JACC: HEART FAILURE VOL. ■, NO. ■, 2018

PUBLISHED BY ELSEVIER

PERSPECTIVE

Opening Opportunities With Open Data

Alexander R. Zheutlin, BS, James Brian Byrd, MD, MS

"It is no use to keep private information which you can't show off."

—Mark Twain (1)

Researchers in many fields have begun to share data. Increasingly, researchers are sharing data without judging data requests as a gatekeeper or requiring authorship. Despite over a decade of calls for data sharing, the culture of medicine changes slowly, and reluctance predominates over willingness to share. A group of cardiovascular clinical trialists recently advocated for 2 years of exclusive access to data generated by their teams, with the principal investigator involved in subsequent vetting of requests. The proposal suggests trepidation about the impact of data sharing on clinical trialists' endeavors. Little if anything has been said about the potential for cardiovascular clinical trialists to benefit from an environment in which sharing data is routine. Yet, evidence is building to suggest that when data sharing begins in earnest, it will be a boon to early career and senior clinical trialists, as well as to the patients who inspire us to study the unknown.

Access to others' datasets can help early career clinical trialists establish their reputation before they are entrusted with the resources to lead a large-scale trial.

There is momentum among early career researchers toward production of shared and open data (2). This shift is sensible because greater access to data enhances training and publication opportunities for early career investigators. In a survey of early career investigators, many reported that they could not access data they requested, limiting the progress of their research and diminishing the sense of collegiality with other researchers (3). No wonder that early career investigators are actively trying to improve the culture of data sharing.

Early career clinical trialists have a special need for access to others' data. Imagine an aspiring academic clinical trialist spends several hundred hours leading recruitment at a site within a multicenter clinical trial. For this time-consuming contribution, she will likely be listed "in the box," that is, the list of contributors in the appendix of a manuscript. This low-visibility type of recognition confers negligible academic credit. If she does not have a strong advocate-style mentor who holds sway with the principal investigators, she may well be in the position of having nothing to show, in academic terms, for the hours she spent on this project. Thus, she needs the opportunity to analyze data from the trial on which she worked—or other high-quality clinical trials, which are perfectly reasonable "substitute goods" for establishing her bona fides.

Now, contrast the current ecosystem with an environment in which data sharing is common. The data-rich environment will ensure early career trialists have more opportunities to perform secondary analyses of the most important clinical trials. Early career clinical trialists will be able to pursue questions fitted to their intellectual passions, rather than constrained by the datasets they can access through lobbying or favors.

Since clinical trials often take years to conduct, junior clinical trialists typically do not have publishable data from their own work very quickly. A common arrangement has been for data access to be granted by a senior clinical trialist in reciprocity for the junior trialist's work on an ongoing trial led by the senior investigator.

From the Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan. Dr. Byrd is supported by grant K23HL128909 from the National Heart, Lung, and Blood Institute. Mr. Zheutlin has reported that he has no relationships relevant to the contents of this paper to disclose.

Manuscript received December 18, 2017; accepted December 29, 2017.

ISSN 2213-1779/$36.00 https://doi.org/10.1016/j.jchf.2017.12.019

A potential concern is that if the junior investigator doesn't need a senior investigator to obtain clinical trial data, the junior investigator might be less likely to do things like recruit for a clinical trial.

It's true that authorship can be a benefit of serving as a site investigator for a clinical trial—although it is far from guaranteed. But the opportunity to publish will be expanded, not diminished in an era of data sharing. It is possible that some opportunities that would have been available exclusively to members of the clinical trial team will be diluted by others' access to the data. Fortunately, new opportunities provided by increased data availability will offset this concern. At the same time, publications are not the sole motivator for early career researchers who choose to work on clinical trials. Curiosity, irreverence for the current state of knowledge, and the desire to improve medicine motivate bright young people to work on new clinical trials, as has been the case for generations before.

Access to others' datasets can help senior investigators apply their unique expertise to relevant data from other groups, including groups that might find the research question "out of scope" for their involvement.

The benefits of clinical trial data sharing are in no way limited to the early career clinical investigator. When data from clinical trials groups are made available with only privacy-preserving restrictions and not promises of copublication, senior investigators will also have access to a much richer data environment. This circumstance will allow investigators with a special interest in a topic to ask unique questions outside the scope of interest of the research group who created the dataset. Furthermore, open data allows for a reconceptualization of collaboration between domain experts who may never have crossed paths otherwise. Collaborations will likely be based on a more pure consideration of the value added by each collaborators' expertise, rather than by who can access which datasets.

