



HHS Public Access

Author manuscript

Account Res. Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:

Account Res. 2017 ; 24(2): 116–123. doi:10.1080/08989621.2016.1257387.

Reproducibility and Research Integrity

David B. Resnik, JD, PhD and

National Institute of Environmental Health Sciences, National Institutes of Health

Adil E. Shamoo, PhD

University of Maryland School of Medicine, University of Maryland, Baltimore

Reproducibility—the ability of independent researchers to obtain the same (or similar) results when repeating an experiment or test—is one of the hallmarks of good science (Popper 1959). Reproducibility provides scientists with evidence that research results are objective and reliable and not due to bias or chance (Rooney et al 2016). Irreproducibility, by contrast, may indicate a problem with any of the steps involved in the research such as, but not limited to, the experimental design, variability of biological materials (such as cells, tissues or animal or human subjects), data quality or integrity, statistical analysis, or study description (Landis et al 2012, Shamoo and Resnik 2015). Although researchers have understood the importance of reproducibility for quite some time (Shamoo and Annau 1987), in the last few years the issue has become a pressing concern because of an increasing awareness among scientists and the public that the results of many studies published in peer-reviewed journals are not reproducible (Ioannidis 2005, The Economist 2013, Collins and Tabak 2014, McNutt 2014, Baker 2015).

Some of the irreproducibility in scientific research may be due to data fabrication or falsification (Shamoo 2013, 2016; Collins and Tabak 2014, Kornfeld and Titus 2016). Misconduct or suspected misconduct accounts for more than two-thirds of retractions (Fang et al 2012). The website Retraction Watch (2016) keeps track of papers which have been retracted due to misconduct or other problems. Approximately 2% of scientists claim that they have fabricated or falsified data at some point in their careers (Fanelli 2009). This percentage may underestimate the actual rate of misconduct because respondents may be unwilling to admit to engaging in illegal or unethical behavior, even in anonymous surveys. Even if the rate of misconduct is low, it still represents a serious ethical problem that can undermine the reproducibility, integrity and trustworthiness of research (Shamoo and Resnik, 2015).

For example, Gottmann et al (2001) compared 121 rodent carcinogenicity assays from the National Cancer Institute/National Toxicology Program and the Carcinogenic Potency Database and found that reproducibility was only 57%. To explain the declining success rates in Phase II drug trials documented by Arrowsmith (2011), scientists from a large pharmaceutical company speculated that poor quality pre-clinical research may be at fault. To test their hypothesis, they analyzed 67 in-house drug target validation projects and found that only 20–25% were reproducible (Prinz et al 2012). In response to evidence of irreproducible results in psychology, a group of 270 researchers attempted to reproduce 100 experiments published in three top psychology journals in 2008. They found that the

percentage of studies reporting statistically significant results declined from 97% for the original studies to 36% for the replications, and effects sizes decreased by 50% (Open Science Collaboration 2015). Irreproducibility of behavioral studies may be due, in part, to variations in populations and difficulties with accurately measuring behavioral parameters (Shamoo 2016). While most of the attention has focused on irreproducibility in biomedical (Kilkenny et al 2009, Landis et al 2012, Pusztai et al 2013, Begley and Ioannidis 2015) and psychological research (Open Science Collaboration 2015), many are concerned that reproducibility problems may also plague “hard” sciences like physics and chemistry (The Economist 2013, Nature 2016).

Scientific journals, funding agencies, and researchers have responded to the reproducibility “crisis” by articulating standards for designing experiments, analyzing data, and reporting methods, materials, data, and results (Landis et al 2012, Pusztai et al 2013, Collins and Tabak 2014, McNutt 2014, The Science Exchange Network 2014, Nature 2014a, National Institutes of Health 2016, Rooney et al 2016). While many of these standards tend to be discipline-specific, some apply across disciplines. For example, statistical power analysis can help to ensure that one’s sample is large enough to detect a significant effect in biomedical, physicochemical, or behavioral research; randomization and blinding can control for bias in clinical trials or animal experiments; and auditing of data and other research records can help reduce errors and inconsistencies in many fields of science (Shamoo and Resnik 2015, National Institutes of Health 2016, Shamoo 2016).

