Open Science and Data Sharing in Clinical Research
Basing Informed Decisions on the Totality of the Evidence

Harlan M. Krumholz, MD, SM

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, prepared for the discussion by having conscientiously reviewed the medical literature, examined the evidence-based systematic reviews, and highlighted the relevant information. The ensuing conversation reflects on the patient’s treatment goals and the published information about the benefits and harms of the available 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 other healthcare professionals to interpret the medical evidence and assist them in making individualized decisions.

Unfortunately, we are learning that what is published in the medical literature represents only a portion of the evidence that is 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. Those who own 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.1–5

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 were published within 2 years of completion.4 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.5 Lee and colleagues found that fewer than half of the new drug trials submitted to the US Food and Drug Administration were published within 5 years of drug approval.7 Although pivotal trials were more likely to be published, 24% remained unpublished at 5 years.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.”6 Others have published checklists to promote high-quality methods; 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 data on the results of meta-analyses of drug trials.10 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,11 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 made available through litigation, revealed the concern that ultimately led to US Food and Drug Administration restrictions on the use of the drug.12,13

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 disease or cognitive impairment.14 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.15 They concluded that, “By altering the apparent risk-benefit ratio of drugs, selective publication can lead

© 2012 American Heart Association, Inc.

*Circ Cardiovasc Qual Outcomes is available at http://circoutcomes.ahajournals.org
DOI: 10.1161/CIRCOUTCOMES.112.965848

The opinions expressed in this article are not necessarily those of the American Heart Association.

From the 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, Conn.

Correspondence to Harlan Krumholz, 1 Church St, Suite 200, New Haven, CT 06510. E-mail harlan.krumholz@yale.edu
(Circ Cardiovasc Qual Outcomes. 2012;5:141-142.)
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. For example, 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 nonpublication 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 US Food and Drug Administration has access to these data, but other 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, 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, the study protocol for 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 de-identified and made available for investigators globally.
5. Identify, within all systematic reviews, trials that are not published, using sources such as 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 information that is 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.

Sources of Funding
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.

Disclosures
Dr Krumholz 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.

References

Keywords: evidence-based medicine • data sharing • medical decision-making • open access
Open Science and Data Sharing in Clinical Research: Basing Informed Decisions on the Totality of the Evidence
Harlan M. Krumholz

*Circ Cardiovasc Qual Outcomes*. 2012;5:141-142
doi: 10.1161/CIRCOUTCOMES.112.965848
*Circulation: Cardiovascular Quality and Outcomes* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/5/2/141

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Circulation: Cardiovascular Quality and Outcomes* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Circulation: Cardiovascular Quality and Outcomes* is online at:
http://circoutcomes.ahajournals.org/subscriptions/