

The Importance of Clinical Trial Data Sharing: Toward More Open Science

Running Title: Clinical Trial Data Sharing

Joseph S. Ross, MD, MHS^{1,2}; Richard Lehman, MRCGP²; Cary P. Gross, MD¹

¹ Section of General Internal Medicine, Department of Medicine, Yale University School of Medicine; ² Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, both New Haven, CT, USA.

Correspondence: Joseph S. Ross, MD, MHS, Section of General Internal Medicine, Yale University School of Medicine, P.O. Box 208093, New Haven, CT 06520, Phone: (203) 785-2987, Fax: (203) 737-3306, E-mail: joseph.ross@yale.edu.

Funding/support and role of the sponsor: This manuscript was not supported by any external funds. All authors receive support from Medtronic, Inc., to develop methods for clinical trial data sharing. The ideas and opinions expressed are the authors'. The content of this publication does not necessarily reflect the views or policies of Medtronic, Inc. and was not subject to review or approval prior to submission or publication. Dr. Ross is supported by the National Institute on Aging (K08 AG032886) and by the American Federation for Aging Research through the Paul B. Beeson Career Development Award Program.

Conflicts of interest: Drs. Ross and Gross are members of a scientific advisory board for FAIR Health, Inc.

Author contributions: Dr. Ross drafted the manuscript and all authors critically revised the manuscript for important intellectual content.

Word Count (text only): 1995

References: 28

In cardiovascular medicine as in all other medical disciplines, realizing the full value of clinical trial research data requires that the data be accessible to the research community and others who might be able to use them. Traditionally, the dissemination of knowledge derived from clinical research has been limited in scope: investigators who have designed and conducted clinical trials make the decisions about which statistical analyses to conduct and then publish peer-reviewed articles to disseminate their findings. Clinical trial data are considered the property of the investigators and the entities that sponsored the research, with little or no opportunity for investigators external to the original study team to access the data. This traditional model is based on dissemination via print publication, the origins of which date back to the seventeenth century.

By continued adherence to this model in the age of electronic knowledge exchange, our understanding of clinical interventions is limited by our lack of access to comprehensive data from all clinical trials in several ways. First, a select number of individuals decide which analyses to conduct, choosing some at the exclusion of others, while an analysis that might have been of great interest to another investigator – and which may have a direct bearing on clinical practice – may not be performed. Second, among these findings generated, a select number might be included in any peer-reviewed publication, leaving the research community and clinicians at a loss to know about what findings were generated and not disseminated. In fact, by comparing published articles to trial protocols, 50% of efficacy and 65% of harm outcomes per trial have been shown to be incompletely reported and biased toward the reporting of statistically-significant findings.¹ Third, among all trials conducted, there may be significant publication delays, as happened with the Ezetimibe and Simvastatin in Hypercholesterolemia Enhances Atherosclerosis Regression [ENHANCE] trial,² which was completed in April 2006 but the findings of which were not released until after substantial coverage in the news media in January 2008.³ Finally, only a limited number of trials are eventually published. By examining trials registered with an Institutional Review Board or the publicly-available trial registry ClinicalTrials.gov,⁴ submitted to the U.S.

Food and Drug Administration as part of new drug applications, or presented as research abstracts at National Scientific meetings, it has been estimated that between 25% and 50% of completed trials remain unpublished.⁵⁻¹²

The cumulative effect is that patients, physicians and other health care professionals, and the research community are placed in the position of making clinical or research decisions with access to only a fraction of the relevant clinical evidence that might otherwise be available. Making clinical research data available outside individual pharmaceutical companies or clinical research groups has obvious value, in terms of validation, reproduction, and optimization of new knowledge generated from clinical research. But why are data not made more widely available to the scientific community? In this commentary, we will review some of the common concerns about data sharing, share some prominent examples of data sharing currently underway in cardiovascular clinical research, and conclude with our expectations for more open scientific and information exchange through data sharing that will increase the value of all clinical trial research.

Data Sharing Trials and Tribulations

Data sharing is increasingly common in some areas of medical research, particularly among genomics investigators and research groups engaged in systematic reviews and meta-analyses. However, individual, patient-level clinical trial data sharing is less common because of concerns among investigators and challenges with the actual act of data sharing. The principal concern, voiced primarily by investigators, is that a substantial amount of individual time and effort has been invested to design the trial and collect the data and that, in return, they deserve ample opportunity to conduct their analyses and disseminate their findings. Without question, investigators do deserve some period of respite during which they can prioritize their analyses and publish their work. However, a recent study found that fewer than half of trials funded by the National Institutes of Health (NIH) are published

within 2 and a half years of completion.¹² Dissemination delays exceeding 2 years inevitably slow and diminish the impact of any research. While investigators may be concerned about being “beaten to the punch” with their own data, they should focus their attention on the fact that the time and effort they have invested has not resulted in data that is being fully used to further scientific understanding and improve patient care.

