

## Open data are not enough to realize full transparency

Lex M. Bouter\*

*Department of Epidemiology and Biostatistics, VU University Medical Center, P.O. Box 7057, 1007MB Amsterdam, The Netherlands*

Accepted 6 May 2015; Published online 8 July 2015

The plea by Robert West to invite authors of clinical and behavioral studies to publish their data sets and command files is clearly important in context of the prevention of research waste [1,2]. I fully agree to his proposal, but I also firmly believe we need to go substantially further. West focuses on voluntary transparency regarding the data and the analyses underlying the article at issue. He provides three reasons why this is important: to protect against fraud and misrepresentation, to reduce the error rate, and to facilitate additional analysis. I will argue that the need for transparency is much broader. Subsequently, I shall comment on the three reasons given by West. Finally, I will propose two potentially effective measures to increase transparency.

Science is based on trust. Society must be able to trust scientists, and scientists should have good reasons to trust their colleagues [3]. To deserve trust, clinical research needs to be open, honest, and transparent. The record should be complete and verifiable. Besides being the basis for trust, that also will serve as a powerful antidote against selective reporting. Nonpublication and selective publication of study outcomes may be the single most important source of research waste [4–6]. It is also the Achilles heel of systematic reviews because these rely on the published reports of research projects. There is evidence that selective reporting increasingly leads to an overrepresentation of positive significant findings in the scientific literature [7,8]. Furthermore, selective reporting is unethical in the sense that the efforts of patients participating in the study are wasted. Transparency concerns the whole trajectory: study protocol, the process of data collection, data sets, data analysis, report of findings, amendments made underway, financial and intellectual conflicts of interest, and so forth [9,10]. The ideal is to make all this information prospectively and publicly available. The proposal by West to publish the data and the syntax together with the article at issue offers only limited transparency and will not help a lot in the prevention of selective reporting. Without a study protocol that was made publicly available before the start of

the data collection, it is very hard to judge whether all planned research questions are answered in the published report. Equally, a data analysis plan that was publicly deposited before the data were collected is necessary to judge whether the statistical analysis was not partly data driven.

I agree that publishing data and syntaxes may serve in the identification of errors and misrepresentation. However, it will not do much for the identification of fraud as the data published may still be fabricated or manipulated. It enables replication of the data-analyses done by the authors of the publication at issue and also provides an opportunity to explore alternative approaches with different cutoff points, categorizations, or statistical techniques. This certainly is useful for establishing the robustness of the published findings [11,12]. And if the published data set contains more than what the authors used for their report, it can also help in identifying instances of selective publication. Please note that replication of the data analysis is only one of the forms replication can take. Other perhaps more important forms of replication are the collection of new data with the same study protocol and attempts to answer the same research questions with another study design and/or in another setting. Replication by collecting new data is indicated when the aggregated data from available studies are insufficient to answer the research question at issue with adequate validity and precision. If there is already enough data, the collection of new data is unethical and a waste of resources.

West makes a distinction between data disclosure and data sharing. He argues that others have a right to look for flaws in the data analysis and to publish them when found. But, he says that the intellectual property rights should be respected, which means that colleagues will need permission to use the data to answer other research questions. I respectfully disagree. I firmly believe that data collected among volunteering participants of clinical research belong to the public domain. Of course, some months of embargo can be reasonable, proper acknowledgments should be made, and maybe the original investigators should be offered the opportunity to participate in the secondary analyses. In addition, I agree with West that

\* Corresponding author.

E-mail address: [lm.bouter@vu.nl](mailto:lm.bouter@vu.nl)

published data sets need to contain all relevant information and also that breaches of privacy and misuse of the data ought to be prevented. And it is obvious that for secondary analyses, the same rules for transparency apply, starting with a predefined study protocol. However, all these do not detract from the principle that data from clinical research belong in the open domain.

