8 Peer Review

Arthur L. Caplan and Barbara K. Redman

Many date editorial peer review to the 1752 Royal Society of London's use of a "Committee on Papers" to oversee the review of text for publication in the journal *Philosophical Transactions*. Initially, peer review was created to help editors decide what to publish. In the twentieth century it evolved into a system in which qualified peers not only judge publication merit but also evaluate the quality of scientific work including grant applications, conference proposals, books, and academic personnel actions. Today, it is the major tool in scientific self-regulation. It is often undertaken double 'blinded' so that reviewers do not know the names of those they review and vice versa. Peer reviewers names for undertaking specific tasks are often expected to be confidential.

Reviews can be open, single-blind (reviewer knows author but not vice versa), or double-blind (neither knows the other). Post-publication review is now common, although the mechanisms by which it accomplished are fragmented. PubMed Commons (https://www.ncbi.nlm.nih.gov/pubmed-commons), in which comments are attached to an article's PubMed record, is one such mechanism for post peer-review commentary. So are journals that utilize the format of target articles with extensive commentaries.

Complaints about peer review include erroneous rejection of important findings, unreliability in the detection of errors and fraud, intellectual plagiarism by reviewers, purposeful delay and undisclosed conflict of interest when reviewers and authors compete for the same funds or publications. Poor agreement among reviewers is seen as both a weakness and as a strength in bringing diverse perspectives to bear. Several kinds of reviewer bias have been noted: confirmation bias in which current beliefs are affirmed rather than challenged, publication bias for positive rather than negative outcomes or replications, bias against certain kinds of methodology (qualitative studies), and embargoing clinically important findings until all peer review is completed. (Manchikanti et al. 2015).

Two studies of peer review are helpful. A review of papers submitted to *Annals of Internal Medicine, British Medical Journal*, and *Lancet* concluded that peer review added value by filtering out submissions of poor quality but had problems dealing with exceptional or unconventional papers published later in other journals (Siler et al. 2015). A study in the social sciences found reviewers made considerable useful contributions to manuscript revision, particularly of interpretations of findings (Strang and Siler 2015).

Peer review is a prime duty of being part of a scientific community and enforcing norms of research integrity. Peer review fraud has been uncovered and dealt with. In 2015, Springer retracted 64 articles from ten different journals in which an individual invented fake email addresses and reviewed his own manuscripts (Haug 2015). Peer review will continue to be a major form of quality control in science but reviewers must disclose conflicts of interest and describe any limitations in their ability to undertake peer review to those making requests.

Advice: Expect that peer review will be imperfect but know that you can always learn from reviewers' comments. Address them directly and explicitly when you revise a manuscript or grant application for resubmission.

Sometimes reviewer comments mean that your manuscript or application is a mismatch with a journal or funding source so you should find other alternatives. Mentors should spend time explaining how to do peer review, and if they do not, mentees should ask before undertaking peer review work.

A. L. Caplan · B. K. Redman (✉)
New York University Langone Medical Center, New York, NY, USA
e-mail: Arthur.Caplan@nyumc.org

© Springer International Publishing AG, part of Springer Nature 2018
A. L. Caplan, B. K. Redman (eds.), *Getting to Good*, https://doi.org/10.1007/978-3-319-51358-4_8

8.1 Let's Make Peer Review Scientific

Drummond Rennie

Rennie, D. Let's make peer review scientific. *Nature* 535, 31–33 (2016). © 2016 Macmillan Publishers Limited. All rights reserved. An imprint of SpringerNature.

Illustration reproduced courtesy of David Parkins.

COMMENT

uncertainty is needed, as is standard practice in other fields — even bathroom scales come with uncertainties printed on them. A mark should signify that the sensor meets a minimum quality standard

If such a stamp of approval sounds bureaucratic, think of how the data might be used. People with asthma might use their local sensor data to make personal decisions on medication; an air-pollution sensor is not meant as a medical device, but its real-world application could make it function like one. Privately owned sensor data could trigger legal actions in areas that apparently exceed local air-quality standards. The economic and socially disruptive costs of closing roads or banning cars based on live sensor data would be huge.

NEXT STEPS

The academic air-pollution community must do the hard yards in the lab and field on calibration and testing. It must also find ways to overcome some measurement challenges. Researchers should take the lead on evaluating sensor performance, creating better devices and designing research applications that are suited to the quantified capabilities of sensors.

More creativity is needed in experimental design. If the long-term performance of sensors is a problem, as is likely, then we need to design shorter-term experiments that can be performed reliably. For example, a fine-scale but qualitative measure of pollution might help to simulate the turbulent flows of pollution in street canyons or tree canopies over a few days. There might be experiments in which a fast-responding bulk sensor — one that measures the sum of many organic compounds, for example — might be able to track rapid temporal changes that add context to a slower but more quantitative instrument, such as a gas chromatograph or diffusion tube. Statistical and machine-learning methods might be developed to enable better extraction of signals from a mix of pollutants[8].

> *"Manufacturers and regulators need to define how and where sensors can be used."*

However, academics should not become gatekeepers or validation bodies. This is a job for manufacturers and regulators, who need to define how and where sensors can and cannot be used effectively.

Governments must provide advice now to potential 'professional users', such as in cities and regional environmental agencies. For sensors that might be used for public policy, health studies or any type of infrastructure control, independent testing and verification is essential, as is already being done through long-standing environment-agency committees and national air-pollution schemes. Even sensors that are designed for entertainment or awareness-raising need appropriate labelling to define their capabilities.

Well designed sensor experiments, that acknowledge the limitations of the technologies as well as the strengths, have the potential to simultaneously advance basic science, monitor air pollution — and bring the public along. ■

Alastair Lewis *is a science director at the National Centre for Atmospheric Science in Leeds, UK, and professor of atmospheric chemistry at the University of York, UK.* **Peter Edwards** *is a research fellow in the Wolfson Atmospheric Chemistry Laboratories at the University of York, UK.*
e-mails: ally.lewis@ncas.ac.uk; pete.edwards@york.ac.uk

1. Lim, S. S. *et al. Lancet* **380,** 2224–2260 (2012).
2. Kumar, P. *et al. Environ. Int.* **75,** 199–205 (2015).
3. Piedrahita, R. *et al. Atmos. Meas. Tech.* **7,** 3325–3336 (2014).
4. Nieuwenhuijsen, M. J. *et al. Environ. Sci. Technol.* **49,** 2977–2982 (2015).
5. Mead, M. I. *et al. Atmos. Environ.* **70,** 186–203 (2013).
6. Kamionka, M., Breuil, P. & Pijolat, C. *Sens. Actuators B Chem.* **118,** 323–327 (2006).
7. Lewis, A. C. *et al. Faraday Discuss.* http://dx.doi.org/10.1039/C5FD00201J (2015).
8. De Vito, S., Piga, M., Martinotto, L. & Di Francia, G. *Sens. Actuators B Chem.* **143,** 182–191 (2009).

