

17

Why We Could Stop Worrying About Gaming Metrics If We Stopped Using Journal Articles for Publishing Scientific Research

Elizabeth Wager

The problem of “gaming” metrics arises because journal articles are used both to report research and to measure the productivity of researchers. Since publications in peer-reviewed journals are viewed as the “currency” of academia, we have created an incentive for researchers to cheat. This cheating ranges from questionable practices such as “salami-sliced” publications, redundant publications, and gift authorship to serious misconduct such as data fabrication and plagiarism. Pressure to publish in “high-impact” journals also creates incentives for data falsification and misleading reporting, since these highly selective journals seek unusual and interesting findings. (This may be one reason why retraction rates have been shown to correlate with Journal Impact Factors [Fang and Casadevall, 2011].)

One radical solution to the problem of such metrics gaming is to stop publishing scientific research in the form of traditional journal articles and to develop new methods both for disseminating findings and for measuring research productivity. For research that generates numeric or digital data, abandoning journal articles as the medium for disseminating primary results would have several benefits. This chapter uses the reporting of clinical trials to consider and exemplify the problems with traditional journal articles and the benefits of abandoning them in favor of alternative methods of dissemination.

A paradox of medical publication is that, while researchers report being under intense pressure to publish, nevertheless many clinical trials are never published. This causes the medical literature to be seriously skewed. Trials that fail to produce a statistically significant result, or that give results that are disappointing to the sponsor or investigator, are the most likely not to be published. This causes serious publication bias, making the published evidence base unreliable. Estimates of the extent of nonpublication vary,

but several studies have suggested that as many as half of all clinical trials are not published (Song et al., 2010; Chan et al., 2014).

Reasons for nonpublication vary but the difficulties of preparing a journal article and then of getting it accepted in a “good” journal undoubtedly contribute (Smyth et al., 2011). Although there are now several journals that explicitly welcome reports of small studies, replication experiments, and nonsignificant findings, academics either continue to believe that such work is unlikely to be accepted, or fear that publishing it may harm their reputation or career. Pressure from sponsors to suppress inconvenient findings and lack of incentives to publish all trials are probably also factors. Current systems of measuring academic productivity focus almost exclusively on the publication of journal articles, often with an emphasis on high-impact journals. Researchers may therefore be incentivized to abandon disappointing avenues of research, leaving them unpublished, to focus on new work, which they hope will be more successful. Competition between researchers may also create disincentives for reporting failed methods.

Another problem with traditional publication models is that many trials are only partially published. While a three-thousand-word article may be convenient for many readers, the space constraints imposed mean that many details of clinical trials are not included in journal articles. Deficiencies in describing both the methods and the results have been clearly documented and are worryingly prevalent (Hoffmann et al., 2014). Missing details about the methods prevent findings from being replicated, trial quality from being properly assessed, and new techniques from being implemented. Partial reporting of results may contribute to publication bias since statistically significant or “positive” findings are more likely to be included than inconclusive or inconvenient ones (Chan et al., 2004). Incomplete reporting (whether deliberate or unintentional) can rarely be detected from journal articles alone and cannot be spotted during peer review unless reviewers have access to, and are prepared to carefully check, the protocol.

The lack of linkage between study protocols, underlying data, and research reports also creates the possibility for other problems. Peer review of a journal article is often done without reference to the study protocol and data analysis plan, which describe the original design for the research and are therefore essential for interpreting the findings. Comparisons of protocols and journal articles have shown that a worryingly high proportion of published articles (around sixty percent according to some studies) do not report the primary outcome specified in the

original study design, or switch primary and secondary outcomes (Chan et al., 2004). This undermines the reliability of the analysis and may reduce the statistical power of the analysis. Linking documents such as protocols, investigator brochures, patient information leaflets, and the various presentations of findings, including clinical trial reports prepared for regulators, conference presentations, results postings on trial registers, and descriptive articles, would not only be more efficient, but also produce a more reliable account of the research (Glasziou et al., 2014). Such linkage is technically possible and may be partially automated, but could be achieved more effectively if supported by funders and regulators (Goldacre and Gray, 2016).