Reproducibility is not just a scientific issue; it is also an ethical one. When scientists cannot reproduce a research result, they may suspect data fabrication or falsification. In several well-known cases, reproducibility issues led to allegations of data fabrication or falsification. For example, in 1986, postdoctoral researcher Margot O’Toole accused her supervisor, Tufts University pathology assistant professor Thereza Imanishi-Kari, of fabricating and falsifying data in a National Institutes of Health (NIH)-funded study on using foreign genes to stimulate antibody production in mice, published in the journal *Cell*. O’Toole became suspicious of the research after she was unable to reproduce a key experiment conducted by Imanishi-Kari and found discrepancies between the data recorded in Imanishi-Kari’s laboratory notebooks and the data reported in the paper. The case made national headlines, in part, because Nobel Prize-winning molecular biologist David Baltimore was one of the coauthors on the paper, even though he was never implicated in the scandal. A Congressional committee headed by Rep. John Dingell discussed the case during its hearings on fraud in federally-funded research. In 1994, the Office of Research Integrity, which oversees NIH-funded research, found that Imanishi-Kari committed misconduct, but a federal appeals panel overturned this ruling in 1996 (Shamoo and Resnik 2015).

In March 1989, University of Utah chemistry professor Stanley Pons and Southampton University chemistry professor Martin Fleischmann announced at a press conference that they had developed a method for producing nuclear fusion at room temperatures (i.e. “cold fusion”). Pons and Fleischmann bypassed the peer review process and reported their results directly to the public in order to protect their claims to priority and intellectual property. When physicists and chemists around the world tried, unsuccessfully, to reproduce these exciting results, many of them accused Pons and Fleischman of conducting research that was

sloppy, careless, or fraudulent. While it does not appear that Pons and Fleischmann fabricated or falsified data, one of the ethical problems with their work was that they did not provide enough detail in their press release to enable scientists to reproduce their experiments. By leading their colleagues on a wild goose chase, they wasted the scientific community's time and resources and tainted cold fusion research for years to come (Shamoo and Resnik 2015).

In March 2014, Haruko Obokata, a biochemist at the RIKEN Center for Developmental Biology in Kobe, Japan, and coauthors published two high-profile papers in *Nature* describing a method for converting adult spleen cells in mice into pluripotent stem cells by means of chemical stimulation and physical stress. Several weeks after the papers were published, researchers at the RIKEN Center were unable to reproduce the results and they accused Obokata, who was the lead author on the papers, of misconduct. The journal retracted both papers in July after an investigation by the RIKEN center found that Obokata had fabricated and falsified data. Later that year, Obokata's advisor, Yoshiki Sasai, committed suicide by hanging himself (Cyranoski 2014).

When irreproducibility does not result from misconduct, it can still have serious consequences for science and society. Irreproducible results could cause severe harms in medicine, public health, engineering, aviation, and other fields in which practitioners or regulators rely on published research to make decisions affecting public safety and well-being (Horton 2015). Irreproducible research, which does not have any immediate practical applications, may still have a negative impact on science by causing researchers to rely on invalid data. Researchers need to be able to trust that published data are reliable, and reproducibility problems can undermine that trust (Shamoo and Resnik 2015). Moreover, the public needs to be able to have confidence in the reliability and integrity of science, and irreproducible research can undermine that trust.

Pressures to produce results and publish may be important factors in science's reproducibility problems (Horton 2015, Shamoo and Resnik 2015). Researchers who are trying to publish results to advance their careers or meet deadlines imposed by supervisors or sponsors may cut corners when designing and implementing experiments. For example, if a researcher has obtained a result that is marginally statistically significant, he or she may decide to go ahead and publish the result without replicating the result and carefully considering whether it is a false positive.

The growing number of for-profit, open access scientific journals which charge high publication fees, i.e. "predatory journals," may also exacerbate reproducibility problems (Clark and Smith 2015). These journals often promise rapid publication and have negligible peer review. While it is not known how many articles published in these journals report irreproducible results, the poor peer review standards found in these journals present a significant threat to the quality and integrity of published research (Beall 2016).