Make peer review scientific

Thirty years on from the first congress on peer review, **Drummond Rennie** reflects on the improvements brought about by research into the process — and calls for more.

Peer review is touted as a demonstration of the self-critical nature of science. But it is a human system. Everybody involved brings prejudices, misunderstandings and gaps in knowledge, so no one should be surprised that peer review is often biased and inefficient. It is occasionally corrupt, sometimes a charade, an open temptation to plagiarists. Even with the best of intentions, how and whether peer review identifies high-quality science is unknown. It is, in short, unscientific.

A long time ago, scientists moved from alchemy to chemistry, from astrology to astronomy. But our reverence for peer review still often borders on mysticism. For the past three decades, I have advocated for research to improve peer review and thus the quality of the scientific literature. Here are some reflections on that winding, rocky path, and some thoughts about the road ahead.

I trained as a physician, studying the pathophysiology of exposure to high altitudes. In 1977, I became deputy editor of *The New England Journal of Medicine* (*NEJM*), working with what I assumed was a smoothly oiled peer-review system. I found myself driving an enormous machine whose operation was sometimes interrupted by startling hiccups. The first big one occurred a year after I arrived. An author who had submitted a paper to our journal accused one of our reviewers, who worked at a competing lab, of plagiarizing parts of her paper. She sent us a manuscript that her lab chief had been sent to assess for another journal, one that I could see had been typed on the same typewriter that the reviewer had used to write his review. I was told to sort it out.

This was more than a decade before a formal definition of research misconduct and systems for its investigation were established. Several careers fell apart. That of the actual plagiarist, and also that of his chief, our reviewer, who was the senior co-author of the manuscript that contained the plagiarism. Tragically, our innocent submitting author also gave up research when her accusations were rebuffed, and she was bullied and demeaned for her persistence and integrity.

This slow-motion catastrophe angered me. How common was such incompetence, confusion and corruption? Did peer review root it out — or just lob it down the road? A few years later, revelations of fabricated data in scores of papers by US cardiologist John Darsee, in *NEJM* and other journals, showed that peer review was usually helpless in detecting gross fraud. More recently, the cases of Dutch psychologist Diederik Stapel and US-based cancer researcher Anil Potti underline how easily false data continue to get through the system. Even if peer review could not detect outright

© 2016 Macmillan Publishers Limited. All rights reserved.

COMMENT

SELECTING GOOD SCIENCE

Milestones in modern peer review and reporting.

1978–79 Revelations of scientific fraud at Yale and Harvard universities publicizes the issue.

1978–92 The Oxford Database of Perinatal Trials is set up by Iain Chalmers. He later establishes the Cochrane Collaboration and its systematic analyses.

1986 Studies demonstrate publication bias in clinical trials; it is caused by the failure of trial authors to submit results for publication.

1989 Regulations defining scientific misconduct and a procedure to address allegations are codified into US law. Peer review is revealed to be ineffective against misconduct.

1989 The first Peer Review Congress held in Chicago, Illinois. It includes a trial of blinding reviewers to authors' identities.

1993 The Cochrane Collaboration, founded to review published reports relevant to health, reveals inherent biases.

1996 The CONSORT statement on reporting clinical trials is released, with a checklist to assist authors and reviewers.

1999 The *British Medical Journal* adopts open peer review on the basis of evidence from randomized trials of the practice.

2000–PRESENT Online-only journals rise in prominence along with new models of peer review.

2004 Clinical-trial pre-registration is made a condition of publication.

2006 The EQUATOR Network is founded to assemble reporting guidelines.

2010 'Beall's list' warns against 'predatory' journals with questionable peer review.

2014–PRESENT Groups (including ORCID, CASRAI, F1000 working group) are founded to support and credit reviewers.

2017 Eighth Peer Review Congress to be held in Chicago.

fabrications, could it sniff out error in honest scientific work, I wondered? There had to be a way to find out.

QUESTIONS ASKED

In 1985, an influential commentary[1] asserted that "the arbiters of rigor, quality, and innovation in scientific reports" did not "apply to their own work the standards they use in judging the work of others". Ouch! Peer review had to be studied, it said, and the most urgent need was leadership within the scientific community.

I had been working at *The Journal of the American Medical Association* (*JAMA*) since 1983. The chief editor was interested in holding a conference on peer review; I jumped at the chance. I insisted that all presentations describe research — and then worried whether we would get a single abstract.

The inaugural Peer Review Congress was held in a distinctly shabby hotel in Chicago, Illinois, in 1989. It was engaging and contentious: presenters studied the demography of reviewers at various journals, how often individuals conducted reviews, blinding, statistical reporting and much more. I was thrilled to see actual data.

A distinguished editor in the audience took another view, excoriating presentation after presentation. Finally, Iain Chalmers (who later co-founded the Cochrane Collaboration) stood and addressed him: "We have listened to your incessant criticisms of everyone who has gone to the trouble of obtaining data. What we have not heard from you is one single piece of evidence for your opinions." There was loud applause, and the future of these congresses was assured. They have taken place every four years since — in much better hotels.

Thanks to such research, we now know a great deal about the mechanics of peer review — the time taken to appraise papers, rates of disagreement between reviewers, the cost at certain journals, even the occurrence of misconduct during review.

Research has brought clear improvement to the biased reporting of clinical trials. Randomized clinical trials cost millions of dollars, are rarely repeated, and greatly influence what treatments patients receive. My colleagues and I showed that most trial results in submitted manuscripts favoured the treatment tested, and this was reflected in the results that were published[2]. Other work revealed that more than 90% of the bias was due to authors failing to submit manuscripts that are unfavourable to the treatment, and that commercial sponsorship drove decisions not to submit[3]. Although any single trial might have been conducted well, the system was skewed. Publication bias made drugs look better than they were.

This line of investigation provided evidence that convinced journals to require that clinical trials be 'pre-registered' at inception. Compliance is still patchy, but journal editors now routinely check that trials were announced publicly (typically at ClincialTrials.gov) before results were collected. We can now expect that when drugs are found to cause serious harm during the trials, the existence of those trials will no longer be hidden from the world.