Other objections to data sharing are also frequently raised,¹³ including concerns that multiple analyses by various independent research groups will produce analyses with differing results, either because of human error or because external investigators conduct inappropriate analyses; that clinical trials are designed with pre-specified study protocols and that additional analyses amount to “data-dredging”; and that data ownership belongs by right to the original investigator team. However, the scientific community is well positioned to review and put into context differing results from the same trial data, as well as to judge whether data has been “dredged” or appropriately analyzed. Regarding ownership, Vickers has posed the rhetorical question “whose data set is it anyway?”¹⁴ He posits that while the data legally belong to the investigators, science, particularly medical science, is essentially an enterprise conducted for moral reasons.¹⁴

Data sharing is a complex undertaking; the scientific community must reach a consensus about several critical points before the promise of sharing clinical research data can be realized. First, what are the responsibilities of the original investigator team? To share data effectively, they must produce a clean, well-described, and accurate data file that can be used by others and protects patient confidentiality. Second, who supports their effort to create this data source? Third, what if there are subsequent questions and inquiries – who bears responsibility for the shared data?

Another broad issue is the question of who owns the data, and who should be allowed to access the data. Is access unfettered or should some minimal application or registration system be used to minimize data-dredging and incentivize pre-specification of analyses using shared clinical trial data? If

the latter, who is responsible for reviewing these applications? Should there be a commitment to publish these analyses, or at least report results on a central repository akin to ClinicalTrials.gov?⁴

Finally, where should the data be placed for others to access? Is it the responsibility of journals to house data from affiliated publications, as is currently done by the journal *Trials*,¹⁵ an open access, peer-reviewed, online journal that publishes on all aspects of the performance and findings of randomized controlled trials? In this and other models, who should support this effort and pay the costs of maintaining internet accessibility?

While many of these questions remain unanswered, the scientific community has begun the process of developing standards and solutions to common problems. The Institute of Medicine set forth recommendations on managing research data in the information age.¹⁶ The Wellcome Trust convened a number of research funders to develop a coherent vision, principles and goals to promote the sharing of research data to improve public health,¹⁷ supporting organizations which include the World Bank, the NIH, and the Bill and Melinda Gates Foundation.¹⁸ Journal editors have developed guidance for the preparation of raw clinical data for publication.¹⁹

Current Data Sharing Initiatives

There are several prominent examples of data sharing currently underway in cardiovascular clinical research that are illustrative and can inform our expectations for open scientific and data exchange.

NIH Data Sharing requirements and NHLBI

The NIH has implemented a policy to share clinical trial data, placing a priority on making the results and accomplishments of the activities that it funds publicly available. All investigator-initiated applications with direct costs greater than \$500,000 in any single year are expected to address data

sharing within their grant proposals.²⁰ However, in practice, the sharing of data varies widely among the Institutes, and the National Heart, Lung, and Blood Institute (NHLBI) clearly leads the way. Alone among the Institutes, NHLBI has established the Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC),²¹ which provides centralized access to more than 100 clinical trials originally funded by NHLBI, including Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), Coronary Artery Risk Development in Young Adults (CARDIA), and Multiple Risk Factor Intervention Trial for the Prevention of Coronary Heart Disease (MRFIT). To access this data, one need only register as a BioLINCC user and submit a simple request form for review by an NHLBI official that essentially requires a brief overview of your research needs, a research plan or protocol, proof of ethics committee review, the principal investigator's curriculum vitae, and a research materials distribution agreement.

The International Stroke Trial

The investigators who designed and conducted the International Stroke Trial (IST) have also been leaders in data sharing. The IST, conducted between 1991 and 1996, was a large, prospective, randomized controlled trial of nearly 20,000 individuals to determine whether early administration of aspirin, heparin, both or neither influenced the clinical course of acute ischemic stroke.²² This trial was originally funded by multiple agencies, most prominently by the UK Medical Research Council, the UK Stroke Association, and the European Union BIOMED-1 program. These data has now been made available for public use, to facilitate the planning of future trials and to permit additional secondary analyses.²³ As part of making these trial data available, the investigators explain the process of anonymisation of the data and ethics committee review. This is a critical issue in data sharing, given that consent for publication of raw data is not routinely obtained from study subjects.

The Yale University Open Data Access Project

We have engaged in a project to promote and facilitate sharing of industry clinical trial data, led by Principal Investigator Dr. Harlan Krumholz, the Editor of *Circulation: Cardiovascular Quality and Outcomes*. The Yale University Open Data Access project, while not specifically directed at cardiovascular research, aims to create a model that can be applied to the complete range of medical interventions.^{24, 25} Through a grant from Medtronic, Inc., we have developed a model to facilitate access to patient-level clinical research data to promote wider availability of clinical trial data and independent analysis by external investigators. As an initial effort, we are coordinating reviews of the safety and effectiveness of INFUSE[®], Medtronic's recombinant bone morphogenetic protein-2 (rhBMP-2) product, by two independent research groups and will subsequently disseminate all clinical research data on INFUSE[®] provided to us by Medtronic to external investigators. This effort is intended to provide a means to ensure access to comprehensive clinical research data currently owned by industry or sponsored by any other funder.

