Implications of clinical trial data sharing for medical writers

Joseph S. Ross,[1,2,3] Harlan M. Krumholz[2,3,4,5]

[1]*Department of Medicine, Section of General Internal Medicine, Yale University School of Medicine, USA*
[2]*Department of Medicine, Robert Wood Johnson Foundation Clinical Scholars Program, Yale University School of Medicine, USA*
[3]*Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, Yale University School of Medicine, USA*
[4]*Department of Medicine, Section of Cardiovascular Medicine, Yale University School of Medicine, USA*
[5]*Department of Epidemiology and Public Health, Section of Health Policy and Administration, Yale University School of Medicine, USA*

Correspondence to:

Joseph S. Ross, Section of General Internal Medicine, Yale University School of Medicine, New Haven, CT 0520, USA

Abstract

Major clinical research funders are increasingly adopting policies supporting or mandating data sharing. These moves should improve the transparency and availability of clinical trial data and are likely to impact the work and responsibilities of medical writers. Medical writers are likely to play a prominent role in standardising policies and procedures and have the opportunity to lead the development of an efficient and feasible system for promoting clinical trial data sharing. These efforts will ensure that the research community can derive the full benefit from the enormous resources devoted to human clinical trial research and will help build patient trust in the research process.

Keywords: Clinical trials as topic, Information dissemination, Access to information, Peer review, Research

Over the past few years, a number of major clinical research funders have adopted policies supporting or mandating data sharing. These include the US National Institutes of Health,[1] the UK Medical Research Council,[2] and the Bill and Melinda Gates Foundation.[3] Similarly, major regulators, most notably the European Medicines Agency (EMA),[4] are contemplating the adoption of open data access policies, as are several companies in the pharmaceutical and medical device industries.[5–7]

These moves toward greater transparency and availability of clinical trial data are likely to impact the work and responsibilities of medical writers, who play a major role in preparing regulatory documentation and peer-reviewed manuscripts on behalf of the clinical research team. This commentary introduces the concept of data sharing and discusses implications of clinical trial data sharing for medical writers.

Why share clinical trial data?

Most of the data generated and information elicited from clinical trials are currently not available to the scientific and clinical communities.[8–10] Today, physicians and other clinicians often recommend treatment options to patients on the basis of information that is incomplete: not all clinical trials are published and made available and not all outcomes collected during a clinical trial are reported and made available, even when the trial is published.[9,11,12]

In clinical research, data sharing is the practice of a research team making trial data available to individuals with whom they are not collaborating. Before discussing the specifics of data sharing, it is worth pausing to consider the rationale for engaging in such an exchange. Briefly, sharing clinical trial data increases the value of all clinical trial research by encouraging the use of data already collected. Sharing clinical trial data also reduces the potential for inaccurate or incomplete reporting of study outcomes that distort the medical evidence, and it ultimately ensures the reliability of the evidence base upon which clinical decisions are made by patients and physicians.[13–15] As the number of data sharing initiatives grow, the scientific community will be able to adopt a more open approach to research.

DOI: 10.1179/2047480612Z.00000000085

When clinical trial data are made more widely available, science can function as a community, continually vetting, critiquing, and building upon each other's ideas.

Clinical trial data sharing

There are two principal methods by which clinical trial data are shared. First, investigators may share trial data on their own terms in response to individual requests. Second, investigators may share trial data by depositing it in a repository, which is an archive of data with terms of access defined by the organisation that maintains it.

Sharing data in response to individual requests is less predictable because it is never clear if or when individual requests for data sharing will be made. Individuals seek out research teams who have generated clinical trial data on a subject in which they are interested, making requests to use the data for analysis, perhaps as part of a larger meta-analysis, to validate the original findings, or even to pursue a secondary question not addressed by the original research team. The decision to share data is made one request at a time and, if approved, the data are transferred directly between the research teams.

In contrast, sharing data through repositories requires a standardised approach and is predictable in practice, as investigators must prepare the data for sharing, regardless of whether the data will ever be accessed. For the repository process to be effective, the data need to be deposited in a publicly accessible repository, such as DRYAD,[16] allowing it to be used according to the rules that govern data access. The repository includes the data files, along with accompanying metadata and documentation. The decision to share is made *a priori* and data are transferred after study completion, once all data management and preparation issues have been resolved.

