ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

Annals of Oncology 28: 1734–1737, 2017
doi:10.1093/annonc/mdx123
Published online 5 April 2017

SPECIAL ARTICLE

Statistical controversies in clinical research: data access and sharing—can we be more transparent about clinical research? Let's do what's right for patients

F. W. Rockhold

Department of Biostatistics and Bioinformatics, Duke Clinical Research Institute, Duke University Medical Center, Durham, USA

Correspondence to: Prof. Frank W. Rockhold, Department of Biostatistics and Bioinformatics, Duke Clinical Research Institute, Duke University Medical Center, PO Box 17969, Durham, NC 27715, USA. Tel: +1-919-668-1073; E-mail: frank.rockhold@duke.edu

Calls for greater transparency and 'open data access' in clinical research are widespread, from sources including the Executive Office of the President, which in 2013 called for increased access to the results of federally funded research. In 2015, The Institute of Medicine issued a report advocating for a multi-stakeholder effort to foster responsible data sharing, and there are many others. Open science is good for researchers, good for innovation, and good for patients. The question at the center of the open-science efforts for clinical trials should not be *whether* data should be shared, but rather how we can usher in responsible methods for doing so. Unfortunately, there remain numerous perceived barriers to complete transparency around clinical trial data. This paper reviews the current status of data disclosure, the barriers to achieving it and a suggestion for the future.

Key words: open science, transparency, data sharing, data access

Introduction

Sharing patient-level data from clinical trials can improve the quality of research and our understanding of disease and medical treatments both of which can benefit patients. The imperative to share one's data would seem compelling, and the debate of how and when to do so has been active. This topic of data disclosure and transparency in clinical research has been the subject of many recent articles, editorials, policy statements, and indeed regulations over the past few years [1–11]. Of note in this list was the proposal by ICMJE in 2016 to require public disclosure of patient level data for articles published in their journals. Some have expressed concern about the logistics and cost (in particular for academic researchers), others resisted on the grounds that they were the best (only?) one to analyze 'their data'. Apparently, the sentiment that *Scientists would rather share their toothbrush than their data!* [12, 13] still seems to be alive and well in some quarters of the clinical research world [8].

The drive for data disclosure to increase 'transparency' in clinical research is of course not new. It has its roots in the drive to minimize publication bias and the 'file drawer problem'. Recent history started in 2004 with the ICMJE policy to require registration of trials before a manuscript would be accepted for publication [14]. As recently as 2005 the world was still debating the 20 elements to be included in such protocol registrations and then quickly moved on to results posting (FDAAA 2007) [15, 16] (update 2016 [17]). While we still have a way to go in making basic results and publications available [18, 19], the current discussions described in the previous paragraph around individual patient data are part of a natural evolution in disclosure and transparency and not a sudden awakening. It is important to distinguish as well between 'transparency' and 'disclosure or access'. Transparency is a desire and a need while access and disclosure or access is action. The latter are necessary for 'transparency' but not sufficient. Transparency is only achieved if researchers are actually able to utilize the information to further increase the knowledge base. While many of the above referenced authors have reviewed the ethical imperatives for broader data sharing, I highlight some here to provide context and balance for the issues. Access to more detailed patient information, enables further research to benefit

patient care. While we have access to peer reviewed publications and summary results on sites like ClinicalTrials.Gov, the broader release of clinical trial data enables the review of results from individual clinical trials to validate the results, help understand how and why trials might fail and avoid duplication of futile research (may avoid unnecessarily enrolling patients into clinical trials and exposing them to risks). It also strengthens trust in clinical research through enhanced openness and transparency, and honors the commitment to patients who volunteer for trials to fully utilize their data. It is difficult, and maybe impossible, to accomplish these tasks without access to the individual patient data. Finally, and perhaps of most current relevance, is if one believes in the value of 'precision medicine', the realization of that goal would indeed be impossible without access to individual patient data. For these reasons and the emergence of a plethora of data access systems, the question is not whether to provide better data access systems, but how to do so in a more efficient manner.

Issues with data access and sharing

A wide range of concerns have been voiced about individual patient data sharing related to patient privacy, consent, intellectual property, academic credit, costs, infrastructure, data standards, and potentially erroneous conclusions from 'rogue analyses'. The latter issue has to date not turned out to be as big an issue as once feared [20, 21]. Many of the other concerns cannot be totally eliminated, but they can be mitigated and managed. For example, the development of algorithms for de-identification of patient data is becoming more sophisticated to protect patient privacy [22]. The concept of intellectual capital has been a bit more difficult to define and remains an issue. Industry may be concerned that commercially sensitive information could be uncovered in those data, while academic investigators are concerned that future research, publications, and grant funding could be compromised by early release of data. While far from resolved, the impact of these issues is being ameliorated by the use of independent review panels. To address the valid concerns of academic researchers in releasing their data, the Institute of Medicine, journal editors, and others have called for clear citation of data generators work by researchers accessing data for secondary analyses to acknowledge the expertise and effort of the original researchers [4, 10].