Another weakness of traditional journal articles, and the way in which they are peer reviewed, is that this system provides little or no safeguards against incorrect statistical analysis. Such problems have been well documented in published articles even in the top-ranking journals and may occur deliberately, to emphasize or mask effects, or inadvertently through researcher ignorance (Altman, 2002). Inappropriate handling of missing data or outliers can also affect outcomes. Unless reviewers have access to the raw data, and sufficient statistical expertise, this is usually not detected by peer review. The traditional journal article, which shows only the analyzed aggregate data, rarely reveals such problems.

While inappropriate statistical methods appear to be relatively common, deliberate research fraud is probably much rarer, but not so rare that it can safely be ignored. However, sophisticated data falsification or fabrication is very hard to detect from traditional publications (Carlisle, 2012). Reviewers and readers see only the final results of analyses and have no access to the underlying data. Falsification of digital images may be detected if journals have access to original images and screen them, but this is time consuming (Linkert, 2010). Therefore, image manipulation is often detected only after publication by readers. Several cases of fraudulent image manipulation leading to retractions have been identified in this way (e.g., via alerts on PubPeer or directly to the journal).

One explanation given for the longevity of the journal article, which has changed little in 350 years in terms of length or format (Wager, 2006), is that readers like them. However, research articles are read for a variety of purposes, and one size does not necessarily suit all users (Altman, 2015). Many readers prefer a short summary, and journals have responded by including abstracts or even shortening the format (e.g., the *BMJ*'s "Pico" format [Jain, 2014]). However, other readers may seek more detail in the methods, so that they can replicate the findings or use the technique.

Articles are also used to create systematic reviews, which may combine findings using the statistical techniques of meta-analysis. This is often hampered by deficiencies in journal articles, so those compiling the review are forced to contact the authors to seek further information, which is not always forthcoming (Wager, 2006). Also, despite the familiarity with the journal article format, and even after technical editing, journal articles generally remain hard to read (Wager and Middleton, 2002).

Despite the predominance of the journal article, research is usually reported in several formats, such as a conference abstract, journal article, and press release and, increasingly, on trial registers. Other reports may also be prepared but shared only with regulators. Relevant information about the research may also be contained in the trial protocol and other documents such as the investigator brochure and participant information (Chalmers and Altman, 1999). Not only is the production of these multiple formats inefficient, but because versions are not linked, they may be inconsistent (Francis et al., 2013; Glasziou et al., 2014). In such cases, it may be unclear which is the correct version. Discrepancies may be due to simple errors or more complex factors such as different methods of data handling and analysis required by regulatory authorities and journal peer reviewers. The reporting of adverse events is especially prone to this (e.g., regulatory reports may include all adverse events while journal articles may report only those that were considered likely to have been caused by the treatment; classification of adverse events by severity may also vary) (Hughes et al., 2014). Once again, alternative publication models, especially those allowing linkage to the underlying data, might reduce these problems.

Conclusions

Familiarity with the conventional format of journal articles may lull users into a false sense of assurance. Yet, despite its familiarity, this format has many deficiencies as mentioned above and noted by several previous commentators (Smith, 1992; Chalmers and Altman, 1999; Smith and Roberts, 2006; Altman, 2014; Tracz, 2015). Furthermore, because journal articles are the “currency” of academia, and highly profitable for publishers, there is built-in resistance to change. The barriers to radically rethinking the way in which we disseminate research results are more social and cultural than technological.

The ideal system for reporting research would link the underlying data, appropriately labeled, with the full methods in the protocol and the entry on the trial register (Chalmers and Altman, 1999; Glasziou et al.,

2014). Text elements should probably be highly structured and machine readable (Altman, 2015). This would ensure completeness and consistency. The use of structured fields and online reports would mean that many checking functions could be automated (Wager, 2006). Structuring and linkage of information should reduce the burden of producing different formats and increase consistency. The resulting “report” might appear dull (at least to human readers), but that seems a small price to pay for accuracy. Incentives should be shifted away from publishing articles in journals toward systems that ensure that all research is publicly posted (for example, by funders withholding the final part of grants until this has been done). This radically different method of reporting research would also bring opportunities for new quality control mechanisms that would, most likely, replace traditional prepublication peer review, which has also been shown to have serious shortcomings and is not well suited to reviewing datasets (Wager and Jefferson, 2001; Tracz, 2015).