Data sharing – the role of medical writers

Both data-sharing methods are likely to increasingly involve medical writers, particularly for industry-funded trials and as the method of data sharing becomes standardised. Medical writers can play a critical role in several steps of the data-sharing process, but their chief responsibility is likely to be preparing the metadata and documentation needed by secondary users of shared clinical trial data. Structural metadata include the design and specification of data structures, whereas descriptive metadata include information about the data content. Both are required by any secondary user to orient them to the data structures and content, thereby improving the usability of the clinical trial data.

The research team needs to prepare a clean, well-annotated data set for deposit that includes supporting documentation to allow for secondary analysis. The data set needs to be clearly organised and follow standard logic and coding formats. Variables need to be clearly defined with no ambiguity. Without strict adherence to best practice for data preparation and documentation, the practice of data sharing will be largely limited. Moreover, standards for data definition will need to be adopted to ensure comparability of the information generated across trials and to ensure that data can be pooled for summary analysis. For example, if several different trials studying acute myocardial infarction events all use different definitions for the endpoint, the ability to summarise and interpret data across multiple studies will be limited.

Medical writers can also play a critical role in identifying the most appropriate and effective policies and advocating for funders to move towards standardised procedures. Currently, policies and procedures for preparing data for deposit in a repository vary among clinical trial research funders, and few funders are making a concerted effort to adopt uniform policies and procedures. If all clinical trial research funders adopt different policies and procedures for data sharing, such as different data definitions and documentation requirements, data sharing may devolve into an inefficient use of resources, time, and energy. As the field evolves, medical writers can lead the way towards standardised policies and procedures.

Managing data sharing

Beyond these two chief responsibilities, medical writers may be forced to consider several other technical issues when preparing data for sharing individually or through deposit in a repository.[17] These issues are likely best managed by medical writers in conjunction with other investigators from the research team as well as from the data analysts. Several of these issues are covered in brief below to introduce the potential challenges facing the field.

Defining the data

Far more data are generally collected within the context of a clinical trial than are reported in any single biomedical journal article. So how is the data set defined? Limited guidance suggests that the data set is the aggregated collection of patient observations (including socio-demographic and

clinical information) used to produce the summary statistical findings presented in the main report of the research project, whether previously published or not.[17] Thus, the data set should include all pre-specified and intentionally collected primary and secondary outcomes as well as safety endpoints, in agreement with the clinical trial registration and results reporting requirements of ClinicalTrials.gov.[18,19]

De-identification

To protect patient confidentiality, all direct and indirect patient identifiers must be removed from the data set. Experts disagree on the complexity of the task; recent presentations at a workshop at the Institute of Medicine suggested that de-identification could require an hour to an afternoon to weeks.[20] However, it is clear that patient confidentiality issues are of greater concern for small trials of rare diseases than for large, multi-centre trials of common conditions where identifying patients is extremely difficult.

Copyright/licensing agreements

Data ownership must be resolved prior to deposit. Clinical research funders may decide to prospectively address this issue by requiring that data be deposited in a repository, transferring any ownership by the funder or research team to the public. Moreover, as most clinical research data are generated through collaborations between multiple researchers (nearly all of whom are paid for their effort), it may not be possible to determine who actually owns the data.[21] Others contend that patients are the rightful owners.[22] The best guidance on this issue is from recommendations on the publication of raw data in journals, which recommend that copyright should be transferred to the publisher for publishing data sets as supplementary material, that the supporting data should be separated from the article itself, and that transfer of copyright for the data is not required as a condition of publication.[17] However, individual repositories are likely to enact their own policies.

Patient consent

Data-sharing plans are rarely discussed with patients as part of the informed consent process. While de-identification of the data may preclude the need for patient consent prior to sharing, going forward, institutional review boards and ethics committees should encourage clinical trial researchers to discuss data-sharing plans when obtaining informed consent, along with any safeguards that will be instituted to protect patient privacy. Everyone's interests would be best served if patients explicitly consented to the sharing of their de-identified clinical research data.

Conclusion

For data sharing to be successful, policies and procedures need to be standardised. Medical writers are likely to play a prominent role in these initiatives, not only by preparing data and documentation for sharing but also by establishing the standards for data definition and documentation and advocating the adoption of procedures that best support effective data sharing. Many outstanding issues remain. Medical writers have the opportunity to lead the development of an efficient and feasible system for promoting clinical trial data sharing. These efforts will ensure that the research community can derive the full benefit from the enormous resources devoted to human clinical trial research and will help build patient trust in the research process. The data generated and information elicited from clinical trials needs to be available to the scientific and clinical communities so that treatment decisions are based on all known information.