The potential for data sharing to increase the impact of a clinical trial team's work is becoming evident. Recently, one of us (J.B.B.) collaborated with a computational biologist, who created 2 neural networks in competition, one to generate privacy-preserving synthetic SPRINT (Systolic Blood Pressure Intervention Trial) data for analysis and one to "bust" the other's synthetic data (4). These neural networks produced synthetic data that were not distinguishable by a hypertension expert from genuine SPRINT trial data. This progress in creating analyzable data that can be shared while preserving privacy resulted from the relatively open availability of the SPRINT trial data. When creative projects of this type get done, science progresses at an accelerated pace—and without loss to the senior investigators involved in the clinical trial. Among publications describing microarray data from cancer clinical trials, authors who made data available to the public were cited 69% more often compared with those who did not (5). When data sharing is widespread, the principal investigator of the group that has generated the data will benefit from greater data availability and will be able to point to the reuse of their data as additional evidence of the value they create.

The availability of data builds the credibility of clinical researchers. As funding agencies, the public, and media become more interested in seeing data shared, early mover advantage will exist, but will not persist.

The coming revolution in clinical trials data sharing will bolster the reputation of trialists and will help ensure the long-term health of clinical trialists' careers. By sharing their data, clinical trialists' amplify their imprint on the world, and further honor their intention to improve patients' lives. The future may see new ways of recognizing those who share data in addition to the traditional means of conferring credit.

Among the many ideas that will help bring this cultural shift about, journals could recognize those who generated clinical trial data by providing a new article type, an invited reflection on how the data have been used some years after the data are made public (6). In addition, the new Annual Research Symbiont Awards co-founded by one of the authors (J.B.B.) recognizes excellence in data sharing (6). The nominees exemplify success via sharing, not despite sharing. Evidence of this is found in the >100 publications reusing one of the datasets generated by one of the winners, including publications in the highest profile scientific journals. *Who would argue that investment in such an investigator isn't money well spent?*

Increasingly, federal funding agencies and foundations are exploring requirements for data sharing (and pharma has already been sharing data). Those clinical scientists wishing to remain most competitive for funding would be wise to heed the early signs and consider what this trend means. As data sharing transitions from an uncommon practice to the norm, it is likely that a track record of sharing data will influence one's likelihood of future funding. Metrics for monitoring data sharing are already being developed (7).

Although the benefits to clinical trialists who share data will remain nonobvious to many for

some time to come, the advantages in terms of funding will accrue—especially to those who get in on the ground level—and will greatly exceed the risks. Those who change before they have to will be in a "class of their own," and they will have no regrets.

ADDRESS FOR CORRESPONDENCE: Dr. James Brian Byrd, University of Michigan Medical School, 5570C MSRB II, 1150 West Medical Center Drive, Ann Arbor, Michigan 48109-5600. E-mail: jbbyrd@med.umich.edu.

REFERENCES

1. Samuel L. Clemens (Mark Twain). An Author's Soldiering. In: Rachels D, editor. Mark Twain's Civil War. Lexington, KY: The University Press of Kentucky, 2007:79-82.

2. Farnham A, Kurz C, Öztürk MA, et al. Early career researchers want Open Science. Genome Biol 2017;18:221.

3. Vogeli C, Yucel R, Bendavid E, et al. Data withholding and the next generation of scientists: results of a national survey. Acad Med 2006;81:128-36.

4. Beaulieu-Jones BK, Wu ZS, Williams C, Byrd JB, Greene CS. Privacy-preserving generative deep neural networks support clinical data sharing 2017. bioRxiv Preprint Service. Available at: https://www.biorxiv.org/content/early/2017/11/15/159756. http://doi.org/10.1101/159756. Accessed December 17, 2017.

5. Piwowar HA, Day RS, Fridsma DB. Sharing detailed research data is associated with increased citation rate. PLoS One 2007;2:e308.

6. Byrd JB, Greene CS. Data-sharing models. N Engl J Med 2017;376:2305.

7. Olfson M, Wall MM, Blanco C. Incentivizing data sharing and collaboration in medical research-the S-Index. JAMA Psychiatry 2017;74:5-6.

KEY WORDS clinical trials, data sharing, randomized controlled trials