

**EDITOR'S PERSPECTIVE**

**Open Science and Data Sharing in Clinical Research:**

**Basing Informed Decisions on the Totality of Evidence**

**Short Title: A New Era of Open Science and Data Sharing**

Harlan M. Krumholz, MD, SM

Section of Cardiovascular Medicine and the Robert Wood Johnson Clinical Scholars Program,  
Department of Medicine, Yale University School of Medicine; Section of Health Policy and  
Administration, Yale School of Public Health; Center for Outcomes Research and Evaluation,  
Yale-New Haven Hospital, New Haven, CT

Dr. Krumholz is supported by grant U01 HL105270-02 (Center for Cardiovascular Outcomes  
Research at Yale University) from the National Heart, Lung, and Blood Institute.

He discloses that he is the recipient of a research grant from Medtronic, Inc. through Yale  
University and is chair of a cardiac scientific advisory board for UnitedHealth.

**Correspondence:**

1 Church Street, Suite 200, New Haven, CT 06510

203-764-5885; (f) 203-764-5653; [harlan.krumholz@yale.edu](mailto:harlan.krumholz@yale.edu)

**Word count:** 1726

A patient tentatively enters the examination room. A decision looms and the patient waits with some trepidation. A doctor soon appears and pulls up a chair. The conscientious physician has prepared for the discussion by reviewing the medical literature—examining the evidence-based systematic reviews and highlighting the relevant information. The ensuing discussion reflects on the patient’s goals and the published information about the benefits and harms of the available treatment options. The tension breaks with a decision and plans for the next steps are put in place.

Every day, patients and their caregivers are faced with difficult decisions about treatment. They turn to physicians and healthcare professionals to interpret the medical evidence and assist them in making decisions tailored to their needs and goals.

Unfortunately, we are learning that what is published in the medical literature represents only a portion of the evidence relevant to the risks and benefits of available treatments. In a profession that seeks to rely on evidence, it is ironic that we tolerate a system that enables evidence to be outside of public view. The owners of data, usually scientists or industry, have the choice of what, where and when to publish. As a result, our medical literature portrays only a partial picture of the evidence about clinical strategies, including drugs and devices. Experts have recently drawn attention to this issue, including contributions in this issue of our journal, but there is resistance to change.<sup>1-3</sup>

Studies document that a remarkable percentage of trials are not published within a reasonable period after they are completed. Ross and colleagues reported that fewer than half of trials are published within 2 years of completion.<sup>4</sup> Although industry-sponsored studies tended to have the lowest publication rates, the problem spans all study sponsors. A subsequent study found that only 46% of clinical trials funded by the National Institutes of Health were published within 30 months of completion.<sup>5</sup> Lee and colleagues found that fewer than half of the new drug trials submitted to the Food and Drug Administration (FDA) are published within 5 years of drug approval.<sup>6</sup> Pivotal trials were more likely to be

published, but 24% remain unpublished at 5 years. Negative trials were less likely to be published, yet 34% of the positive trials were also unpublished at 5 years.

We tout systematic reviews, the linchpin of evidence-based medicine, as a method to distill and interpret the medical literature. The Institute of Medicine recently produced standards for these reviews and noted that, "Healthcare decision makers—including clinicians and other healthcare providers—increasingly turn to systematic reviews for reliable, evidence-based comparisons of health interventions."<sup>7</sup> Others have published checklists to promote high quality methods.<sup>8</sup> However, the elegance of the methods cannot overcome the undermining effect of missing studies.

Missing studies are unlikely to be missing at random and the effect of the missing data on inferences about an intervention is not easy to predict. Hart and colleagues investigated the effect of unpublished trial data on the results of meta-analyses of drug trials.<sup>9</sup> They showed that the unpublished data influenced the summary estimates such that the drug efficacy was lower in 46% of the meta-analyses and greater in 46%. The result was essentially unchanged in only 7% of the studies. Readers of meta-analyses, without the benefit of these types of analyses, are left to wonder how many missing trials might be relevant to the clinical question and what their effect might be.

In some cases, the missing clinical research data may hold important information about risk. A classic example occurred with Vioxx (rofecoxib). Merck had data, mostly unpublished and obtained several years before the drug was withdrawn from the market,<sup>10</sup> demonstrating that Vioxx likely increased the risk of an acute myocardial infarction. No systematic analysis could have detected this harm because most of the data were beyond public view. Similarly, GlaxoSmithKline had data, much of it unpublished, which indicated that Avandia (rosiglitazone) increased the risk of an acute myocardial infarction. Nissen and colleagues, accessing the data through litigation, revealed the concern that ultimately led to FDA restrictions on the use of the drug.<sup>11,12</sup>

Selective publication similarly skews the evidence base. For example, Psaty and Kronmal, using information obtained during litigation, reported that Merck and its consultants employed selective reporting in representing mortality results in trials of Vioxx in patients with Alzheimer’s disease or cognitive impairment.<sup>13</sup> The published studies failed to report that an intention-to-treat analysis showed Vioxx to be associated with a significant increase in all-cause mortality. Turner and colleagues showed that trials of antidepressant drugs highlighted positive results as if they were the primary outcomes, contrary to the original study protocols.<sup>14</sup> They concluded that, “By altering the apparent risk–benefit ratio of drugs, selective publication can lead doctors to make inappropriate prescribing decisions that may not be in the best interest of their patients and, thus, the public health.”