Data sharing systems

While the benefits of secondary analyses of patient-level data may be hard to measure, the voices of support for data sharing seem to be outweighing the doubters. In spite of, or perhaps because of the debate, a number of major initiatives have been launched over the past 3–4 years to move the field of data sharing in the clinical research sector forward. Project Data Sphere [23, 24], CSDR [21, 25], Yale University Open Data Access (YODA) [26], Supporting Open Access for Researchers (SOAR) [27], to name a few. The last three are collections of individual patient data from trials receive from 13, 2, and 1 commercial sponsors, respectively.

The Oncology community has also been quite active in the area of data sharing and several platforms have evolved specific to oncology. Data sharing has also been put center stage by Vice President Joe Biden as part of his Cancer Moonshot program. Driven by the sense of urgency to get treatments to patients, these sites can be a source of data and information for researchers and care givers alike. The Cancer Data Access system [CDAS, https://biometry.nci.nih.gov/cdas/ (28 March 2017, date last accessed)] from NCI is one large effort. While there are also many oncology trials with patient level data listed in the platforms mentioned in the previous paragraph, one platform of particular interest to the oncology community is the Project Data Sphere Initiative which houses individual patient data from control groups of a selection of oncology trials primarily from industry [23, 24]. This has been valuable for study design and evaluation of baseline covariates and potential enrichment criteria. As it is focused on a single disease area it is not as large as some of other initiatives and is limited to only control group data. It does, however, stand out as the most 'open' of the systems available in that no intermediate review panel was required and the data were easily accessible from the data portal. In addition, there has been a more rapid availability of publications [28]. Another benefit of the system is the ability to combine studies across sponsors. The limitations of the system are the relatively small size the lack of data from the experimental arm. It is hoped that more trials will be added and perhaps at some point the experimental arm patient data as well as this would benefit the oncology community.

Difficulties with current data access systems

Interestingly, in spite of all this progress in response to the resounding cry for openness and availability of data, the jury is still out on the utility of the approaches being taken. A DCRI study [20] of the use of open-access platforms found that although 3000 trials were available to investigators, access to only 15.5% had been requested. The study examined three open access platforms, ClinicalStudyDataRequest.com, the YODA Project, and SOAR. Most proposals did not focus on validating the primary results of the trial, instead suggesting secondary uses such as epidemiological studies, subgroup analyses, analyses of the disease state, or predictors of treatment response. The lack of publications was also highlighted recently in another review by the CSDR Independent Review Panel [29]. The reasons for underutilization of data from open-access platforms may include lack of knowledge about the existence of these resources, lack of funding for analyses, or the length of time needed to prepare and submit publications. Difficulties in accessing data, lack of ability to download datasets, conducting analyses of data from trials that used different structures and standards for data and metadata may also be slowing the release of additional studies. The lack of visibility of data availability seems at odds with the drive for transparency and it is hoped that improved efficiency of the data access process will drive up usage.

Moving ahead

The clinical research community is at an important point in time with regard to data access and sharing. There are clearly a growing

number of people who believe that sharing data are the right thing to do and that we need to find the best ways to realize those benefits while minimizing the risks to patients and researchers. Instead of resisting this movement is many, researchers are now instead debating when and how data should be shared after publication of the primary manuscript. This is largely good news, but the unintended consequence of developing multiple approaches and systems will create a fragmented, complex, and confusing landscape in which the full benefits of data sharing will not be realized [5]. If we do not react this proliferation soon, we may be in the paradoxical position of disclosure *without* transparency. Thus, while we will have moved past denying access, the patients and researchers will not realize the benefits of truly accessible and transparent information and we will move from overtly hiding data to hiding it in plain sight.

One possible approach is for everyone to move to a common portal where sponsors or investigators send study details, data, or both to an independent custodian who manages scientific review, privacy, and other aspects [30]. This approach might require a number of existing systems, but it would realize economies of scale, helping to address cost barriers. Alternatively, the provision of data-sharing services for some sponsors could be combined with a federated model offering a central portal linking to other systems or portals, that may be independently set up within therapeutic or disease areas such as discussed above. This approach would allow researchers to more efficiently access data from multiple existing platforms. This such a portal would have to be more than simply a directory of systems. No matter which approach is ultimately developed, there would be a need for common data standards [31] and at least agreed minimum standards for data use agreements and individual patient level data de-identification.

The funding mechanism for such a system is an important consideration. I would point to government, regulators, health authorities, professional societies, to take the leadership and direction from the principles outlined in the IOM review previously mentioned. While such systems are not inexpensive, we must put the cost in perspective. The cost of a central access system would in all likelihood add up to the fraction of the cost of a *single* large confirmatory clinical trial amortized over the ability to access thousands of trials and certainly less expensive than setting up multiple individual systems.

The clinical research enterprise needs to come together to build on what currently exists and create simpler 'one-stop shop' platform(s) for clinical trial data sharing. If we get this system right, it could provide a basis for sharing other types of data, such as preclinical data and real-world epidemiologic data. If we allow inevitable differences in systems or processes to produce a fragmented, uncoordinated approach, we will have missed the opportunity to realize great value for patients.

