



OPINION

Nov 17, 2013 [Editor](#) [1 Comment](#)

Resolving Irreproducibility in Empirical and Computational Research

Victoria Stodden writes:

The reproducibility of published findings is becoming a hot topic. From reports in the popular press to congressional activity, and from scholarly society engagement to academic publications and editorials, there has been an upsurge in attention to this issue. I will offer some explanations of the concept itself, suggest reasons why this topic is suddenly front and center, and outline ways the field of statistics can contribute to resolving the underlying issues all this attention is bringing to the fore.

Unpacking Reproducibility

The concept of reproducibility is getting attention in mainstream discussions. On October 19, *The Economist* magazine opened a Briefing on “Unreliable Research” with a quote from Nobel Laureate Daniel Kahneman, “I see a train wreck looming,” referring to the irreproducibility of certain psychological experiments [1]. On October 27, the *Los Angeles Times* informed us that “Science has lost its way” since it cannot be relied upon to generate “verifiable facts” [2]. Reproducibility is also discussed in scholarly communications [3–8]. In 2011 *Science Magazine* began requiring authors to remit code and data upon request for articles it publishes [9], and in April of this year *Nature* published an editorial entitled “Reducing our irreproducibility,” in which they encouraged researchers to make raw data available and follow a checklist for reporting methods, while extending the methods section to accommodate [10]. These are just a few examples.

These discussions have emerged from a wide variety of scientific disciplines, each with different practices that contextualize the meaning of reproducibility differently. At one end of the spectrum is the traditional scientific notion of experimental researchers capturing descriptive information about (non-computational) aspects of their research protocols and methods, labeled empirical reproducibility. For example, a spotlight was placed on empirical cancer research in 2011 when Bayer HealthCare in Germany could not validate the published findings upon which 67 of their in-house projects were based [11]. In 2012 Amgen’s attempts to replicate studies were published, and they claimed to have only been able to do so for 6 of 53 articles [12]. These results rocked the research community and, in part, prompted *Nature* to encourage authors to communicate their methods more completely. These efforts could be described as attempts to adhere more closely to the long-established standards of communication, as reflected in the title of the *Nature* editorial of March 2012: “Must try harder.” [13]

At the other end of the spectrum are the very different concerns arising from research communities that have adopted computational methods, labeled computational reproducibility [5, 14–16]. These voices call for new standards of scientific communication that include digital scholarly objects such as data and code, asserting that the traditional research article alone fails to capture the computational details and other information necessary for others to replicate the findings. Irreproducible computational results from genomics research at Duke University crystallized attention to this issue [17]. As a result the Institute of Medicine of the National Academies published a report in 2012 recommending new standards for clinical trials approval for computational tests arising from omics-based research [18]. The report recommended for the first time that the software associated with a computational test be fixed at the beginning of the approval process, and made “sustainably available.” In December of 2012 a workshop on “Reproducibility in Computational and Experimental Mathematics” produced recommendations regarding information to include with publications of computational findings, including access to code, data, and implementation details [19–22]. The distinction between these two types of reproducibility is important in order to understand their sources and appropriate solutions.

Resolving Irreproducibility

The reasons, and therefore the remedies, differ depending on the type of reproducibility. In the case of computational reproducibility, issues arises from an exogenous shift in the scientific research process itself — the broad use of computation — and the proposed solution seeks to extend the standards of

Search for:

Search



[RSS feed](#)

WELCOME!

Welcome to the new and improved *IMS Bulletin* website! We are developing the way we communicate news and information more effectively with members. The print *Bulletin* is still with us (free with [IMS membership](#)), and still [available as a PDF to download](#), but in addition, we are placing some of the news, columns and articles on this blog site, which will allow you the opportunity to interact more. We are always keen to hear from IMS members, and encourage you to write articles and reports that other IMS members would find interesting. Contact the *IMS Bulletin* at bulletin@imstat.org

WHAT IS “OPEN FORUM”?

With this new blog website, we are introducing a new feature, the **Open Forum**. Any IMS member can propose a topic for discussion. Email your subject and an opening paragraph (send this to bulletin@imstat.org) and we'll post it to start off the discussion. Other readers can join in the debate by commenting on the post. Search other Open Forum posts by using the Open Forum category link below. Start a discussion today!

CATEGORIES

- [Anirban's Angle](#)
- [From the Editor](#)
- [Hand writing](#)
- [IMS awards](#)
- [IMS news](#)
- [Journal news](#)
- [Lectures and Addresses](#)
- [Letters](#)
- [Meetings](#)

transparency established for empirical science to the computational aspects of the research. In the case of empirical reproducibility, which lacks an obvious change to the underlying research process, we must look further.