Adherence to some commonly recognized principles of responsible conduct of research (RCR) plays an important role in promoting reproducibility in science. One of the key pillars of RCR is that scientific records, including laboratory notebooks, protocols, and other

documents, should describe one's research in sufficient detail to allow others to reproduce it (Schreier et al 2006, Shamoo and Resnik 2015). Records should be accurate, thorough, clear, backed-up, signed, and dated. Failure to record a vital piece of information, such as a change in an experimental design, the pH of a solution, or the type of food fed to an animal, the time of year, can lead to problems with reproducibility (Buck 2015, National Institutes of Health 2016, Firestein 2016).

However, it is important to note that scientists may not always know all the factors that could impact research outcomes. For example, Sorge et al (2014) found that exposure to human male odors, but not female odors, induces pain inhibition and stress in mice and rats. Failure to record the sexes of the researchers conducting experiments on rodents involving measurements of pain responses could therefore undermine reproducibility. Prior to the publication of this finding, many researchers would have not recorded or reported the sexes of the experimenters, or taken this information into account in studies or pain or stress in rodents. If researchers are having difficulty reproducing the results of an experiment, they may need to reexamine methods, materials, and procedures to determine whether they have overlooked an important detail.

Transparency involves the honest and open disclosure of all information related to one's research when submitting it for publication (Landis et al 2012). Information may be disclosed in the materials and methods section of the paper or in appendices. Because many journals have space constraints that limit the length of articles published in print, some disclosures may need to occur online in supporting documents. Many journals require authors to make supporting data available on public websites (Shamoo and Resnik 2015).

While required disclosures depend on the nature of research one is conducting, some general types of information which should be disclosed include: the research design (e.g. controlled trial, prospective cohort study), methods (e.g. blinding, randomization), procedures, techniques, materials, equipment, data analysis methods and tools (including computer programs or codes), study population (for animals or humans), exclusion and inclusion criteria (for animals and humans), ethics committee approvals (if appropriate), theoretical assumptions and potential biases, sources of funding, and conflicts of interest (Landis et al 2012, Nature 2014b, McNutt 2014, Nature 2015, Elliott and Resnik 2015, Rooney et al 2016, Morgan et al 2016). Additional disclosures may need to occur after the research is published to allow independent scientists to obtain information needed to reproduce experiments, reanalyze data, or develop new hypotheses or theories related to the research.

Some researchers may decide to retract a paper or publish an expression of concern if its results cannot be reproduced. For example, Casadevall et al (2014) examined 423 retraction notices indexed in PubMed that cited error (as opposed to misconduct) as the reason for the retraction and found that 16.1% of these were retracted due to irreproducibility. Other types of errors included laboratory error (55.8%), analytical error (18.9%). 9.2% of the retraction notices were classified as "other" error.

The Committee on Publication Ethics (2009) has developed guidelines for retracting articles. According to the guidelines, "Journal editors should consider retracting a publication if: they

have clear evidence that the findings are unreliable, either as a result of misconduct (e.g. data fabrication) or honest error (e.g. miscalculation or experimental error); the findings have previously been published elsewhere without proper crossreferencing, permission or justification (i.e. cases of redundant publication); it constitutes plagiarism; or it reports unethical research (Committee on Publication Ethics 2009: 1).” The purpose of retracting an article is to provide a “mechanism for correcting the literature and alerting readers to publications that contain such seriously flawed or erroneous data that their findings and conclusions cannot be relied upon (Committee on Publication Ethics 2009: 2).” Retraction notices should provide a clear explanation of the reasons for the retraction and should be linked to the original article. Retracted articles should be clearly identified in electronic versions of the journal and bibliographic databases (Committee on Publication Ethics 2009: 2).”

To help students and trainees better understand how to promote reproducibility in their work, courses in research methodology and RCR should include sections devoted to the importance of reproducibility and issues that can impact it, such as experimental design, record-keeping, biological variability, data analysis, and transparency (Titus et al 2008, Shamoo and Resnik 2015, National Institutes of Health 2016). The NIH has developed some online reproducibility training modules which are available to the public (National Institutes of Health 2015). The modules address topics such as blinding, randomization, transparency, record-keeping, bias, sample size, and data analysis. Several World Conferences on Research Integrity (2016) have provided forums for international discussion of ethics issues in research and science education, including those that impact reproducibility.