Meta-research has revealed other sources of distortion. For instance, when trial reports fail to account for control patients or do not fully describe methods for randomization and blinding, they are also more likely to report exaggerated effects.

Such observations led to new standards for reporting clinical trials. An early version of the guidelines was tested in *JAMA* and produced a report that our readers found unreadable[4]. The next version of the guidelines, published in 1996 and called CONSORT (Consolidated Standards of Reporting Trials, of which I am a co-organizer), was much better accepted. These proved a highly successful model for reporting, say, epidemiologic studies, or reports of assessing clinical tests[5]. A collection of more than 300 reporting guidelines have been gathered into the EQUATOR Network (www.equator-network.org), and their use is spreading widely among biomedical researchers, journals and reviewers.

Meta-research on clinical trials has been further advanced by the Cochrane Collaboration, which systematically collects studies across disease types to weigh up the evidence. Cochrane has developed 'risk of bias' assessments to help its reviewers to evaluate possible weaknesses in trial reports.

OPEN REVIEW

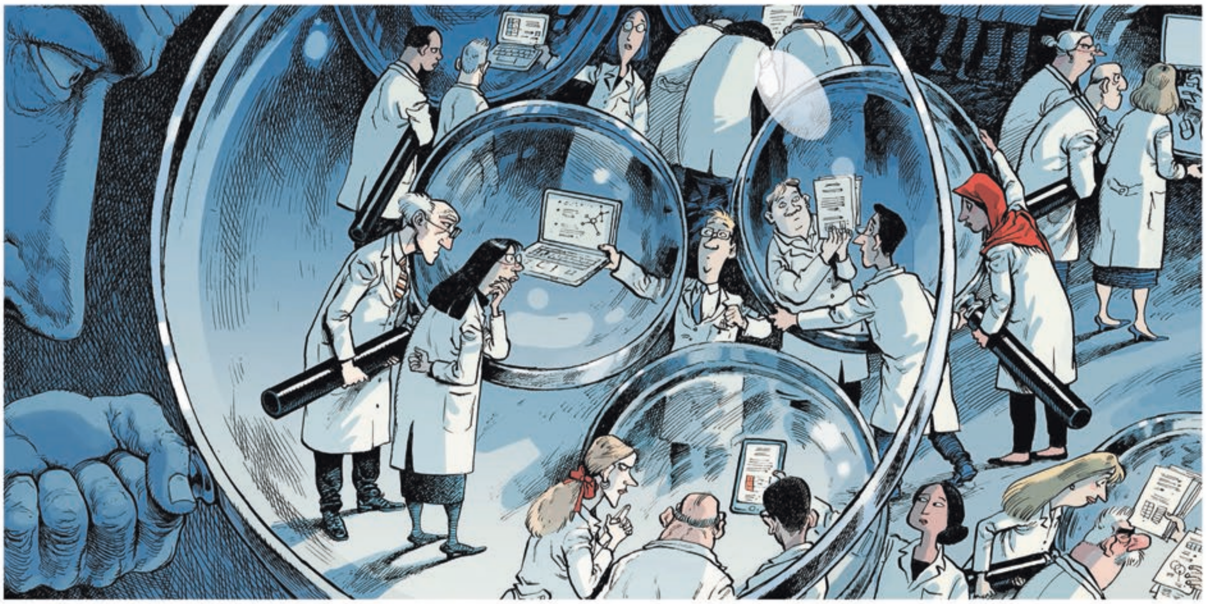
Blinding of reviews is another fertile area of study. In 1998, my colleagues and I conducted a five-journal trial[6] of double-blind peer review (neither author nor reviewer knows the identity of the other). We found no difference in the quality of reviews. What's more, attempts to mask authors' identities were often ineffective and imposed a considerable bureaucratic burden. We concluded that the only potential benefit to a (largely unsuccessful) policy of masking is the appearance, not the reality, of fairness. Since then, online technologies for blinding have increased, as have numbers of scientists (and thus the difficulty of guessing who authors may be). It will be interesting to see how similar studies work out now, and whether double-blind reviewing affects acceptance rates for women and under-represented minorities.

More than a decade ago, the *British Medical Journal* (*BMJ*) ran trials in which the identities of both author and reviewer were disclosed to each other during review, and, if the paper was published, the reviewers' names were made public. The *BMJ* did not suffer a loss of manuscripts or reviewers, and

© 2016 Macmillan Publishers Limited. All rights reserved.

COMMENT

ILLUSTRATION BY DAVID PARKINS



now makes such disclosures compulsory. Its experience suggests that how questions are posed is crucial. If a survey asks: "Would you like to sign your review?", most will decline. But if an editor says: "Our journal requires signed reviews. Will you review?", the *BMJ*'s experience is that very few will refuse[7]. I believe that this brand of open review is the most ethical variety, and its practicability is established. In the present system, authors frequently misidentify reviewers with complete confidence, so blame falls on innocent bystanders.

THE FUTURE

The past 15 years have seen an exciting surge of experimentation with new models of peer review — open, blinded, pre- and post-publication, portable and so on[8]. Some of these systems were tried and abandoned decades ago, before the Internet eased testing and logistics.

We need rigorous studies to tell us the pros and cons of these approaches today. Until then any advertised advantages of new arrangements are unsupported assertions. A 2015 survey[9] of more than 1,000 manuscripts was encouraging about the ability of review to identify important papers, but still found lapses.

After all, online technologies don't give reviewers more time or stamina. A common claim of new journals, whether legitimate or 'predatory' (those that charge fees to publish, but that do not offer standard publishing services), is rapid review and publication. This is a powerful pull for authors, but the detailed attention and mature reflection required for a constructive review takes time.

So what now? In my field, and perhaps in many others: follow the triallists. First, develop evidence-based lists of items to be included in reporting (mission-sort-of-accomplished for many clinical journals). Journals must accept and promote these guidelines and ensure that reviewers hold authors to them; perhaps they should facilitate training in peer review, which has been shown to improve performance. Finally, manuscript editors and copy editors must uphold the standards. For example, we now routinely reject trial reports that cannot prove registration before inception. This change is large for all involved — authors, reviewers and journal staff — and it is taking years.

And we must continue to study what we have done. Assessment of review is more likely now than ever before. The two-year-old Meta-Research Innovation Center (METRICS) Institute at Stanford University in California, which is devoted to researching and improving the process of science, shows that the field is maturing and gaining respect. So does last year's launch of the journal *Research Integrity and Peer Review*, a home for research on the topic.

> "We need rigorous studies to tell us the pros and cons of these approaches."