Several reasons have been postulated for the reported lack of reproducibility in empirical research, beyond mistakes or misconduct such as outright fraud or falsification. Small study size, inherently small effect sizes, early or novel research without previously established evidence, poorly designed protocols that permit flexibility during the study, conflicts of interest, or the trendiness of the research topic have been previously suggested as contributing to irreproducibility in the life sciences [4]. Others include social reasons such as publication bias toward positive findings or established authors, or ineffective peer review [24]. Statistical biases may stem from misapplied methodology, incorrect use of p-values, a failure to adjust for multiple comparisons, or overgeneralization of the results. [25-27]. Statistical methods have varying degrees of sensitivity to perturbations in the underlying data, and can produce different findings in replication contexts [28-29]. Many fields have been inundated with vast amounts of data, often collected in novel ways or from new sources, rapidly shifting the context within which statistical methods must operate. Developing a research agenda within the statistical community to address issues surrounding reproducibility is imperative.

New Research Directions

Addressing issues of reproducibility through improvements to the research dissemination process is important, but insufficient. Research directions that would contribute to resolving these new methodological questions could include: new measures to assess the reliability and stability of empirical inferences, including developing new validation measures; expanding the field of uncertainty quantification to develop measures of statistical confidence and a better understanding of sources of error, especially when large multi-source datasets or massive simulations are involved [30-31]; and detecting biases arising from statistical reporting conventions. In addition, advances in understanding how to best archive software and data for replication purposes, and the development of best research practices are essential. This is not an exhaustive list, but intended to jumpstart thinking about the importance a research agenda in reproducibility as it relates to developing, asserting, and interpreting statistical findings.

For students and others wishing to learn more about reproducible research further information is available my wiki page <http://wiki.stodden.net/>. For an example of teaching reproducible research, see Gary King's course website, where students replicate findings from a published article [32]. I have taught a similar course at Columbia [33].

References

- [1] "Unreliable Research: Trouble at the Lab," The Economist, Oct 19, 2013. <http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble>
- [2] "Science has lost its way, at a big cost to humanity," Los Angeles Times, Oct 27, 2013. <http://www.latimes.com/business/la-fi-hiltzik-20131027,0,1228881.column>
- [3] D. L. Donoho, A. Maleki, I. U. Rahman, M. Shahram and V. Stodden, Reproducible Research in Computational Harmonic Analysis, Computing in Science and Engineering, vol. 11, no. 1, pp. 8-18, Jan./Feb. 2009, doi:10.1109/MCSE.2009.15
- [4] Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med 2(8): e124. doi:10.1371/journal.pmed.0020124 <http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0020124>
- [5] Stodden, V., "Reproducible Research: Tools and Strategies for Scientific Computing," July/August 2012 (vol. 14 no. 4), pp. 11-12. <http://www.computer.org/csdl/mags/cs/2012/04/mcs2012040011-abs.html>
- [6] R. D. Peng, Reproducible Research in Computational Science, Science, Vol. 334, p. 1226-1227, 2011, doi:10.1126/science.1213847
- [7] Sandve GK, Nekrutenko A, Taylor J, Hovig E (2013) Ten Simple Rules for Reproducible Computational Research. PLoS Comput Biol 9(10): e1003285. doi:10.1371/journal.pcbi.1003285
- [8] Brian A. Nosek, Jeffrey R. Spies and Matt Motyl, "Scientific Utopia II. Restructuring Incentives and Practices to Promote Truth Over Publishability," Perspectives on Psychological Science November 2012 vol. 7 no. 6, 615-631. <http://pps.sagepub.com/content/7/6/615.abstract>
- [9] Brooks Hanson, Andrew Sugden, Bruce Alberts, "Making Data Maximally Available," Science, Vol. 331 no. 6018 p. 64, 2011. doi:10.1126/science.1203354 <http://www.sciencemag.org/content/331/6018/649.full>
- [10] Announcement: Reducing our irreproducibility, Nature 496, 398, 25 April 2013. doi:10.1038/496398a <http://www.nature.com/news/announcement-reducing-our-irreproducibility-1.12852>
- [11] Florian Prinz, Thomas Schlange, Khusru Asadullah, "Believe it or not: how much can we rely on published data on potential drug targets?," Nature Reviews Drug Discovery 10, 712 (September 2011) | doi:10.1038/nrd3439-c1 <http://www.nature.com/nrd/journal/v10/n9/full/nrd3439-c1.html>
- [12] C. Glenn Begley & Lee M. Ellis, "Drug development: Raise standards for preclinical cancer research," Nature 483, 531-533 (29 March 2012) doi:10.1038/483531a <http://www.nature.com/nature/journal/v483/n7391/full/483531a.html>
- [13] Editorial, "Must Try Harder," Nature 483, 509 (29 March 2012) doi:10.1038/483509a <http://www.nature.com/nature/journal/v483/n7391/full/483509a.html>
- [14] King, Gary. 1995. "Replication, Replication." Political Science and Politics 28: 443-499. <http://gking.harvard.edu/files/abs/replication-abs.shtml>
- [15] LeVeque, R. J., 2006, "Wave propagation software, computational science, and reproducible

• [Member news](#)

• [Nominations](#)

• [Obituary](#)

