



Replication and Reproducibility and the Self-Correction of Science: What Can *JID Innovations* Do?

In common life, to retract an error even in the beginning, is no easy task...

but in a public station, to have been in an error, and to have persisted in it,

when it is detected, ruins both reputation and fortune.

To this we may add, that disappointment and opposition inflame the minds of men, and attach them, still more, to their mistakes.

Alexander Hamilton

“Let science guide us.” Over the last few years, this simple phrase has been utilized across our society to advocate for a wide range of activities. This statement assumes that the scientific findings that are used to guide future scientific work and public policy are correct. Our confidence and indeed public confidence in scientific findings are based on our expectation that science is self-correcting. This self-correcting capability has traditionally been based on the value of high-quality peer review and the ability to reproduce and replicate scientific observations. However, the concept of self-correction has been challenged over the last two decades or more (Ioannidis, 2012). Replication of previous research relies on new studies that confirm previous results. Unfortunately, studies that just replicate previous observations are not often undertaken, and when done, they are difficult to publish. The lack of published studies that confirm previous observations is a result of the strong preference of funding agencies, tenure committees, and journal editors for new or innovative work (Ioannidis, 2012). This devaluing of confirmatory studies is similar to the fate of contradictory studies. Errington et al. (2021a, 2021b, 2021c, 2014) have sought to assess how often pre-clinical studies in cancer biology could be replicated (<https://validation.scienceexchange.com/#/projects>). In this study, they selected 193 experiments from 53 high-impact papers published from 2010 through 2012 (Errington et al., 2021a, 2021b, 2021c). Errington et al. (2021a, 2021b, 2021c) reviewed the papers; solicited necessary reagents from authors; and when needed, asked authors for assistance in performing experiments. They were only able to repeat 50 of 193 experiments from only 23 of the 53 papers. One barrier to replicating the experiments was the lack of the data needed to calculate effect sizes and to conduct power analysis in the original publication. Also concerning was that even after contacting authors for this information, they only received the

data for 68% of the experiments. None of the 193 experiments had sufficient detail in the papers to allow replication of the experiments, and 41% of authors did not respond or were minimally helpful when details were requested. Finally, when protocols for the experiments were obtained, 67% required significant modification to be completed. Similar findings were reported by Begley and Ellis (2012), who reported on their attempts to replicate the results of 53 landmark pre-clinical research papers in oncology. They reported that the scientific findings of these papers could be confirmed in only 11% of the papers. They also noted that these papers, whose results could not be confirmed, had already been highly cited in the literature (Begley and Ellis, 2012). The combination of the lack of interest of the scientific community in performing and publishing confirmatory studies and the observation that many studies cannot be replicated seriously threatens the concept of self-correcting science. The inability to replicate studies is a complex problem, and multiple issues come into play. The lack of published details about complicated experimental methodology and/or the failure to recognize and/or control important variables in an experiment could result in the failure to be able to confirm the findings of previous studies. Regardless of the etiology of the problem, published studies that cannot be replicated are still cited and used to support future studies and, as a result, negatively impact the efficient progress of scientific discovery and threaten the self-correcting nature of science.

An additional issue that is important in assessing the validity of a study is the reproducibility of its findings. Can another scientist analyze the data presented in a paper and reach the same conclusions (Allison et al., 2016)? A critical issue in reproducing the findings of a paper is the availability of the primary data. Recent developments have led to an increasing trend to make all primary data publicly available. As of January 2023, all National Institutes of Health (NIH)–sponsored research will be required to provide a data-sharing plan, with the goal of public availability of high-quality data sets (<https://sharing.nih.gov/>). Increasingly, journals are requiring that primary data be placed in a public repository (<https://journals.plos.org/plosone/s/data-availability>). However, it is clear that analysis of that data is often flawed to the point of a lack of reproducibility of the conclusion of the papers (Brown et al., 2018; Halsey et al., 2015; Peng, 2015). The failure to appropriately analyze data is apparent across the entire spectrum of basic, translational, and clinical research (Günel Karadeniz et al., 2019; Misra et al., 2021; Real, 2016; Strasak et al., 2007; Weissgerber et al., 2016). Sources of these errors range from a faulty initial design of the study and the hypothesis that will be tested, to the use of the wrong statistical testing, and to inappropriate documentation and presentation of the findings (Peng, 2015; Strasak et al., 2007). Despite the