In 1986, we were lucky with our timing. The peer-review congresses came just as others were trying to see what could be learned from the literature to arrive at the best treatments for patients, developing methods for systematic review, and nailing down the biases that pervade clinical research (see 'Selecting good science'). These people did the work.

To announce that first Peer Review Congress, I wrote: "There are scarcely any bars to eventual publication. There seems to be no study too fragmented, no hypothesis too trivial, no literature citation too biased or too egotistical, no design too warped, no methodology too bungled, no presentation of results too inaccurate, too obscure, and too contradictory, no analysis too self-serving, no argument too circular, no conclusions too trifling or too unjustified, and no grammar and syntax too offensive for a paper to end up in print"[10].

Unfortunately, that statement is still true today, and I'm not just talking about predatory journals. That said, I am confident that the Peer Review Congress scheduled for 2017 will be asking more incisive, actionable questions than ever before. ■

Drummond Rennie *is a co-organizer of CONSORT, a former member of the Commission on Research Integrity for the US Public Health Service, and former president of the World Association of Medical Editors.*
e-mail: drummond.rennie@ucsf.edu

1. Bailar, J. C. & Patterson, K. *N. Engl. J. Med.* **312,** 654–657 (1985).
2. Olson, C. M. *et al. J. Am. Med. Assoc.* **287,** 2825–2828 (2002).
3. Dickersin, K. & Rennie, D. *J. Am. Med. Assoc.* **290,** 516–523 (2003).
4. Rennie, D. *J. Am. Med. Assoc.* **273,** 1054–1055 (1995).
5. Begg, C. *et al. J. Am. Med. Assoc.* **276,** 637–639 (1996).
6. Justice, A. C. *et al. J. Am. Med. Assoc.* **280,** 240–242 (1998).
7. Groves, T. *Br. Med. J.* **341,** c6424 (2010).
8. Paglione, L. D. & Lawrence, R. N. *Learn. Publ.* **28,** 309–316 (2015).
9. Siler, K., Lee, K. & Bero, L. *Proc. Natl Acad. Sci. USA* **112,** 360–365 (2015).
10. Rennie, D. *J. Am. Med. Assoc.* **256,** 2391–2392 (1986).

© 2016 Macmillan Publishers Limited. All rights reserved.

8.2 A Stronger Post-Publication Culture Is Needed for Better Science

Hilda Bastian

Bastian H (2014) A Stronger Post-Publication Culture Is Needed for Better Science. PLoS Med 11(12): e1001772. doi:https://doi.org/10.1371/journal.pmed.1001772

OPEN ACCESS Freely available online

PLOS | MEDICINE

Editorial

A Stronger Post-Publication Culture Is Needed for Better Science

Hilda Bastian*

Scientist and Editor, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, United States of America

A research report or idea needs to clamber over more than the hurdle of publication to move science, practice, or policy forward. It's not only a matter of authors waiting for kudos and citations to roll in. If their work is not to sink into oblivion, or be acted on when it shouldn't be, publication is just the beginning. Both improving research quality [1,2] and reducing waste in science [3] require a stronger post-publication culture.

Early Enlightenment science was rooted in ongoing discussion among scientists. Scientific discourse in a small, widely scattered community was in person and via books and the "erudite letters" that were the precursor of journal articles [4]. The journal system, capturing fragments of research, enabled massive expansion and acceleration of scientific activity [4].

These days the system does not keep up well with the speed of activity and the volume of research from a vast community. Articles are, by and large, too uncorrectable and unconnected [5], and much significant intellectual effort is not captured at all. Substantive discussions in journal clubs, in email lists, in social media, and at conferences are not distilled into a concise, permanent, accessible record. Most of the unaddressed content of pre-publication peer review is also lost.

Post-publication evaluation is highly fragmented. It often appears within future articles, either embedded in the introduction and discussion sections, or in formal research syntheses. Dedicated review journals (and journal sections) select, summarize, and critique publications, usually in an "expert picks" way. There are also rigorously structured systems of post-publication evaluation inside and outside journals [6,7].

There are more immediate channels to respond to published research, such as letters and comments to the editor, commentaries, and editorials in journals, and discussion in blogs. Dedicated websites have been developed for discussing and sharing research among authors [8], and PubMed Commons (for which I am editor) enables post-publication commenting and linkages by the PubMed authorship community and journal clubs [9].

Somewhere within this activity is the amorphous phenomenon that people call post-publication peer review. For some, post-publication peer review is simply shifting pre-publication peer review to after an article's release [10]. For others, it's any evaluation of an article that is similar to pre-publication peer review. Post-publication peer review overlaps with post-publication commenting, but does not encompass all of that activity.

Post-Publication Commenting

Many associate post-publication commentary with only the negative "yin" of criticism, correction, retraction, and failed replication. It is essential to prevent research-led error, harm, and futile studies. But there is a vital positive "yang" aspect, too, incorporating research aftercare [11]. Answers to questions may be critical for other studies, for adequate research assessment and synthesis, and for considering practice and policy implications [12]. Discussion can build, apply, connect, and update ideas and ongoing work.

For some, though, the success of post-publication commentary is concerned only with the "yin" of correction and retraction. For others, post-publication evaluation only "works" if it occurs for all articles, making pre-publication peer review redundant. From these perspectives, post-publication evaluation would always be shortchanged, and be seen to fall short. However, success includes rescuing important work from obscurity, and building work and capacity, not just tearing it down. Updating is at least as critical as correction to improving published research.

Furthermore, the scientific evidence base for effects of routine pre-publication peer review on article quality remains weak [13]. Pre-publication peer review can also worsen the quality of research, as when peer reviewers demand unplanned analyses of clinical trials [14]. With an oversaturation of publication in many areas, assessing it all only exacerbates the waste. Post-publication review faces the same problems.

Cultural Challenges to Post-Publication Activity

Many are wary or worse about post-publication culture. For some, any un-peer-reviewed response to peer-reviewed work is impertinent, and the Internet's removal of constraints to adding both substantive and trivial post-publication commentary to the public space is hard to accept. The Internet has also increased the quantity of incivility out in public view (Figure 1).

Disputes between scientists have always been common, and it has always been the case that "in the bitter social conflict that ensues, the standards governing behavior

Citation: Bastian H (2014) A Stronger Post-Publication Culture Is Needed for Better Science. PLoS Med 11(12): e1001772. doi:10.1371/journal.pmed.1001772

Published December 30, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: HB has received funding from the Intramural Research Program of the National Center for Biotechnology Information at the National Library of Medicine, National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: HB is a member of the Editorial Board of *PLOS Medicine*, and lead editor for PubMed Commons, a forum for scientific discourse.