If scholarly journals were no longer the medium for publishing primary research findings, journal articles would cease to be the “currency” by which research output was measured. This would give an opportunity to develop better metrics. There are already proposals of new systems to give credit for data “authorship” (Bierer et al., 2017) and best practice for publishing raw clinical trial data (Hrynaszkiewicz and Altman, 2009). While structured reports and raw data might be highly efficient methods for disseminating research findings, there would still be a role for commentary, interpretation, and synthesis, all of which could be provided by scholarly journals.

Viewed in this way, the deficiencies of the current system for publishing research seem enormous, and the potential benefits from radical change seem obvious. Nevertheless, change has been technically possible and clearly proposed for almost twenty years. At the moment, it remains unclear whether dissatisfaction with journal articles might kill the current system of metrics, or whether concerns about gaming and unfair metrics could kill the journal article.

References

- Altman, Douglas G. 2002. “Poor-Quality Medical Research. What Can Journals Do?” *JAMA* 287:2765–2767.
- Altman, Douglas G. 2015. “Making Research Articles Fit for Purpose: Structured Reporting of Key Methods and Findings.” *Trials* 16:53.
- Bierer, Barbara E., Mercè Crosas, and Heather H. Pierce. 2017. “Data Authorship as an Incentive to Data Sharing.” *New England Journal of Medicine* 376:1684–1687.

Carlisle, John. 2012. "The Analysis of 168 Randomised Controlled Trials to Test Data Integrity." *Anaesthesia* 67:521–537.

Chalmers Ian, and Douglas G. Altman. 1999. "How Can Medical Journals Help Prevent Poor Medical Research? Some Opportunities Presented by Electronic Publishing." *Lancet* 353:490–493.

Chan, An-Wen, Asbjørn Hróbjartsson, Mette T. Haahr, Peter C. Gøtzsche, and Douglas G. Altman. 2004. "Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials: Comparison of Protocols to Published Articles." *JAMA* 291:2457–2465.

Chan, An-Wen, Fujian Song, Andrew Vickers, Tom Jefferson, Kay Dickersin, Peter C. Gøtzsche, Harlan M. Krumholz, Davina Gherzi, and H. Bart van der Worp. 2014. "Increasing Value and Reducing Waste: Addressing Inaccessible Research." *The Lancet* 383:257–266.

Fang, Ferric C., and Arturo Casadevall. 2011. "Retracted Science and the Retraction Index." *Infection and Immunity* 79:3855–3859.

Francis, Darrel P., Michael Mielewicz, David Zargarán, and Graham D. Cole. 2013. "Autologous Bone Marrow-Derived Stem Cell Therapy in Heart Disease: Discrepancies and Contradictions." *International Journal of Cardiology* 168:3381–3403.

Goldacre, Ben, and Jonathan Gray. "OpenTrials: Towards a Collaborative Open Database of All Available Information on All Clinical Trials." *Trials* 17:164.

Glasziou, Paul, Douglas G. Altman, Patrick Bossuyt, Isabelle Boutron, Mike Clarke, Steven Julious, Susan Michie, David Moher, and Elizabeth Wager. 2014. "Reducing Waste from Incomplete or Unusable Reports of Biomedical Research." *The Lancet* 383:267–276.

Hoffmann, Tammy C., Chrissy Erueti, and Paul P. Glasziou. 2013. "Poor Description of Non-Pharmacological Interventions: Analysis of Consecutive Sample of Randomised Trials." *BMJ* 347:f3755.

Hrynaskiewicz, Iain, and Douglas G. Altman. 2009. "Towards Agreement on Best Practice for Publishing Raw Clinical Trial Data." *Trials* 10:17.

Hughes, Shannon, David Cohen, and Rachel Jaggi. 2014. "Differences in Reporting Serious Adverse Events in Industry Sponsored Clinical Trial Registries and Journal Articles on Antidepressant and Antipsychotic Drugs: A Cross-Sectional Study." *BMJ Open* 4:e005535.