widespread recognition that high-quality data analysis is critical in the peer review of scientific manuscripts, journals most often do not utilize statistical experts in the review of submitted manuscripts. In 2020, [Hardwicke and Goodman \(2020\)](#) repeated a survey conducted by [Goodman et al. \(1998\)](#). They surveyed 364 biomedical journals, receiving 107 replies, and found that 34% of journals utilized statistical review for only 10% or less of submissions, whereas another 34% used it for 10–50% of submissions ([Hardwicke and Goodman, 2020](#)). Despite the increased realization of the critical importance of statistics in the literature, these findings were little changed from the 1998 survey ([Goodman et al., 1998](#)).

The realization that we have a significant problem with both the replication and reproducibility of the scientific literature has been widely recognized, and over the last two decades, several attempts to address the problem have been documented. In 2010, [Kilkenny et al. \(2010\)](#) developed the ARRIVE (Animal Research: Reporting of In Vivo Experiments) guidelines for reporting animal research; these guidelines have goals similar to that of the CONSORT (Consolidated Standards of Reporting Trials) guidelines for clinical research. [Kilkenny et al. \(2021\)](#) developed a checklist for authors covering 10 essential points to be followed in preclinical animal research. These points addressed frequent issues in study design, statistical methods, and the reporting of results that authors needed to address in all studies involving animal research (<https://arriveguidelines.org/>). In 2014, the United States NIH announced plans to enhance reproducibility but recognized that “Efforts by the NIH alone will not be sufficient to effect real change in this unhealthy environment.” ([Collins and Tabak, 2014](#)). Also in 2014, a group of 30 editors of major journals, representatives of funding agencies, and scientific leaders met and reported a series of recommendations to develop guidelines for reporting preclinical research (<https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research>). In 2018, *Nature* published a collection of commentaries under the title “Challenges in Irreproducible Research” (<https://www.nature.com/collections/prbfkwmwvz>). Despite these and many other efforts, the problem has proven difficult to resolve. Although journals such as *PLoS* and *Nature* group journals published and supported the ARRIVE standards, [Baker et al. \(2014\)](#) found in an analysis of articles published in 2012 “...very little improvement in reporting standards...” ([Baker et al., 2014](#)) They concluded that “the editorial endorsement (of ARRIVE Checklist) is yet to be effectively implemented” ([Baker et al., 2014](#)).

JID Innovations is committed to publishing research that has rigor, is reproducible, and can be replicated. We believe that the process of replication and reproducibility of research is critical for the efficient advancement of science and for the self-correcting nature of science to be effective. We realize that science is never settled and that it is only through the process of effective peer review and the process of replication of results that our knowledge can truly advance.

Toward that end, *JID Innovations* has endorsed several policies that we believe are crucial to approaching these goals. *JID Innovations* believes in publishing negative studies and studies that may contradict (or not replicate) previous studies. We believe that making our scientific community

aware of this work will increase the efficiency as well as the reliability of the studies we publish. As an open-access publication, we have no limit on the length of the methods sections of the articles we publish. Indeed, we often ask authors to add more detail to their methods section, in sufficient detail to allow others to replicate their findings. We have also implemented double-anonymized peer review, where reviewers are not aware of the authors’ identity or the site where the research was done. We believe that this will help to assure authors and readers that articles published in *JID Innovations* are judged on the content of the submitted manuscript and not on where the work was done or who did it.

In January of 2023, we are implementing another innovation. We have established a statistical review board on our Editorial Board. These editors are experts in statistical analysis and data science. All manuscripts submitted to *JID Innovations* will be reviewed by one of these editors early in the review process. The purpose of this review is to provide an expert evaluation of the study design, data analysis, and statistical evaluation reported. It is our expectation that this will not only improve the quality of our publications but will also help educate our entire community about the importance of including statisticians and data scientists on study teams.