* Email: hilda.bastian@nih.gov

Provenance: Commissioned; not externally peer reviewed

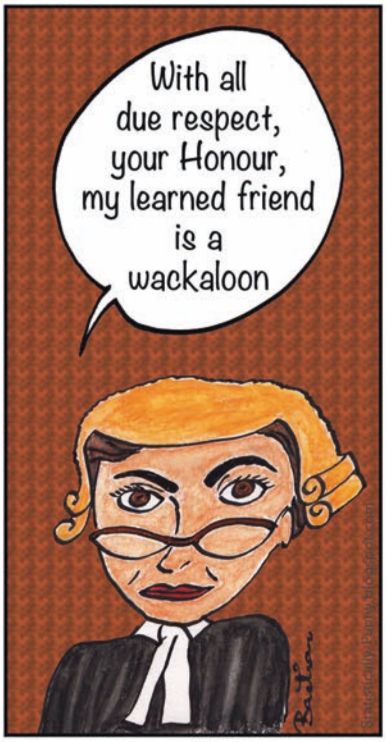


Figure 1. The melding of Internet culture and traditional communication. (Wackaloon: Internet slang for a kook; believed to derive from wacky and loon [33].)
doi:10.1371/journal.pmed.1001772.g001

deteriorate" [15]. According to sociologist Robert Merton, the example of Edmond Halley calling another astronomer a "lazy and malicious thief" in the 17th century was, and remains, more commonplace than aberrant [15]. He saw these conflicts as arising from the same "deep devotion to the advancement of knowledge" that fuels the passion for engaging in intellectual labor. We need to study and improve the way we communicate and cope with our errors and criticisms of our work.

The fear of repercussions for junior scientists in particular is high. This fear lies at the heart of the contentious issue of anonymous post-publication commenting. Some argue, though, that the risks for young scientists of openly commenting on others' work do not necessarily outweigh the advantages of visibility and recognition [16].

Even if the cost is reticence about participating, I believe the balance tips towards the requirement for transparency. Readers need to be able to judge whether writers are commenting outside their areas of expertise. Concerned readers need to have a chance of recognizing writers who have conflicts of interest, or be able to investigate whether or not potential conflicts exist.

However, addressing the obstacle to scientific progress posed by social dominance and aggression is a critical cultural issue, and not only—or even necessarily predominantly—for young researchers. Stereotype threat (anticipating discrimination) and other social issues may deter women scientists and other groups from commenting, too. Social influences can make women less talkative and less assertive than men in mixed gender groups [17], especially where "participants' concerns for self-presentation are heightened" [18].

Women scientists seem to be underrepresented in science activities that make their reflections public. In some fields and countries at least, women may still publish less [19–21], present less at conferences [22,23], and blog less [24,25]. A small body of research since the 1990s has identified some disturbingly low rates of participation by women as peer reviewers [26–28], though double-blind peer review might increase women's participation [29].

During the first year of PubMed Commons, less than 20% of those commenting were women. Research on gender bias in research and editorial peer review has been somewhat reassuring [26,27,30]. But the subject of this research has been the effect on publication fairness. The effect of under-participation on the development of confidence with the core science career skill of formulating valuable and effective critique was not considered.

I don't think that anonymity is a good solution. We need to consider skill development in critiquing research [13,31]. That may also be valuable for those who are not scientific peers, but have contributions to make [32]. Developing a much more encouraging communication climate about errors and weaknesses of scientific communication is critical. This situation reminds me of the imperative identified decades ago to create a safety and quality culture in hospitals. A mature culture of responsiveness to complaints and problem identification is as much a prerequisite for research quality improvement as it was in health care.

Rewards for substantive intellectual effort post-publication and for the aftercare of research publications and sharing of data would help. Formal recognition is also necessary to undo the perverse incentive for authors to keep important insights and additional data until a subsequent publication. Such delays can last for months, if not years.

Passive consumption of scientific papers, and the withholding of adequate information by authors, cannot advance science. Thinking and talking about our responses to research reports is still science's vibrant and compelling intellectual core. Capturing that post-publication intellectual effort more rigorously is essential for better science.

Acknowledgments

The views expressed are personal, and do not necessarily reflect those of the National Institutes of Health. My views and knowledge on post-publication evaluation have benefited greatly from ongoing discussions with David Lipman and Melissa Vaught.

Author Contributions

Wrote the first draft of the manuscript: HB. Wrote the paper: HB. ICMJE criteria for authorship read and met: HB. Agree with manuscript results and conclusions: HB.

References

1. Collins FS, Tabak LA (2014) Policy: NIH plans to enhance reproducibility. Nature 505: 612–613. Available: http://www.nature.com/news/policy-nih-plans-to-enhance-reproducibility-1.14586. Accessed 30 October 2014.
2. Ioannidis JP (2014) How to make more published research true. PLoS Med 11: e1001747. Available: http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001747. Accessed 30 October 2014.
3. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, et al. (2014) Reducing waste from incomplete or unusable reports of biomedical research. Lancet 383: 267–276. Available: http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(13)62228-X/fulltext. Accessed 30 October 2014.
4. Kronick DA (1984) Literature of the life sciences: the historical background. Bull NY Acad Med 60: 857–875. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1911798/. Accessed 30 October 2014.
5. The PLOS Medicine Editors 2013) Getting closer to a fully correctable and connected research literature. PLoS Med 10: e1001408. Available: http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1001408. Accessed 30 October 2014.
6. Haynes RB, Cotoi C, Holland J, Walters L, Wilczynski N, et al. (2006) Second-order peer review of the medical literature for practitioners. JAMA 295: 1801–1808. Available: http://jama.jamanetwork.com/article.aspx?articleid=202708. Accessed 30 October 2014.
7. NHS Centre for Reviews and Dissemination (2002) The Database of Abstracts of Reviews of Effects (DARE). Eff Matters 6: 1–4. Available: http://www.york.ac.uk/inst/crd/EM/em62.pdf. Accessed 30 October 2014.
8. Van Noorden R (2014 Mar 14) The new dilemma of online peer review: too many places to post? Nature News Blog. Available: http://blogs.