One may wonder how transparency can be promoted best. Next to good education on responsible conduct of research on all levels in Academia, there are two approaches I find promising. First, we should look critically at the current reward systems and consider alternatives. Scientists gain prestige and get tenure by collecting as much publications, citations, and grants as possible. Having spectacular and statistically significant results helps them a lot. Current reward systems do neither focus on replication and nor on sharing data. In addition, rewards for publishing study protocols and negative results are nonexistent. Recently, Ioannidis and Khoury [13] proposed an interesting and more balanced alternative to remedy some of these perverse incentives.

Second, transparency could be enforced by a concerted action of granting agencies, institutional review boards, and scientific journals [14]. Demanding a timely public deposition of study protocol, syntax and outcome reports as a condition for the last payment, for permission to perform the study and for accepting the article for publication, respectively, would obviously be strong incentives to behave transparent. In the field of randomized clinical trials, we have seen some progress in that sense during the last 2 decades. However, there is still a lot room for improvement, and other types of studies are lagging behind [15–18]. Especially, the impact of demands for transparency by funding agencies may be substantial [19].

We clearly need to collect some more evidence on how transparency can be realized best. And—as Robert West also mentions—we need to look into potential drawbacks and undesired side effects of the interventions proposed. Including exploring methods to implement transparency procedures on the Web sites of journals, funding agencies, or other organizations. Especially, feasible ways of monitoring the compliance with the rules for transparency need to be developed. Consequently, it makes sense to first experiment on a voluntary basis, with a view to move on

to compulsory measures once we understand better how to nudge and force clinical research in a direction of minimal waste and maximum transparency.

## References

- [1] West R. Promoting greater transparency and accountability in clinical and behavioural research by routinely disclosing data and statistical commands. *J Clin Epidemiol* 2015. [E-pub ahead of print].
- [2] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 2009;374:86–9.
- [3] Resnik DB. Scientific research and the public trust. *Sci Eng Ethics* 2011;17:399–409.
- [4] Bouter LM. Perverse incentives and rotten apples. *Account Res* 2015; 22:148–61.
- [5] 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:e3081.
- [6] Knottnerus JA, Tugwell P. Selection-related bias, an ongoing concern in doing and publishing research. *J Clin Epidemiol* 2014;67: 1057–8.
- [7] van Assen MA, van Aert RC, Nuijten MB, Wicherts JM. Why publishing everything is more effective than selective publishing of statistically significant results. *PLoS one* 2014;9:e84896.
- [8] Fanelli D. Negative results are disappearing from most disciplines and countries. *Scientometrics* 2012;90:891–904.
- [9] Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gøtzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet* 2014;383:257–66.
- [10] Bero L. Nonfinancial influences on the outcomes of systematic reviews and guidelines. *J Clin Epidemiol* 2014;67:1239–41.
- [11] Krumholz HM, Peterson ED. Open access to clinical trial data. *JAMA* 2014;312:1002–3.
- [12] Ebrahim S, Sohani ZN, Montoya L, Agarwal A, Thorlund K, Mills EJ, et al. Reanalyses of randomized clinical trial data. *JAMA* 2014;312:1024–32.
- [13] Ioannidis JP, Khoury MJ. Assessing value in biomedical research: the PQRST of appraisal and reward. *JAMA* 2014;312:483–4.
- [14] Ter Riet G, Bouter LM. How to end selective reporting in animal research. In: animal models in research and development of cancer therapy. (In press).
- [15] Goldacre B. Are clinical trial data shared sufficiently today? *BMJ* 2013;347:f1880.
- [16] Chalmers I, Glasziou P, Godlee F. All trials must be registered and the results published. *BMJ* 2013;346:f105.
- [17] Hudson K. Sharing results of RCTs. *JAMA Published Online* 2014; E1–2.
- [18] Swaen GM, Carmichael N, Doe J. Strengthening the reliability and credibility of observational epidemiology studies by creating an Observational Studies Register. *J Clin Epidemiol* 2011;64:481–6.
- [19] Chinnery F, Young A, Goodman J, Ashton-Key M, Milne R. Time to publication for NIHR HTA programme-funded research: a cohort study. *BMJ Open* 2013;3:e004121.