The Way Forward

Increased open science and information exchange through data sharing will further the value of all clinical trial research. Most of the data from clinical trials in cardiovascular medicine are currently not available to the scientific and clinical communities. When providers recommend treatment options to patients, this is routinely done on the basis of information which is biased and seriously incomplete. This standard of practice is tolerated not because it has any intellectual or ethical justification, but because we are accustomed to it. The clinical and research community often only becomes aware of its shortcomings when safety concerns are raised about a drug, device or other treatment strategy. Past experiences with rofecoxib (Vioxx),²⁶ rosiglitazone (Avandia),²⁷ and oseltamivir (Tamiflu)²⁸ have

illustrated that it is in the public's interest to have access to comprehensive clinical trial data to ensure a complete understanding of drug or device safety and effectiveness.

Yet clinical trial data sharing goes beyond product safety concerns. Science is a community, continually building upon one another's ideas. In the era of electronic knowledge exchange, when open access to data has become an accepted norm in most of science, we need a better way of working together. Only by making individual patient data available to the whole research community can we derive full benefit from the enormous resources devoted to human clinical trial research and maintain patient trust in the research process.

References

1. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291:2457-2465.
2. Kastelein JJ, Akdim F, Stroes ES, Zwinderman AH, Bots ML, Stalenhoef AF, Visseren FL, Sijbrands EJ, Trip MD, Stein EA, Gaudet D, Duivenvoorden R, Veltri EP, Marais AD, de Groot E. Simvastatin with or without ezetimibe in familial hypercholesterolemia. *N Engl J Med*. 2008;358:1431-1443.
3. Greenland P, Lloyd-Jones D. Critical lessons from the ENHANCE trial. *JAMA*. 2008;299:953-955.
4. U.S. National Institutes of Health. ClinicalTrials.gov. Available at: <http://www.clinicaltrials.gov/>. Accessed February 5, 2012.
5. Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet*. 1991;337:867-872.
6. Krzyzanowska MK, Pintilie M, Tannock IF. Factors associated with failure to publish large randomized trials presented at an oncology meeting. *JAMA*. 2003;290:495-501.
7. 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;358:252-260.
8. Decullier E, Lheritier V, Chapuis F. Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ*. 2005;331:19.
9. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA*. 1992;267:374-378.
10. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, Decullier E, Easterbrook PJ, Von Elm E, Gamble C, Ghera D, Ioannidis JP, Simes J, Williamson PR. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS ONE*. 2008;3:e3081.

11. 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.
12. 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.
13. Kirwan JR. Making original data from clinical studies available for alternative analysis. *J Rheumatol.* 1997;24:822-825.
14. Vickers AJ. Whose data set is it anyway? Sharing raw data from randomized trials. *Trials.* 2006;7:15.
15. Hrynaszkiwicz I, Altman DG. Towards agreement on best practice for publishing raw clinical trial data. *Trials.* 2009;10:17.
16. Institute of Medicine. *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age.* Washington, DC: National Academy Press; 2009.
17. Walport M, Brest P. Sharing research data to improve public health. *Lancet.* 2011;377:537-539.
18. Wellcome Trust. Sharing research data to improve public health: joint statement of purpose, signatories to the joint statement. Available at: <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/Signatories-to-the-joint-statement/index.htm>. Accessed February 6, 2012.
19. Hrynaszkiwicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *BMJ.* 2010;340:c181.
20. U.S. Department of Health and Human Services, National Institutes of Health, Office of Extramural Research. NIH Data Sharing Policy. April 17, 2007. Available at: http://grants.nih.gov/grants/policy/data_sharing/. Accessed February 6, 2012.
21. U.S. Department of Health and Human Services, National Institutes of Health, National Heart Lung and Blood Institute. Biologic Specimen and Data Repository Information Coordinating

- Center April 17, 2007. Available at: <https://biolincc.nhlbi.nih.gov/home/>. Accessed February 6, 2012.
22. The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke. International Stroke Trial Collaborative Group. *Lancet*. 1997;349:1569-1581.
 23. Sandercock PA, Niewada M, Czlonkowska A. The International Stroke Trial database. *Trials*. 2011;12:101.
 24. Krumholz HM, Ross JS. A model for dissemination and independent analysis of industry data. *JAMA*. 2011;306:1593-1594.
 25. Yale School of Medicine. Yale University Open Data Access (YODA) Project. Available at: <http://medicine.yale.edu/core/projects/yodap/index.aspx>. Accessed February 6, 2012.
 26. 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;169:1976-1985.
 27. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med*. 2007;356:2457-2471.
 28. Doshi P. Neuraminidase inhibitors--the story behind the Cochrane review. *BMJ*. 2009;339:b5164.