nature.com/news/2014/03/the-new-dilemma-of-online-peer-review-too-many-places-to-post.html. Accessed 30 October 2014.
9. National Center for Biotechnology Information (2014) PubMed Commons. Available: http://www.ncbi.nlm.nih.gov/pubmedcommons/. Accessed 30 October 2014.
10. Hunter J (2012) Post-publication peer review: opening up scientific conversation. Front Comput Neurosci 6: 63. Available: http://journal.frontiersin.org/Journal/10.3389/fncom.2012.00063/full. Accessed 30 October 2014.
11. Rennie D (1998) Freedom and responsibility in medical publication: setting the balance right. JAMA 280: 300–302. Available: http://jama.jamanetwork.com/article.aspx?articleid=187765. Accessed 30 October 2014.
12. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, et al. (2014) Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ 348: g1687. Available: http://www.bmj.com/content/348/bmj.g1687.long. Accessed 30 October 2014.
13. Jefferson T, Rudin M, Brodney Folse S, Davidoff F (2007) Editorial peer review for improving the quality of reports of biomedical studies. Cochrane Database Syst Rev 2007: MR000016.
14. Hopewell S, Collins GS, Boutron I, Yu LM, Cook J, et al. (2014) Impact of peer review on reports of randomized trials published in open peer review journals: retrospective before and after study. BMJ 349: g4145. Available: http://www.bmj.com/content/349/bmj.g4145.long. Accessed 30 October 2014.
15. Merton RK (1963) The ambivalence of scientists. Bull Johns Hopkins Hosp 112: 77–97.
16. Goetz A (2014) Reexamining reviewer anonymity—more costs than benefits. Open Science Collaboration. Available: http://osc.centerforopenscience.org/2014/10/22/reexamining-reviewer-anonymity/. Accessed 30 October 2014.
17. Leaper C, Robnett RD (2011) Women are more likely than men to use tentative language, aren't they? A meta-analysis testing for gender differences and moderators. Psychol Women Q 35: 129–142.
18. Leaper C, Ayres MM (2007) A meta-analytic review of gender variations in adults' language use: talkativeness, affiliative speech, and assertive speech. Pers Soc Psychol Rev 11: 328.
19. Jagsi R, Guancial EA, Worobey CC, Henault LE, Chang Y, et al. (2006) The "gender gap" in authorship of academic medical literature—a 35-year perspective. N Engl J Med 355: 281–287. Available: http://www.nejm.org/doi/full/10.1056/NEJMsa053910. Accessed 30 October 2014.
20. Kelly CD, Jennions MD (2006) The *h index* and career assessment by numbers. Trends Ecol Evol 21: 167–170.
21. Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR (2013) Bibliometrics: global gender disparities in science. Nature 504: 211–213. Available: http://www.nature.com/news/bibliometrics-global-gender-disparities-in-science-1.14321. Accessed 30 October 2014.
22. Isbell LA, Young TP, Harcourt AH (2012) Stag parties linger: continued gender bias in a female-rich scientific discipline. PLoS ONE 7: e49682. Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0049682. Accessed 30 October 2014.
23. Jones TM, Fanson KV, Lanfear R, Symonds MR, Higgie M (2014) Gender differences in conference presentations: a consequence of self-selection? PeerJ 2: e627. Available: https://peerj.com/articles/627/. Accessed 30 October 2014.
24. Shema H, Bar-Ilan J, Thelwall M (2012) Research blogs and the discussion of scholarly information. PLoS ONE 7: e35869. Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0035869. Accessed 30 October 2014.
25. Maynard A (2014 Mar 24) Gender balance in science blog networks: how has the male/female science blogger ration changed over the past three and a half years? Medium. Available: https://medium.com/2020-science-comms/gender-balance-in-science-blog-networks-71a1efb79958. Accessed 30 October 2014.
26. Gilbert JR, Williams ES, Lundberg GD (1994) Is there gender bias in JAMA's peer review process? JAMA 272: 139–142.
27. Zuber MA (2001) [Underrepresentation of women among peer reviewers and textbook authors in medicine in Germany.] Med Klin (Munich) 96: 173–180.
28. Mutz R, Bornmann L, Daniel HD (2012) Does gender matter in grant peer review? Z Psychol 220: 121–129. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3414231/. Accessed 30 October 2014.
29. Budden AE, Tregenza T, Aarssen LW, Koricheva J, Leimu R (2008) Double-blind review favours increased representation of female authors. Trends Ecol Evol 23: 4–6.
30. Grod ON, Budden AE, Tregenza, Koricheva J, Leimu R, et al. (2008) Systematic variation in reviewer practice according to country and gender in the field of ecology and evolution. PLoS ONE 3: e3202. Available: http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0003202. Accessed 30 October 2014.
31. Galipeau J, Moher D, Skidmore B, Campbell C, Hendry P, et al. (2013) Systematic review of the effectiveness of training programs in writing for scholarly publication, journal editing, and manuscript peer review (protocol). Syst Rev 2: 41. Available: http://www.systematicreviewsjournal.com/content/2/1/41. Accessed 30 October 2014.
32. Bastian H (2003) Non-peer review: consumer involvement in research review. In Godlee F, Jefferson T, editors. Peer review in health sciences, 2nd edition. London: BMJ Books. pp. 248–262.
33. Wiktionary (2014) Wackaloon. Available: https://en.wiktionary.org/w/index.php?title=wackaloon&oldid=21379171. Accessed 4 November 2014.

8.3 Reviewing Post-Publication Peer Review

Paul Knoepfler

Knoepfler, P. Reviewing post-publication peer review. *Trends in Genetics* 31(3), 221–223 (2015). © 2015 Elsevier Ltd. All rights reserved.

Reprinted with permission from Elsevier.

Scientific Life

CellPress

Reviewing post-publication peer review

Paul Knoepfler[1,2,3]

[1] Department of Cell Biology and Human Anatomy, University of California Davis School of Medicine, 4303 Tupper Hall, Davis, CA 95616, USA
[2] Genome Center, University of California Davis School of Medicine, 451 Health Sciences Drive, Davis, CA 95616, USA
[3] Institute of Pediatric Regenerative Medicine, Shriners Hospital For Children Northern California, 2425 Stockton Boulevard, Sacramento, CA 95817, USA

Post-publication peer review (PPPR) is transforming how the life sciences community evaluates published manuscripts and data. Unsurprisingly, however, PPPR is experiencing growing pains, and some elements of the process distinct from standard pre-publication review remain controversial. I discuss the rapid evolution of PPPR, its impact, and the challenges associated with it.

The rise of PPPR in the life sciences

PPPR is having a rapidly increasing impact on science. Rigorous post-publication assessment of papers is crucial for the filtering and potential integration of meritorious data into the scientific collective. It is also faster than traditional forms of evaluation. Despite this, adoption of PPPR has been relatively slow in the life sciences. As early as 2007 Todd Gibson suggested that post-publication review could be helpful [1], but it did not really catch on until recently. It now shows every sign of continuing to have a major influence on the life sciences.