Jain, Anita. 2014. "Evolving Trends in Research Communication in a Digital Age: Examples from the BMJ." *Keio Journal of Medicine* 63:67–68.

Linkert, M., C. T. Rueden, C. Allan, J. M. Burel, W. Moore, A. Patterson, B. Lorange, J. Moore, C. Neves, D. Macdonald, A. Tarkowska, C. Sticco, E. Hill, M. Rossner, K. W. Eliceiri, and J. R. Swedlow. 2010. "Metadata Matters: Access to Image Data in the Real World." *Journal of Cell Biology* 189:777–782.

Smith, Richard. 1992. "The End of Scientific Journals?" *BMJ* 28:304.

Smith, Richard, and Ian Roberts. 2006. "Patient Safety Requires a New Way to Publish Clinical Trials." *PLoS Clinical Trials* 2006;1:e6.

Smyth, R. M. D., J. J. Kirkham, A. Jacoby, D. G. Altman, C. Gamble, and P. R. Williamson. 2011. "Frequency and Reasons for Outcome Reporting Bias in Clinical Trials: Interviews with Trialists." *BMJ* 342:c7153.

Song, Fujian, S. Parekh, L. Hooper, Y. K. Loke, J. Ryder, A. J. Sutton, C. Hing, C. S. Kwok, C. Pang, and I. Harvey. 2010. "Dissemination and Publication of Research Findings: An Updated Review of Related Biases." *Health Technology Assessment* 2010; Vol. 14: No. 8. http://www.journalslibrary.nihr.ac.uk/_data/assets/pdf_file/0005/64751/FullReport-hta14080.pdf. Accessed February 29, 2016.

Tracz, Vitek. 2015. "The Five Deadly Sins of Science Publishing." *F1000 Research* 4:112.

Wager, Elizabeth. 2006. "Publishing Clinical Trial Results: The Future Beckons." *PLoS Clinical Trials* 1:e31.

Wager, Elizabeth, and Philippa Middleton. 2002. "Effects of Technical Editing in Biomedical Journals: A Systematic Review." *JAMA* 287:2821–2824.

Wager, Elizabeth, and Tom Jefferson. 2001. "Shortcomings of Peer Review in Biomedical Journals." *Learned Publishing* 14:257–263.

This is a section of [doi:10.7551/mitpress/11087.001.0001](https://doi.org/10.7551/mitpress/11087.001.0001)

Gaming the Metrics

Misconduct and Manipulation in Academic Research

Edited by: Mario Biagioli, Alexandra Lippman

Citation:

Gaming the Metrics: Misconduct and Manipulation in Academic Research

Edited by: Mario Biagioli, Alexandra Lippman

DOI: 10.7551/mitpress/11087.001.0001

ISBN (electronic): 9780262356565

Publisher: The MIT Press

Published: 2020

This title is freely available as an open access edition thanks to the TOME initiative and the generous support of the University of California, Davis. Learn more at openmonographs.org



The MIT Press

© 2020 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC BY-NC-ND license.



Subject to such license, all rights are reserved.

This title is freely available as an open access edition thanks to the TOME initiative and the generous support of the University of California, Davis. Learn more at openmonographs.org.

This book was set in Sabon by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Biagioli, Mario, 1946- editor. | Lippman, Alexandra, editor.

Title: Gaming the metrics : misconduct and manipulation in academic research / edited by Mario Biagioli and Alexandra Lippman.

Description: Cambridge, MA : MIT Press, [2020] | Series: Infrastructures | Includes bibliographical references and index.

Identifiers: LCCN 2019010150 | ISBN 9780262537933 (pbk. : alk. paper)

Subjects: LCSH: Scholarly publishing—Corrupt practices. | Learning and scholarship—Corrupt practices. | Research—Corrupt practices. |

Communication in learning and scholarship—Moral and ethical aspects.

Classification: LCC Z286.S37 G36 2020 | DDC 070.5—dc23

LC record available at <https://lccn.loc.gov/2019010150>

10 9 8 7 6 5 4 3 2 1