• [Open Forum](#)

• [Opinion](#)

• [Other news](#)

• [Rick's Ramblings](#)

• [Robert Adler](#)

• [Statistics2013](#)

• [Stéphane Boucheron](#)

• [Student Puzzle Corner](#)

• [Terence's Stuff](#)

• [Vlada's Point](#)

• [Welcome](#)

• [XL Files](#)

RECENT POSTS

• [US National Academy of Sciences elects new Members](#)

• [Edward Waymire selected to receive Carver Medal](#)

• [Members News: June/July 2014](#)

• [IMS-ASC 2014 in Sydney](#)

• [Vlada's Point: The WorkSHOP](#)

ABOUT IMS

The Institute of Mathematical Statistics is an international scholarly society devoted to the development and dissemination of the theory and applications of statistics and probability. We have about 4,500 members around the world. Visit IMS at <http://imstat.org>

LINKS

• [Employment opportunities](#)

• [IMS homepage](#)

• [Latest IMS Bulletin issue \(PDF file\)](#)

• [Meetings](#)

• [RSS feed](#)

research," Proc. International Congress of Mathematicians, (M Sanz-Sole, J. Soria, J. L. Varona and J. Verdera, eds.) Madrid, August 22-30, 2006, pp. 1227-1254.

<http://faculty.washington.edu/rjl/pubs/icm06/icm06leveque.pdf>

[16] David L. Donoho, Arian Maleki, Inam Ur Rahman, Morteza Shahram, Victoria Stodden, "Reproducible Research in Computational Harmonic Analysis", IEEE Computing in Science and Engineering, 11(1), January 2009, p.8-18.

<http://www.computer.org/csdl/mags/cs/2009/01/mcs2009010008-abs.html>

[17] Kaiser, J., "Panel Calls for Closer Oversight of Biomarker Tests," ScienceInsider, March 23, 2012.

<http://news.sciencemag.org/education/2012/03/panel-calls-closer-oversight-biomarker-tests>

[18] "Evolution of Translational Omics: Lessons Learned and the Path Forward," Institute of Medicine Report, National Academies of Science, 2012. <http://www.iom.edu/reports/2012/evolution-of-translational-omics.aspx>

[19] ICERM Workshop, "Reproducibility in Computational and Experimental Mathematics," December 10-14, 2012. <http://icerm.brown.edu/tw12-5-rcem>

[20] ICERM Workshop Report, Appendix D, http://stodden.net/icerm_report.pdf

[21] D. H. Bailey, J. M. Borwein, Victoria Stodden, "Set the Default to 'Open'," Notices of the American Mathematical Society, June/July 2013. <http://www.ams.org/notices/201306/moti-p679.pdf>

[22] Victoria Stodden, Jonathan Borwein, and David H. Bailey, "'Setting the Default to Reproducible' in Computational Science Research," SIAM News, June 3, 2013. <http://www.siam.org/news/news.php?id=2078>

[24] Francis, G. (2012). The psychology of replication and replication in psychology. Perspectives on Psychological Science, 7, 585– 594. <http://pps.sagepub.com/content/7/6/585.abstract>

[25] J. Berger, "Reproducibility of Science, P-values and Multiplicity," SBSS Webinar, Oct 4, 2012.

[26] Gelman A., "The connection between varying treatment effects and the crisis of unreplicable research: A Bayesian perspective," forthcoming in the Journal of Management, 2014.

http://www.stat.columbia.edu/~gelman/research/published/bayes_management.pdf

[27] Simmons J., Nelson L., Simonsohn U. (2011) "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allow Presenting Anything as Significant", Psychological Science, V22(11), 1359-1366. http://papers.ssm.com/sol3/papers.cfm?abstract_id=1850704

[28] Yu, B., "Stability," Bernoulli, Volume 19, Number 4 (2013), 1484-1500.

<http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.bj/1377612862>

[29] Chingway Lim, Bin Yu, "Estimation Stability with Cross Validation (ESCV)," 2013.

<http://arxiv.org/abs/1303.3128>

[30] National Research Council Report, "Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification," 2012.

http://www.nap.edu/catalog.php?record_id=13395

[31] National Research Council Report, "Frontiers in Massive Data Analysis," 2013.

http://www.nap.edu/catalog.php?record_id=18374

[32] <http://projects.iq.harvard.edu/gov2001/>

[33] <http://www.stodden.net/class/stat8325/STAT8325SyllabusFall2012.pdf>

1 Comment

article: Resolving Irreproducibility in Empirical and Computational Research [REPLY](#)
Research « Reproducible Research Ideas December 9, 2013 at 7:41 pm

[...] nice article on reproducible research by Victoria Stodden in IMS Bulletin: V. Stodden, Resolving Irreproducibility in Empirical and Computational Research, IMS Bulletin, Nov [...]

Leave a comment

Name (required)

Mail (will not be published) (required)

Website

Spam protection: Sum of 9 + 10 ?

*

Comment

POST COMMENT