

Credit for clinical trial data

Industry and its regulators are increasingly finding value in inviting independent scrutiny of clinical trial data at the participant level. In addition to increasing accuracy and trust, accessible trial data can be used to generate new research hypotheses and validate existing research. Academic trial investigators need to be incentivized to catch up with this encouraging trend.

Although third-party reanalysis ensures data integrity and brings many new insights, we think it is a mistake to separate data producers from their data set—they need to get publication credit. If data sharing were mandatory, existing authoring conventions would allow trial investigators to coauthor and take shared responsibility for data reanalyses or, if they disagree with the reanalyses, to be named collaborators, responsible only for the integrity of the data set.

The most commonly expressed fear we have heard from trial investigators is that, when they share data, improperly executed third-party analysis might be rapidly disseminated, creating public confusion or irreversible adverse consequences. However, when data generators are coauthors, they can identify and redirect flawed analyses. As collaborators taking responsibility only for the trial data, they can track the publication of analyses with which they disagree. In the latter case, we believe journals have a responsibility to publish dissenting analyses by data-provider groups in the event that data reuse generates differing analytical interpretations. We would insist that these pass normal peer review and the journal's usual quality filters. In some cases, public interest might require the journal to commission a commentary or take an editorial position to explain different conclusions, but we think that these proposals for mutual responsibility go a long way to encourage independent reanalysis and deter sensational misuse of data.

Another concern is that investigators need sufficient time to analyze the results of their trials. To ask for some moratorium on competing publication is reasonable, but current processing times for trials are not helping this argument. Despite journals and funders mandating registration of trials at ClinicalTrials.gov, fewer than 50% of NIH-funded trials achieve a journal publication within 30 months of the trial's end point (*BMJ* 344, d7292, 2012). We think that a 1-year moratorium is a reasonable request that is compatible with immediate access to results and thorough concurrent analyses.

At a recent meeting organized by the Institute of Medicine (<http://tinyurl.com/IOM-sharing-clinical-data>), we learned that pharmaceutical companies can regard opening their trial data to independent scrutiny as both a form of insurance and a way to build customer trust. Medtronic has funded the

Yale University Open Data Access project to review trial data generated with recombinant human bone morphogenetic protein-2 (http://medicine.yale.edu/core/projects/yodap/medtronic_data/medtronic_data.aspx), and GlaxoSmithKline recently announced that they will allow access to data from their global trials since 2007 and all trials from 2013 (*New York Times* (11 October 2012), <http://tinyurl.com/nytimes-gsk>). The Alzheimer's Disease Neuroimaging Institute (ADNI) trial of imaging in early stages of Alzheimer's disease (<http://www.fnih.org/work/areas/chronic-disease/adni>), funded by the NIH and 20 pharmaceutical companies, has already been accessed by over 400 researchers. Public interest also drives transparency. The European Medicines Agency (Europe's equivalent of the FDA) is making clinical trial data it holds available in response to prompting from the European Ombudsman (<http://tinyurl.com/Reuters-EMA>).

But what will data sharing mean for research trial participants whose consent was obtained with the expectation of privacy? Firstly, we encourage patients and their support groups to engage in research and researchers to treat subjects as partners who should receive credit and information in exchange for participation. If the ethical approval and trial design permit participants to identify their own data, they may derive information that is useful to them in understanding their own condition. However, because trial participation is still very limited and privacy is very important to most people, we think that patients and researchers alike should refrain from attempting to identify individuals, as this is likely to create ethical and legal problems that can discourage future participation.

The FDA is proposing to turn its surveillance tool Mini-Sentinel (<http://mini-sentinel.org/>) into a data-sharing protocol. This system permits queries for deidentified data held by the FDA through a firewall. This proposal is a fine idea and scales up to search data held at distributed sites. We would like to see provenance information follow the data wherever possible, as trial investigators and participants deserve credit for their work and may be able to help analysts understand the formatting and limitations of the data.

Genomic data generated in randomized controlled trials will inevitably be standard evidence to change medical practice. Genome sequences are a way to characterize the biological properties of the people who make up trial populations and thereby reduce costs and increase trial effectiveness. Genome data come with a research culture of data access. We think the academic culture of giving credit is an incentive system that should not be lost as we anonymize and aggregate large data sets. ■