

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<sup>8</sup>.

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

**“Manufacturers and regulators need to define how and where sensors can be used.”**

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

# 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<sup>1</sup> 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<sup>2</sup>. 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<sup>3</sup>. 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 ClinicalTrials.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<sup>4</sup>. 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<sup>5</sup>. A collection of more than 300 reporting guidelines have been gathered into the EQUATOR Network ([www.equator-network.org](http://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<sup>6</sup> 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





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<sup>7</sup>. 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<sup>8</sup>. 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<sup>9</sup> 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 trialists. 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.

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”<sup>10</sup>.

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

**“We need rigorous studies to tell us the pros and cons of these approaches.”**

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).