The goal of increasing the rigor, replicability, and reproducibility of the work of the scientists in our community will ultimately improve the quality of science and the efficiency of future work. It will assure the public that in these times of increasing questioning of scientific findings, we are committed to producing only the highest-quality science. We ask that you join *JID Innovations* as we seek to adhere to the standards of rigor and reproducibility of our studies and to encourage the publication of studies that replicate previous work or report the inability to confirm previous work. Together we must agree that these studies are vital for a healthy scientific community. We are convinced that this effort will move us closer to realizing the goal of the self-correcting nature of science, more efficiently advance our knowledge, and improve the understanding of how science works to improve our society.

CONFLICT OF INTEREST

The author states no conflict of interest.

Russell P. Hall III, MD

Editor*

*Correspondence: *Russell P. Hall III, MD. e-mail: InnovationsEditor@sidnet.org*

Cite this article as: *JID Innovations* 2023;3:100188.

REFERENCES

- Allison DB, Brown AW, George BJ, Kaiser KA. Reproducibility: a tragedy of errors. *Nature* 2016;530:27–9.
- Baker D, Lidster K, Sottomayor A, Amor S. Two years later: journals are not yet enforcing the ARRIVE guidelines on reporting standards for pre-clinical animal studies. *PLoS Biol* 2014;12:e1001756.
- Begley CG, Ellis LM. Raise standards for preclinical cancer research. *Nature* 2012;483:531–3.
- Brown AW, Kaiser KA, Allison DB. Issues with data and analyses: errors, underlying themes, and potential solutions. 2018;115:2563–2570.
- Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature* 2014;505:612–3.

- Errington TM, Denis A, Perfito N, Iorns E, Nosek BA. Challenges for assessing replicability in preclinical cancer biology. *Elife* 2021a;10:e67995.
- Errington TM, Denis A, Allison AB, Araiza R, Aza-Blanc P, Bower LR, et al. Experiments from unfinished Registered Reports in the Reproducibility Project: Cancer Biology. *Elife* 2021b;10:e73430.
- Errington TM, Iorns E, Gunn W, Tan FE, Lomax J, Nosek BA. An open investigation of the reproducibility of cancer biology research. *Elife* 2014;3:e04333.
- Errington TM, Mathur M, Soderberg CK, Denis A, Perfito N, Iorns E, et al. Investigating the replicability of preclinical cancer biology. *Elife* 2021c;10:e71601.
- Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals: caveat lector? *J Gen Intern Med* 1998;13:753–6.
- Günel Karadeniz P, Uzabacı E, Atış Kuyuk S, Kaskır Kesin F, Can FE, Seçil M, et al. Statistical errors in articles published in radiology journals. *Diagn Interv Radiol* 2019;25(2):102–8.
- Halsey LG, Curran-Everett D, Vowler SL, Drummond GB. The fickle P value generates irreproducible results. *Nat Methods* 2015;12:179–85.
- Hardwicke TE, Goodman SN. How often do leading biomedical journals use statistical experts to evaluate statistical methods? The results of a survey. *PLoS One* 2020;15:e0239598.
- Ioannidis JP. Why science is not necessarily self-correcting. *Perspect Psychol Sci* 2012;7:645–54.
- Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. *PLoS Biol* 2010;8:e1000412.
- Misra DP, Zimba O, Gasparyan AY. Statistical data presentation: a primer for rheumatology researchers. *Rheumatol Int* 2021;41:43–55.
- Peng R. The reproducibility crisis in science. A statistical counterattack 2015;12:30–2.
- Real J, Forné C, Roso-Llorach A, Martínez-Sánchez JM. Quality reporting of multivariable regression models in observational studies: review of a representative sample of articles published in biomedical journals. *Medicine (Baltimore)* 2016;95:e3653.
- Strasak AM, Zaman Q, Pfeiffer KP, Göbel G, Ulmer H. Statistical errors in medical research - a review of common pitfalls. *Swiss Med Wkly* 2007;137:44–9.
- Weissgerber TL, Garovic VD, Milin-Lazovic JS, Winham SJ, Obradovic Z, Trzeciakowski JP, et al. Reinventing biostatistics education for basic scientists. *PLoS Biol* 2016;14:e1002430.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>