This rapid growth in PPPR has been made possible by several key factors. First, although cultural acceptance within the life science community of PPPR had consistently been rather minimal for decades, it has grown substantially in the past few years, largely due to the broader, perhaps generational shift towards the Internet culture. Second, PPPR is also gaining traction because of the wider availability of popular web platforms where the review can readily take place, such as Faculty of 1000 (F1000), ResearchGate, and PubPeer, as well as blogs (Table 1). The US National Institutes of Health (NIH) is even getting into the act. PubMed Commons now allows and even encourages comments on any article in the database. Sometimes PPPR even happens in real time on social media platforms such as Twitter. Websites that are wholly or in part dedicated to PPPR are popular and influential, as evidenced by their relatively high ranking on the web, which is often similar to or higher than that of journal websites (Table 1).

Together, these factors have shifted laboratory journal club type discussions of new papers out of the confines of conference rooms into the public domain online where commentary can be rapidly disseminated and discussions with any interested individual can be facilitated. Although a quantitative assessment of the influence of this invigorated post-publication review in the life sciences is currently difficult [2], direct observations 'in the field' of the phenomenon suggest a strongly growing influence. For example, numerous article retractions and corrections have been catalyzed by PPPR, attracting the attention of journal editors, and some authors are directly responding to criticisms in the same online platforms in the public domain.

Fast and furious?

In the stem cell field there has been significant debate over so-called 'ground state pluripotency' of human cells and the role of the factor MBD3 (methyl-CpG-binding domain protein 3) in cellular reprogramming to make induced pluripotent stem cells (IPSC). Surprisingly, much of that debate has played out on PubPeer (dubbed the 'stem cell shoot out' https://pubpeer.com/topics/1/2B2B490DD36C55707411830470926D), as well as on bioRxiv, a preprint server for biology, where PPPR is occurring as well. The two main scientists involved in this debate, Jose Silva and Jacob Hanna, are engaged in an almost real-time, public PPPR and scientific interaction (http://biorxiv.org/content/early/2015/01/16/013904) that seems unprecedented in biology. Hanna has even publicly addressed specific criticisms of his papers and as a result submitted corrections to journals (https://pubpeer.com/publications/C278F3DE939616C4ADBDB9C15DB268#fb21519) only weeks or months after the issues were first raised via PPPR, demonstrating the extraordinary speed at which this process can catalyze concrete outcomes.

Another illustrative recent example of problematic issues in science being resolved strikingly fast largely via PPPR also comes from the stem cell arena in the form of the stimulus-triggered acquisition of pluripotency (STAP) cell case. In late January 2014, two papers on so-called STAP cells were published in *Nature* reporting a seemingly too good to be true method of cellular reprogramming [3,4]. On PubPeer and other sites, including my own blog, the STAP story quickly started to unravel, ultimately leading to the retraction of those papers and correction of the scientific record with an unprecedented rapidity of only a few months (Table 1) [5,6]. If the STAP cell papers had been published 5 or 10 years ago, I believe it would have taken several years for the record to be corrected. In the meantime valuable resources would have been squandered on STAP and trainee careers redirected to work on STAP could have been in serious jeopardy. Fortunately that did not happen, and I believe that PPPR deserves much of the credit.

Corresponding author: Knoepfler, P. (knoepfler@ucdavis.edu).

0168-9525/
© 2015 Elsevier Ltd. All rights reserved. http://dx.doi.org/10.1016/j.tig.2015.03.006

CrossMark

Scientific Life

Trends in Genetics May 2015, Vol. 31, No. 5

Table 1. Ranking and influence of PPPR sites and blogs

	Started	MozRank[a]	Link	Notes
F1000	2002	5.958	http://www.f1000.com	Early adopter, focused on positive reviews
Tree of Life Blog	2005	5.407	http://phylogenomics.blogspot.com/	Jonathan Eisen blog, some PPPR
RRResearch	2006	5.01	http://rrresearch.fieldofscience.com/	Rosie Redfield blog, debunked arsenic life
ResearchGate	2008	6.387	http://www.researchgate.net/	Community focused, non-anonymous
Wiring the Brain	2009	5.008	http://www.wiringthebrain.com/	Kevin Mitchell brain research-focused blog
Knoepfler Blog	2010	5.261	http://www.ipscell.com	Author's blog
PubPeer	2012	4.601	http://www.pubpeer.com	Largely anonymous post-publication review site
PubMed Commons	2013	6.718	http://www.ncbi.nlm.nih.gov/pubmedcommons/	NIH moderated venue for post-publication comments
bioRxiv	2014	5.102	http://biorxiv.org/	Pre-print server that includes PPPR
Trends in Genetics	–	4.52	http://www.cell.com/trends/genetics/home	Example reference site for MozRank

[a]The MozRank tool is an indicator of online authority and popularity in which higher numbers reflect relatively higher predicted impact. MozRank data shown are from February 2015.

Certainly, problematic life science and a corrective role for PPPR are not limited to the stem cell field. Another valuable, earlier example is the 'arsenic life' story. Scientist Felisa Wolfe-Simon at the US National Aeronautics and Space Administration (NASA) led a team reporting that they had found a microorganism that could live on arsenic instead of phosphorous. The work was eventually published in *Science* in 2011 [7]. Both in PPPR on her blog (Table 1) and in traditional publication format [8], Rosie Redfield debunked the arsenic life story in a rapid manner that limited the negative fallout from the flawed science. Even so, it is notable that the original arsenic life paper in *Science* has to date not been retracted or even corrected.

Although a clear majority of respondents to a poll I carried out on attitudes regarding PPPR was generally positive about it, a minority expressed concern over a 'gotcha' mentality (http://www.ipscell.com/2015/01/thumbs-up-for-post-pub-review-in-poll-dissenters-fault-gotcha-mentality/). Indeed, the vast majority of PPPR is negative and sometimes intensely so. In part this inclination may not be surprising given that many view it as a corrective mechanism for dealing with hyped science and inadequacies of standard peer review, particularly for high-profile papers that are perceived to have been given a 'soft' review. A potential example is the first paper on successful human therapeutic cloning, which was published in *Cell* after only a 4 day review process; it contained numerous image duplications rapidly identified on PubPeer (http://news.sciencemag.org/people-events/2013/05/cell-investigating-breakthrough-stem-cell-paper) [9].