The myriad of reasons why trials are selectively or never published include professional bias, profit, journal bias against negative studies, loss of interest, lack of funding, and competing interests. When results do not confirm the beliefs of investigators, the motivation for their dissemination may weaken, as may the willingness of the investigator to devote time and resources to the publication. Companies may also lose interest when the findings are counter to embedded beliefs about a product. The end result is a scientific culture in which many experiments performed on patients are absent from the medical literature.

Another issue relevant to selective and non-publication is that access to a research database is commonly restricted to the principal investigator or the funders. For many trials, there is no opportunity for independent replication of the analyses or evaluation of the raw data. The FDA has access to these data, but outside experts do not.

The sharing of trial data, as of yet uncommon except through mechanisms by some funders such as the National Heart, Lung and Blood Institute,<sup>15</sup> could provide an opportunity to leverage the strength of the global community of investigators. Many trials yield only a fraction of the knowledge that could be produced with more resources and creativity.

Now is the time to bring data sharing and open science into the mainstream of clinical research, particularly with respect to trials that contain information about the risks and benefits of treatments in current use. This could be accomplished through the following steps:

1. Post in the public domain detailed study protocols with each published trial. The protocol should be comprehensive and include policies and procedures relevant to actions taken in the trial.
2. Develop mechanisms for those who own trial data to share their raw data and individual patient data.
3. Encourage industry to commit to place all its clinical research data relevant to approved products in the public domain. This action would acknowledge that the privilege of selling products is accompanied by a responsibility to share all the clinical research data relevant to the products' benefits and harms.
4. Develop a culture within academics that values data sharing and open science. After a period in which the original investigators can complete their funded studies, the data should be deidentified and made available for investigators globally.
5. Identify, within all systematic reviews, trials that are not published, using sources such as [clinicaltrials.gov](http://clinicaltrials.gov) and regulatory postings to determine what is missing.
6. Share data.

The path is not easy. We lack an infrastructure and funding, and incentives in the system run counter to sharing. Those who share may be giving advantage to 'competitors.' The imperative, however, derives from what is best for society and our obligation to respect the contributions made by the subjects who agreed to participate in the research studies.

Patients facing a decision deserve for information to be based on all of the evidence. A new era of data sharing and open science would allow us to leverage existing investments

to provide more and better evidence that will increase the possibility that patients' decisions will help them obtain the results they desire.

## References

1. Gøtzsche P. Strengthening and opening up health research by sharing our raw data. *Circ Cardiovasc Qual Outcomes*. 2012;5:xxx-xxx.
2. Lehman R, Loder E. Missing clinical trial data. *BMJ*. 2012;344:d8158.
3. Ross JS, Lehman R, Gross CP. The importance of clinical trial data sharing: toward more open science. *Circ Cardiovasc Qual Outcomes*. 2012;5:xxx-xxx.
4. Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. Trial publication after registration in ClinicalTrials.gov: a cross-sectional analysis. *PLoS Med*. 2009;6:e1000144.
5. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *BMJ*. 2012;344:d7292.
6. Lee K, Bacchetti P, Sim I. Publication of clinical trials supporting successful new drug applications: a literature analysis. *PLoS Med*. 2008;5:e191.
7. Committee on Standards for Systematic Reviews of Comparative Effectiveness Research, Institute of Medicine. Finding what works in health care: standards for systematic reviews. The National Academies Press; 2011.  
<http://www.iom.edu/Reports/2011/Finding-What-Works-in-Health-Care-Standards-for-Systematic-Reviews.aspx>. Accessed February 21, 2012.
8. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA*. 2000;283:2008-12.
9. Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. *BMJ*. 2012;344:d7202.

10. Ross JS, Madigan D, Hill KP, Egilman DS, Wang Y, Krumholz HM. Pooled analysis of rofecoxib placebo-controlled clinical trial data: lessons for postmarket pharmaceutical safety surveillance. *Arch Intern Med*. 2009;169:1976-85.
11. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*. 2007;356:2457-71.
12. United States Food and Drug Administration. FDA drug safety communication: updated risk evaluation and mitigation strategy (REMS) to restrict access to rosiglitazone-containing medicines including Avandia, Avandamet, and Avandaryl. <http://www.fda.gov/Drugs/DrugSafety/ucm255005.htm>. May 18, 2011. Accessed February 21, 2012.
13. Psaty BM, Kronmal RA. Reporting mortality findings in trials of rofecoxib for Alzheimer disease or cognitive impairment: a case study based on documents from rofecoxib litigation. *JAMA*. 2008;299:1813-7.
14. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358:252-60.
15. National Heart, Lung, and Blood Institute. NHLBI policy for data sharing from clinical trials and epidemiological studies. <http://www.nhlbi.nih.gov/funding/datasharing.htm>. Accessed February 21, 2012.