward a data-sharing culture. If we can develop global standards and associated tools; harmonize processes and policies for data requests, data-set citation, and data-sharing credit; ensure adherence to terms and conditions; and promote a global conversation about benefits, risks, and methods of data sharing, all will benefit. With respect to the specifics of Vivli, we invite comments on this proposal at www.vivli.org.

Patient-level data are the lifeblood of clinical research, and their value and benefit can be maximized if they are responsibly, routinely and easily shared. We owe nothing less to participants who have endured risk for the benefit of all.

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

From the Multi-Regional Clincal Trials Center of Brigham and Women's Hospital and Harvard University (B.E.B., R.L., M.B.), Boston and Cambridge; the Department of Medicine, Brigham and Women's Hospital (B.E.B., R.L.), the Department of Medicine (B.E.B., R.L.), and the Center for Bioethics (B.E.B., R.L., M.B.), Harvard Medical School, and Ropes and Gray (M.B.) — all in Boston; Yale Law School, New Haven (M.B.); and the University of California, San Francisco, San Francisco (I.S.).

This article was published on May 11, 2016, at NEJM.org.

1. Institute of Medicine. Sharing clinical trial data: maximizing benefits, minimizing risk. Washington, DC: National Academies Press, 2015.
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.

DOI: 10.1056/NEJMp1605348