The poor availability of research data for reanalysis has just become much poorer

Jelte M. Wicherts & Marjan Bakker

University of Amsterdam.

Correspondence concerning this manuscript should be addressed to Jelte M. Wicherts, Department of Psychology, Psychological Methods, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands, telephone: +31 205257067, fax: +31 206390026, J.M.Wicherts@uva.nl.

The preparation of this article was supported by VENI grant no. 451-07-016 from the Netherlands Organization for Scientific Research (NWO).

The newly revised, 6th Edition of the Publication Manual of the American Psychological Association (APA, 2010) is currently (September 1st, 2009) the best-selling book in all categories of books sold by Amazom.com. Its previous version was published in 2001 and has remained in the Top 100 of Amazon bestsellers for the last two years. Given its selling record, it is safe to say that the Publication Manual ranks among the most important books in the whole of the behavioral sciences. The Publication Manual not only specifies how manuscripts to be submitted to peer-reviewed journals should be structured and written, but also deals with matters of ethics surrounding publication of research. For instance, the Publication Manual refers to Standard 8.12 of the Ethics Code of the APA, which states that after research results are published, the authors have an obligation to share their data with other researchers who wish to verify the conclusions through reanalysis of the data. Although this requirement was already explicated in previous editions of the Publication Manual, and despite the fact that the APA requires authors undersign a statement to that effect before it publishes their work, we (Wicherts, Borsboom, Kats, & Molenaar, 2006) found that 73% of researchers who had published papers in high-impact APA journals did not comply with this guideline and refused to share their data for reanalysis. Our results suggest that approximately three quarters of the data published in APA journals are not open to verification by other researchers. Yet the analyses as presented in the publication often do not represent the final word on the empirical findings. As anyone who has ever analyzed psychological research data knows, data often lend themselves to much more than one type of statistical analysis. Data often do not solely fit the preconceived ideas and cherished conclusions of researchers, but may look different when reanalyzed by others (Firebaugh, 2007). Additional reasons to make data available for reanalysis are that statistical analyses are characterized by disturbingly large error rates (Berle & Starcevic, 2007; Rossi, 1987; Wolins, 1962), and that raw data are often needed for inclusion in meta-analyses.

In 2006, we proposed that the APA should take steps towards improving data availability, for instance by a general requirement to publish research data alongside papers published by the APA (e.g., as supplementary online material). *Nature* (2006) followed with an editorial that mostly subscribed to our views. Indeed, at the time the APA indicated to us that it was in the process of revising the Publication Manual and was considering ways to improve compliance with its Ethical Standard 8.12. Needless to say, we were excited when the Amazon box with the new Publication Manual fell on our doormat.

Unfortunately, we came to realize quickly that the APA has not only missed an opportunity to amend the Publication Manual to promote data availability for reanalysis, but has actually made it even harder to access data. According to the 6th Edition of the Publication Manual, research data should shared with the object of re-analyses only following a written agreement between the original researcher and the requesting researcher on "the conditions under which the data are to be shared" (APA, 2010, p. 12). These conditions include (1) a careful consideration of "copyright restrictions, consent provided by subjects, requirements of funding agencies, and rules promulgated by the employer or of the holder of the data"; (2) a priori explication of the specific use of the data; "e.g., verification of already published results, for inclusion in meta-analytic studies, for secondary analysis"; (3) a formal agreement about the limits on the distribution of the shared data; (4) a specification of the limits on the dissemination of the results of the re-analysis (e.g., conference abstracts, internal reports, journal articles, book chapters, etc.); and (5) a specification of the authorship expectations.

These conditions, especially points 4 and 5, appear to have been inspired by a scenario in which the requester wants to write a new paper on the existing data with a secondary analysis aimed at a new research hypothesis. It is reasonable that in this case the original author has the opportunity to lay down conditions on potential co-authorship and on

how and where the results are published. However, these newly explicated conditions hinder the sharing of research data for reanalysis if the requester is critical towards the original researcher's analyses and presentation of results.

Suppose, for instance, that Dr. Y reads a paper by Dr. X and submises that the statistical analyses are incorrect. He wants to check whether the conclusions drawn by Dr. X still stand using the appropriate statistical analysis. Dr. Y writes a polite email to Dr. X including passages like "With much interest I have read your article on…", "your analyses may however be suboptimal", "I would like to verify the robustness of your conclusions against the use of an alternative statistical method", and "I promise not to share the data with anyone else and will use the data only for this stated purpose". One the basis of our experiences, we would estimate the chances of Dr X. responding positively to this email below 30%. Now, in the light of the guidelines in the new Publication Manual, Dr. X may respond to the request by writing: "Thank you for your interest in our work. Of course, I would be happy to share with you these data, but only if I become a co-author on a publication in a peer-reviewed journal. I can't waste my time with some internal report which I cannot put on my resume. Cheers, Dr. X". Dr. X can do so with impunity because his response is in line with the guidelines in the 6th Edition of the Publication Manual. Dr. X. could even stipulate in the written agreement that the results of any re-analysis cannot be published at all (Point 3), that the reanalysis should not be different from the ones conducted in the original paper (Point 2), or that the results may only be published with Dr. X as a co-author (Point 5). The present guidelines are such that there is every chance that scientists will not come to a written agreement about data sharing, especially when the re-analyses are motivated by possible errors in the original publication. Disagreements on data analyses and errors therein were the prime reasons for proposing guidelines with respect to data availability. The written agreement as stipulated in the new Publication Manual tilts the

balance of power on how the data should be analyzed strongly towards the original researcher and away from those who wish to replicate the result with alternative methods of analysis. In practice, these conditions may make it virtually impossible for a critical researcher like Dr. Y to reanalyze data published in an APA journal if he deems the original analyses to be questionable in any way. Scientists like Dr. X may not wish to co-author a paper that reveals possible errors in his own work.

We ourselves have experiences like Dr. Y's in the field of ethnic differences in IQ and are familiar with similar experiences of many others in other fields of study. As it stands, the APA has provided primary authors with more power *not* to share their data for reanalysis. We had already demonstrated that the majority of researchers are unconcerned about violating this basic ethical principle of science. If anything the current APA guidelines have made matters worse. We urge the APA to take real steps to open up the obscurity with which data analyses are conducted in psychology.

References

American Psychological Association. (2010). *Publication Manual of the American Psychological Association. Sixth Edition*. Washington, DC: American Psychological Association.

Berle, D., & Starcevic, V. (2007). Inconsistencies between reported test statistics and p-values in two psychiatry journals. *International Journal of Methods in Psychiatric Research, 16*, 202-207.

Firebaugh, G. (2007). Replication data sets and favored-hypothesis bias. *Sociological Methods & Research, 36*, 200-209.

Nature. (2006). A fair share. Editorial. *Nature, 444*, 653-654.

Rossi, J. S. (1987). How Often Are Our Statistics Wrong - a Statistics Class Exercise. *Teaching of Psychology, 14*, 98-101.

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist, 61*, 726-728.

Wolins, L. (1962). Responsibility for raw data. *American Psychologist, 17*, 657-658.