

Embargoed until 3pm CT/ 4pm ET, Tuesday, March 20, 2012

**Editor's Perspective –  
The Double-edged Sword of Open Access to Research Data**

John Spertus MD MPH  
Saint Luke's Mid America Heart Institute/UMKC  
4401 Wornall Road, 5<sup>th</sup> Floor MAHI  
Kansas City, MO 64111  
[spertusj@umkc.edu](mailto:spertusj@umkc.edu)

Word Count: 1005

In this issue of *Circulation: Cardiovascular Quality and Outcomes*, 2 articles have made a compelling case for an aggressive move to widespread dissemination of study data that can support other scientists' independent analyses and publications. The first article, by Gøtzsche, is a revised transcript of a speech made to the European Parliament in Brussels and the second is a thoughtful commentary by Ross and colleagues. Both of these articles make a compelling argument that open access to clinical study data would accelerate science, illuminate new opportunities to improve care, optimally leverage existing resources to generate the most knowledge in the most cost-effective manner, and are an ethical responsibility to the patients who participated in the original studies so that their experiences can help the greatest number of individuals. These are all important points and I wholeheartedly embrace the vision of these scholars. However, a number of practical issues need to be considered for such an approach to become a reality.

First, there needs to be a means of insuring that data are appropriately analyzed so that the results of any published findings are, in fact, true. While the primary analysis of a randomized trial is straightforward, investigators often use these data to examine other treatments and outcomes within a longitudinal study. While there are methods for comparing treatments or patient characteristics that are not randomly distributed in the study, such as propensity or instrumental variable techniques, a recent review of the application of these methods in cardiovascular studies revealed that less than 5% were conducted optimally.<sup>1</sup> While the proponents of open science believe that the peer-review process is capable of filtering the quality of research, the other 95% of studies reviewed by Austin were published in some of the most prestigious cardiovascular journals; and these journals' review processes failed to identify the deficiencies that he did. Given the paucity of reviewers capable of understanding and critiquing advancing statistical methods, the challenges for such reviewers to devote uncompensated time to the peer review process and the proliferation of journals, I worry that inaccurate conclusions may be drawn from publically available data and that patients adversely treated on the basis of such fallacious insights will be harmed. Since an important source for such data will be large clinical trials, which have garnered respect from the medical community, incorrect analyses and conclusions may be given added credibility and accelerate adoption of misinformation.

Second, each study has its challenges, whose limitations are best known by the investigators that conducted the study. For example, some data fields may not be adjudicated or to have been collected in an imprecise way. While the original investigators would know of these weaknesses, others accessing the data may not. For example, the National Cardiovascular Data Registry collect elements, such as peri-procedural myocardial infarction, that would understandably be of great interest to clinical investigators. Not knowing the enormous variability in the surveillance of peri-procedural infarcts across institutions might introduce important confounding in analyses of this outcome, and the extensive review process and restricted analyses of NCDR data prevent such mistakes from occurring. Similarly, there is often

missing follow-up data in clinical studies and understanding the magnitude and impact of these data (and properly analyzing such fields) is challenging. cursory analyses that are insensitive to these limitations can lead to erroneous conclusions that could be averted by insuring that the original investigators are involved in the analyses of their studies and they could prevent these types of potential errors.

Finally, the funding of secondary analyses of clinical data is limited. While the recent decision of Medtronic to provide and fund analyses of their entire clinical experience with recombinant bone morphogenetic protein-2 is not only laudable and precedent setting, it is by no means the norm.<sup>3</sup> In fact, recent decisions to eliminate R21 grants by the National Heart Lung and Blood Institute (NHLBI) means that even fewer resources to support such analyses are available.<sup>4</sup> Dr. Ross notes that substantial resources are needed to collate and annotate the data from a study so that future investigators are able to analyze it. Without adequate funding or other mandates, however, I fear that this will be difficult to accomplish. Efforts by the NHLBI to host conferences to educate researchers as to how best to analyze its large epidemiological databases are excellent examples of how this can be done. However, without investing in such an infrastructure, it is hard to imagine how high quality science can emerge from the mere availability of more data.

These concerns are not meant to discourage the movement to open science and the sharing of data. I wholeheartedly concur with the sentiment of the authors of this issue's articles calling for open science. However, we need to define how best to invest in the infrastructure of open science. One model is that used by the Cardiovascular Outcomes Research Consortium, which I co-founded with Dr. Krumholz. Using a web-based infrastructure ([www.cvoutcomes.org](http://www.cvoutcomes.org)), we enable all investigators in the study (and outside investigators who request permission) to post analytic proposals to the website. This enables all involved with the study to become aware of, and to participate in, these research ideas, thus elevating the review and import of the proposals. A Publications Committee then reviews and approves proposals and all analyses are conducted by the original study statisticians who are most familiar with the nuances of the dataset. Articles produced from such analyses then undergo extensive peer-review within the Consortium, prior to journal submission. Creating such an infrastructure within all clinical studies would be a huge advance, although its funding remains a challenge. Whether the funders of the original research, including governmental agencies, foundations and industry, would consider such support is an open question, but one that could clearly maximize the intellectual contributions and impact of their original studies. The dialogue in this issue of *Circulation: Cardiovascular Quality and Outcomes* should serve as an import stimulus to solving the challenges of open science. Clearly our patients and practicing community will benefit from an innovative solution to this important problem.

#### **References:**

1. Austin PC. Primer on statistical interpretation or methods report card on propensity-score matching in the cardiology literature from 2004 to 2006: a systematic review. *Circ Cardiovasc Qual Outcomes*. Sep 2008;1(1):62-67.
3. Krumholz HM, Ross JS. A model for dissemination and independent analysis of industry data. *Jama*. Oct 12 2011;306(14):1593-1594.
4. National Heart Lung and Blood Institute. NHLBI Will No Longer Participate in the Investigator-Initiated Innovative Research Grant (R21) Program. 2011; <http://www.nhlbi.nih.gov/funding/r21.htm>. Accessed February 19, 2012.