Researchers should also informally discuss reproducibility issues with their students and trainees as part of the mentoring process. For example, if a student or trainee is having difficulty repeating an experiment, the mentor should help him or her to understand what may have gone wrong and how to fix the problem. The mentor may also be able to share stories of his or her own experimental successes and failures with the student to illustrate reproducibility concepts. Failure to replicate an experiment need not be a total loss but can be an opportunity to teach students about principles of good science (Firestein 2016).

Acknowledgments

This research was supported, in part, by the National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH). It does not represent the views of the NIEHS, NIH, or US federal government.

References

- Arrowsmith J. Trial watch: Phase II failures: 2008–2010. *Nature Reviews Drug Discovery*. 2011; 10(5):328–329.
- Baker M. The reproducibility crisis is good for science. *Slate*. 2015 Apr 15. [Accessed: May 31, 2016] 2015 Available at: http://www.slate.com/articles/technology/future_tense/2016/04/the_reproducibility_crisis_is_good_for_science.html.
- Beall J. Dangerous predatory publishers threaten medical research. *Journal of Korean Medical Science*. 2016; 31(10):1511–1513. [PubMed: 27550476]
- Begley CG, Ioannidis JP. Reproducibility in science: improving the standard for basic and preclinical research. *Circulation Research*. 2015; 116(1):116–126. [PubMed: 25552691]

- Buck S. Solving reproducibility. *Science*. 2015; 348(6242):1403. [PubMed: 26113692]
- Casadevall A, Steen RG, Fang FC. Sources of error in the retracted scientific literature. *FASEB Journal*. 2014; 28(9):3847–3855. [PubMed: 24928194]
- Clark J, Smith R. Firm action needed on predatory journals. *BMJ*. 2015; 350:h210. [PubMed: 25596387]
- Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature*. 2014; 505(7485):612–613. [PubMed: 24482835]
- Committee on Publication Ethics. Retraction guidelines. 2009 [Accessed: June 21, 2016] Available at: <http://publicationethics.org/files/retraction%20guidelines.pdf>.
- Cyranoski D. Collateral damage: How a case of misconduct brought a leading Japanese biology institute to its knees. *Nature*. 2015; 520(7549):600–603. [PubMed: 25925457]
- Elliott KC, Resnik DB. Scientific reproducibility, human error, and public policy. *Bioscience*. 2015; 65(1):5–6. [PubMed: 26955072]
- Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*. 2009; 4(5):e5738. [PubMed: 19478950]
- Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences USA*. 2012; 109(42):17028–17033.
- Firestein S. Why failure to replicate findings can actually be good for science. *LA Times*. 2016 Feb 14. [Accessed: September 4, 2016] 2016 Available at: <http://www.latimes.com/opinion/op-ed/la-oe-0214-firestein-science-replication-failure-20160214-story.html>.
- Gottmann E, Kramer S, Pfahringer B, Helma C. Data quality in predictive toxicology: reproducibility of rodent carcinogenicity experiments. *Environmental Health Perspectives*. 2001; 109(5):509–514. [PubMed: 11401763]
- Horton R. Offline: What is medicine's 5 sigma? *The Lancet*. 2015; 385:1380.
- Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005; 2:696–701.
- Kilkenny C, Parsons N, Kadyszewski E, Festing MF, Cuthill IC, Fry D, Hutton J, Altman DG. Survey of the quality of experimental design, statistical analysis and reporting of research using animals. *PLoS One*. 2009; 4(11):e7824. [PubMed: 19956596]
- Kornfeld DS, Titus SL. Stop ignoring misconduct. *Nature*. 2016; 537(7618):29–30. [PubMed: 27582206]
- Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG, Darnell RB, Ferrante RJ, Fillit H, Finkelstein R, Fisher M, Gendelman HE, Golub RM, Goudreau JL, Gross RA, Gubitza AK, Hesterlee SE, Howells DW, Huguenard J, Kelner K, Koroshetz W, Krainc D, Latic SE, Levine MS, Macleod MR, McCall JM, Moxley RT 3rd, Narasimhan K, Noble LJ, Perrin S, Porter JD, Steward O, Unger E, Utz U, Silberberg SD. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature*. 2012; 490(7419):187–191. [PubMed: 23060188]
- McNutt M. Reproducibility. *Science*. 2014; 343(6168):229. [PubMed: 24436391]
- Morgan RL, Thayer KA, Bero L, Bruce N, Falck-Ytter Y, Ghersi D, Guyatt G, Hooijmans C, Langendam M, Mandrioli D, Mustafa RA, Rehfues EA, Rooney AA, Shea B, Silbergeld EK, Sutton P, Wolfe MS, Woodruff TJ, Verbeek JH, Holloway AC, Santesso N, Schünemann HJ. GRADE: Assessing the quality of evidence in environmental and occupational health. *Environment International*. 2016; 92–93:611–616.
- National Institutes of Health. [Accessed: June 21, 2016] Clearinghouse for training modules to enhance data reproducibility. 2015. Available at: <https://www.nigms.nih.gov/training/pages/clearinghouse-for-training-modules-to-enhance-data-reproducibility.aspx>
- National Institutes of Health. [Accessed: May 30, 2016] Rigor and Reproducibility. 2016. Available at: <https://www.nih.gov/research-training/rigor-reproducibility>
- Nature. Journals unite for reproducibility. *Nature*. 2014a; 515(7525):7.
- Nature. Code share. *Nature*. 2014b; 514(7524):536. [PubMed: 25355323]
- Nature. Let's think about cognitive bias. *Nature*. 2015; 526(7572):163. [PubMed: 26450020]
- Nature. Reality check on reproducibility. *Nature*. 2016; 533(7604):437.