Disclosure

Consulting/honoraria: Amgen, AbbVie, Adverum, Nabriva, Novo Nordisk, GlaxoSmithKline, California Institute for Regenerative Medicine, PCORI. Equity Interest: GlaxoSmithKline. The study was supported by Astra Zeneca, PCORI, NIH, Janssen and ReNeuron (no grant numbers apply).

References

1. Drazen JM. Data sharing and the journal. N Engl J Med 2016; 374: e24.
2. Horton R. Offline: data sharing—why editors may have got it wrong. Lancet 2016; 388: 1143.
3. Drazen JM, Morrissey S, Malina D et al. The importance—and the complexities—of data sharing. N Engl J Med 2016; 375: 1182–1183.
4. Lo B, DeMets DL. Incentives for clinical trialists to share data. N Engl J Med 2016; 375: 1112–1115.
5. Rockhold F, Nisen P, Freeman A. Data sharing at a crossroads. N Engl J Med 2016; 375: 1115–1117.
6. Longo DL, Drazen JM. Data sharing. N Engl J Med 2016; 374: 276–277.
7. Taichman DB, Backus J, Baethge C et al. Sharing clinical trial data. Br Med J 2016; 532: i255.
8. Toward fairness in data sharing. N Engl J Med 2016; 375: 405–407.
9. Academic Research Organization Consortium for Continuing Evaluation of Scientific S-C, Patel MR, Armstrong PW et al. Sharing data from cardiovascular clinical trials—a proposal. N Engl J Med 2016; 375: 407–409.
10. IOM. Discussion framework for clinical trial data sharing: guiding principles, elements, and activities. In Discussion Framework for Clinical Trial Data Sharing: Guiding Principles, Elements, and Activities. Washington (DC): National Academies Press (US) Copyright 2014 by the National Academy of Sciences. All rights reserved. 2014.
11. EMA. Policy 700 Data Disclosure. 2015.
12. Goble C, Me-Science' the New e-Science. Science Grid This Week 2006.
13. Hodson S, Data-sharing culture has changed. Research Information 2010.
14. De Angelis C, Drazen JM, Frizelle FA et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. N Engl J Med 2004; 351: 1250–1251.
15. Laine C, Horton R, DeAngelis C et al. Clinical trial registration: looking back and moving ahead. N Z Med J 2007; 120: U2586.
16. Bian ZX, Wu TX. Legislation for trial registration and data transparency. Trials 2010; 11: 64.
17. Zarin DA, Tse T, Williams RJ, Carr S. Trial reporting in ClinicalTrials.gov—the final rule. N Engl J Med 2016; 375: 1998–2004.
18. Zarin DA, Tse T, Williams RJ et al. The ClinicalTrials.gov results database–update and key issues. N Engl J Med 2011; 364: 852–860.
19. Ross JS, Tse T, Zarin DA et al. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. Br Med J 2012; 344: d7292.
20. Navar AM, Pencina MJ, Rymer JA et al. Use of open access platforms for clinical trial data. JAMA 2016; 315: 1283–1284.
21. Strom BL, Buyse M, Hughes J, Knoppers BM. Data sharing, year 1–access to data from industry-sponsored clinical trials. N Engl J Med 2014; 371: 2052–2054.
22. Hughes S, Wells K, McSorley P, Freeman A. Preparing individual patient data from clinical trials for sharing: the GlaxoSmithKline approach. Pharmaceut Statist 2014; 13: 179–183.
23. Green AK, Reeder-Hayes KE, Corty RW et al. The project data sphere initiative: accelerating cancer research by sharing data. Oncologist 2015; 20: U464–U424.
24. Abdallah K, Jenkins R, Hugh-Jones C et al. Project data sphere initiative: results from first 6 months of broad access to oncology trials data. J Clin Oncol 2015; 33.
25. Nisen P, Rockhold F. Access to patient-level data from GlaxoSmithKline clinical trials. N Engl J Med 2013; 369: 475–478.
26. Krumholz HM, Waldstreicher J. The Yale Open Data Access (YODA) Project—a mechanism for data sharing. N Engl J Med 2016; 375: 403–405.

27. Pencina MJ, Louzao DM, McCourt BJ et al. Supporting open access to clinical trial data for researchers: the Duke Clinical Research Institute–Bristol-Myers Squibb Supporting Open Access to Researchers Initiative. Am Heart J 2016; 172: 64–69.
28. Halabi S, Rini B, Escudier B et al. Progression-free survival as a surrogate endpoint of overall survival in patients with metastatic renal cell carcinoma. Cancer 2014; 120: 52–60.
29. Strom BL, Buyse ME, Hughes J, Knoppers BM. Data sharing - is the juice worth the squeeze? N Engl J Med 2016; 375: 1608–1609.
30. Bierer BE, Li R, Barnes M, Sim I. A global, neutral platform for sharing trial data. N Engl J Med 2016; 374: 2411–2413.
31. Merson L, Gaye O, Guerin PJ. Avoiding data dumpsters–toward equitable and useful data sharing. N Engl J Med 2016; 374: 2414–2415.