Funding/support and role of the sponsor

This work was not supported by any external grants or funds. Drs Ross and Krumholz receive support from the Centers of Medicare and Medicaid Services (CMS) to develop and maintain hospital performance measures that are used for public reporting. 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. Dr Krumholz is supported by a National Heart Lung Blood Institute Cardiovascular Outcomes Center Award (1U01HL105270-02).

Conflicts of interest

Drs Ross and Krumholz receive research support from Medtronic, Inc. to develop methods to promote data sharing. Dr Ross reports that he is a member of a scientific advisory board for FAIR Health, Inc. Dr Krumholz reports that he chairs a scientific advisory board for UnitedHealthcare.

References

1. National Institutes of Health (NIH). Final NIH statement on sharing research data. 2003. Available from: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html. [accessed 2012 Oct 6].

2. Medical Research Council (MRC). MRC policy on research data-sharing. Available from: http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/Policy/index.htm [accessed 2012 Oct 6].
3. Bill and Melinda Gates Foundation. Our commitment to sharing information. Available from: http://www.gatesfoundation.org/global-health/Pages/our-commitment.aspx [accessed 2012 Oct 6].
4. Hampton T. European drug agency works to improve transparency, but skepticism remains. JAMA 2012;308(9):850–1.
5. Krumholz HM, Ross JS. A model for dissemination and independent analysis of industry data. JAMA 2011;306(14):1593–4.
6. Loder E. Liberating Clinical Trial Data. 2012; Available from: http://blogs.bmj.com/bmj/2012/10/09/elizabeth-loder-liberating-clinical-trial-data/ [accessed 2012 Oct 15].
7. Thomas K. Glaxo opens door to data on research. New York Times, 11 October 2012.
8. 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 Int Med 2009;169(21):1976–85.
9. 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(9):e1000144.
10. 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.
11. 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(20):2457–65.
12. Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, *et al.* Systematic review of the empirical evidence of study publication bias and outcome reporting bias. PloS one 2008;3(8):e3081.
13. Gotzsche PC. Why we need easy access to all data from all clinical trials and how to accomplish it. Trials 2011;12:249.
14. Ross JS, Gross CP, Krumholz HM. Promoting transparency in pharmaceutical industry-sponsored research. Am J Public Health 2012;102(1):72–80.
15. Ross JS, Lehman R, Gross CP. The importance of clinical trial data sharing: toward more open science. Circul. Cardiovasc Qual Outcomes 2012;5(2):238–40.
16. DRYAD. Available from: http://datadryad.org/ [accessed 2012 Oct 19].
17. Hrynaszkiewicz I, Norton ML, Vickers AJ, Altman DG. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. BMJ 2010;340:c181–0.
18. U.S. National Institutes of Health. ClinicalTrials.gov. 2009; Available from: http://www.clinicaltrials.gov/ [accessed 2012 Feb 5].
19. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database–update and key issues. N Engl J Med 2011;364(9):852–60.
20. Institute of Medicine of the National Academies. Sharing clinical research data: a workshop. Available from: http://www.iom.edu/Activities/Research/SharingClinicalResearchData.aspx [accessed 2012 Oct 19].
21. Hrynaszkiewicz I, Altman DG. Towards agreement on best practice for publishing raw clinical trial data. Trials 2009;10:17.
22. Vickers AJ. Whose data set is it anyway? Sharing raw data from randomized trials. Trials 2006;7:15.

Author information

Joseph Ross is an assistant professor in the Section of General Internal Medicine at the Yale University School of Medicine. His research interests include evaluating the impact of state and federal policies on the delivery of appropriate and higher quality care and issues related to conflict of interest, evidence development, and drug safety. Dr Ross received his Master's degree in Health Sciences Research from the Yale University School of Medicine and an M.D. from the Albert Einstein College of Medicine.

Harlan Krumholz is a cardiologist and the Harold H. Hines, Jr Professor of Medicine and Epidemiology and Public Health at the Yale University School of Medicine, as well as the director of Center for Outcomes Research and Evaluation and of the Robert Wood Johnson Clinical Scholars Program at Yale. An international leader in the field of outcomes research, he serves on the Board of Governors of the recently established Patient-Centered Outcomes Research Institute. Krumholz holds a Master's degree in Health Policy and Management from the Harvard School of Public Health and a M.D. from Harvard Medical School.