- Open Science Collaboration. Psychology. Estimating the reproducibility of psychological science. *Science*. 2015; 349(6251):aac4716. [PubMed: 26315443]
- Popper, K. *The Logic of Scientific Discovery*. London, UK: Routledge; 1959.
- Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*. 2011; 10(9):712.
- Pusztai L, Hatzis C, Andre F. Reproducibility of research and preclinical validation: problems and solutions. *Nature Reviews Clinical Oncology*. 2013; 10(12):720–724.
- Retraction Watch. 2016 [Accessed: September 5, 2016] Available at: <http://retractionwatch.com/>.
- Rooney AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Boyles AL, Ratcliffe JM, Kraft AD, Schünemann HJ, Schwingl P, Walker TD, Thayer KA, Lunn RM. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. *Environment International*. 2016; 92–93:617–629.
- Schreier AA, Wilson K, Resnik D. Academic research record-keeping: best practices for individuals, group leaders, and institutions. *Academic Medicine*. 2006; 81(1):42–47. [PubMed: 16377817]
- Shamoo AE. Data audit as a way to prevent/contain misconduct, *Accountability in Research*. 2013; 20(5–6):369–379.
- Shamoo AE. Audit of research data, *Accountability in Research*. 2016; 23(1):1–3.
- Shamoo, AE., Resnik, DB. *Responsible Conduct of Research*. 3rd. New York, NY: Oxford University Press; 2015.
- Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, Tuttle AH, Wieskopf JS, Acland EL, Dokova A, Kadoura B, Leger P, Mapplebeck JC, McPhail M, Delaney A, Wigerblad G, Schumann AP, Quinn T, Frasnelli J, Svensson CI, Sternberg WF, Mogil JS. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nature Methods*. 2014; 11(6):629–632. [PubMed: 24776635]
- The Economist. Trouble at the lab. 2013 Oct 19. [Accessed: May 31, 2016] 2013 Available at: <http://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarms-degree-it-not-trouble>.
- The Science Exchange Network. Reproducibility initiative. 2014 [Accessed: September 5, 2016] Available at: <http://validation.scienceexchange.com/#/reproducibility-initiative>.
- Titus SL, Wells JA, Rhoades LJ. Repairing research integrity. *Nature*. 2008; 453(7198):980–982. [PubMed: 18563131]
- World Conferences on Research Integrity. 2016. Available at: <http://www.researchintegrity.org/>