Challenges for PPPR

A difficult issue frequently raised regarding PPPR that enables the 'gotcha' mentality that surfaces at times is the fact that the reviewers who participate are often anonymous. Although anonymity protects reviewers during both pre- and post-publication peer review from potential retaliation from authors, there is also a possible cost associated with anonymity. Some anonymous participants in PPPR feel emboldened to cross the line to engage in non-constructive criticism. In some cases PPPR comments have seemed targeted at specific individuals, and negative comments about researchers have even been sent to institutions – with negative repercussions leading to litigation against PubPeer (http://news.sciencemag.org/scientific-community/2014/10/researcher-files-lawsuit-over-anonymous-pubpeer-comments). It would be beneficial if more post-publication reviews noted the strengths of papers, and this does occur at times on blogs and on sites such as F1000, but realistically PPPR is likely to continue to be negative more often than not. The scientific community needs to consider how this inclination could limit the positive impact of PPPR and brainstorm ways to balance this culture.

These types of issues likely take place in pre-publication review as well, but in principle the fact that editors know the identity of the reviewers is a partial deterrent. In PPPR that safety net is at best incomplete, and often entirely inoperative, because commenter identities can be masked with pseudonyms and blocked IP addresses. Anonymity also can be a roadblock to fruitful give-and-take discussions between different scientists that largely depend on knowing with whom you are engaged. So-called 'sockpuppetry', where commenters are not merely pseudonymous but sometimes actively take on false identities, or even the identities of real people, has also emerged at times in anonymous PPPR and has had a negative impact. Notably, there has recently been some constructive dialogue and brainstorming about ways to manage the potential downsides to anonymity, including better moderation, comment filtering, a set of standards, and a proposed PPPR editorial board (https://pubpeer.com/publications/F2A7891E2259B6AAD71E7F5BDA1849).

An additional concern about PPPR centers around the role that unpublished data could play. Commenters might be reluctant to publicly back-up challenges to published data with unpublished data of their own for fear of being scooped, by others or by themselves. For example, it remains unclear if a journal might consider the posting of such data online to be 'prior publication'. This very real concern limits the extent of data-based give-and-take during PPPR.

The power of a new paradigm in peer review

Skeptics or outright opponents of PPPR point out that science is already self-correcting, and that scientists can comment on each other's articles via what are supposed to be relatively rapid journal-based mechanisms such as

Scientific Life Trends in Genetics May 2015, Vol. 31, No. 5

letters or similar formats. However, the reality is that such mechanisms are sometimes slow, and face their own challenges. For example, journals might be reluctant [10] to publish such responses if they challenge research that the journal has published, which might in some cases even lead to retractions, because no journal is likely to want to see increased retraction numbers. In the STAP case, a response article rebutting the original findings was submitted to *Nature* by Kenneth Lee, but the journal rejected it without clearly articulating why; it was only later published elsewhere [11]. Although there could have been many valid reasons why *Nature* rejected the Lee piece, this example is indicative of the complex interplay between multiple stakeholders that can in some cases tend to slow down this type of journal-centered post-publication communication, a limitation that is largely avoided in the dynamic interactions that post-publication review so nicely facilitates.

Rapid PPPR is here to stay, and, if anything, it is only likely to grow in influence and speed. A case has been made that, despite the hurdles remaining, PPPR will improve the quality of research and reduce waste in science [12]. I agree with that sentiment. Ultimately the goal is to make science more efficient, accurate, and reproducible. However, that does not mean that the evolution of PPPR will be painless or simple. Instead it is likely to be a fascinating rollercoaster ride with many twists awaiting us along the way. Hang on.

References

1 Gibson, T.A. (2007) Post-publication review could aid skills and quality. *Nature* 448, 408
2 Eyre-Walker, A. and Stoletzki, N. (2013) The assessment of science: the relative merits of post-publication review, the impact factor, and the number of citations. *PLoS Biol.* 11, e1001675
3 Obokata, H. *et al.* (2014) Bidirectional developmental potential in reprogrammed cells with acquired pluripotency. *Nature* 505, 676–680
4 Obokata, H. *et al.* (2014) Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature* 505, 641–647
5 Obokata, H. *et al.* (2014) Retraction: bidirectional developmental potential in reprogrammed cells with acquired pluripotency. *Nature* 511, 112
6 Obokata, H. *et al.* (2014) Retraction: stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature* 511, 112
7 Wolfe-Simon, F. *et al.* (2011) A bacterium that can grow by using arsenic instead of phosphorus. *Science* 332, 1163–1166
8 Reaves, M.L. *et al.* (2012) Absence of detectable arsenate in DNA from arsenate-grown GFAJ-1 cells. *Science* 337, 470–473
9 Tachibana, M. *et al.* (2013) Human embryonic stem cells derived by somatic cell nuclear transfer. *Cell* 153, 1228–1238
10 Macbeth, F.R. (2010) Post-publication review. A tale of woe. *BMJ* 341, c5147
11 Tang, M.K. *et al.* (2014) Transient acid treatment cannot induce neonatal somatic cells to become pluripotent stem cells. *F1000Research* 3, 102
12 Bastian, H. (2014) A stronger post-publication culture is needed for better science. *PLoS Med.* 11, e1001772

References

Haug CJ. Peer-review fraud – hacking the scientific publication process. N Engl J Med. 2015;373(25):2393–5.
Manchikanti M, Kaye AD, Boswell M, Hirsch JA. Medical journal peer review: process and bias. Pain Physician. 2015;18:E1–14.
Siler K, Lee K, Bero L. Measuring the effectiveness of scientific gatekeeping. Proc Natl Acad Sci. 2015;112(2):360–5.
Strang D, Siler K. Revising as reframing: original submissions versus published papers in administrative science quarterly, 2005 to 2009. Sociol Theory. 2015;33(1):71–96.

Additional Suggested Reading

Ferreira C, et al. The evolution of peer review as a basis for scientific publication: directional selection towards a robust discipline? Biol Rev. 2016;91(3):597–610. (*Evolution of peer review as a method of quality control reflects a cultural lag.)*
Walker R, da Silva PR. Emerging trends in peer review – a survey. Front Neurosci. 2015;9(109):1–18. (*New channels of pre- and post-publication review are described.)*
Knoepfler P. Reviewing post-publication peer review. Trends Genet. 2015;31(5):221–3. (*Post-publication review, largely stimulated by the Internet, is thriving.)*
Vercellini P, Buggio L, Vigano P, Somigliana E. Peer review in medical journals: beyond quality of reports towards transparency and public scrutiny of the process. Eur J Intern Med. 2016;31:15–9. (*A number of measures could be instituted to improve peer review, including instituting more transparency*.)