

Peer review: a flawed process at the heart of science and journals

Richard Smith

J R Soc Med 2006;99:178–182

Richard Smith was editor of the BMJ and chief executive of the BMJ Publishing Group for 13 years. In his last year at the journal he retreated to a 15th century palazzo in Venice to write a book. The book will be published by RSM Press [www.rsmppress.co.uk], and this is the second in a series of extracts that will be published in the JRSM.

Peer review is at the heart of the processes of not just medical journals but of all of science. It is the method by which grants are allocated, papers published, academics promoted, and Nobel prizes won. Yet it is hard to define. It has until recently been unstudied. And its defects are easier to identify than its attributes. Yet it shows no sign of going away. Famously, it is compared with democracy: a system full of problems but the least worst we have.

When something is peer reviewed it is in some sense blessed. Even journalists recognize this. When the *BMJ* published a highly controversial paper that argued that a new ‘disease’, female sexual dysfunction, was in some ways being created by pharmaceutical companies, a friend who is a journalist was very excited—not least because reporting it gave him a chance to get sex onto the front page of a highly respectable but somewhat priggish newspaper (the *Financial Times*). ‘But,’ the news editor wanted to know, ‘was this paper peer reviewed?’. The implication was that if it had been it was good enough for the front page and if it had not been it was not. Well, had it been? I had read it much more carefully than I read many papers and had asked the author, who happened to be a journalist, to revise the paper and produce more evidence. But this was not peer review, even though I was a peer of the author and had reviewed the paper. Or was it? (I told my friend that it had not been peer reviewed, but it was too late to pull the story from the front page.)

WHAT IS PEER REVIEW?

My point is that peer review is impossible to define in operational terms (an operational definition is one whereby if 50 of us looked at the same process we could all agree most of the time whether or not it was peer review). Peer review is thus like poetry, love, or justice. But it is something to do with a grant application or a paper being

scrutinized by a third party—who is neither the author nor the person making a judgement on whether a grant should be given or a paper published. But who is a peer? Somebody doing exactly the same kind of research (in which case he or she is probably a direct competitor)? Somebody in the same discipline? Somebody who is an expert on methodology? And what is review? Somebody saying ‘The paper looks all right to me’, which is sadly what peer review sometimes seems to be. Or somebody pouring all over the paper, asking for raw data, repeating analyses, checking all the references, and making detailed suggestions for improvement? Such a review is vanishingly rare.

What is clear is that the forms of peer review are protean. Probably the systems of every journal and every grant giving body are different in at least some detail; and some systems are very different. There may even be some journals using the following classic system. The editor looks at the title of the paper and sends it to two friends whom the editor thinks know something about the subject. If both advise publication the editor sends it to the printers. If both advise against publication the editor rejects the paper. If the reviewers disagree the editor sends it to a third reviewer and does whatever he or she advises. This pastiche—which is not far from systems I have seen used—is little better than tossing a coin, because the level of agreement between reviewers on whether a paper should be published is little better than you’d expect by chance.¹

That is why Robbie Fox, the great 20th century editor of the *Lancet*, who was no admirer of peer review, wondered whether anybody would notice if he were to swap the piles marked ‘publish’ and ‘reject’. He also joked that the *Lancet* had a system of throwing a pile of papers down the stairs and publishing those that reached the bottom. When I was editor of the *BMJ* I was challenged by two of the cleverest researchers in Britain to publish an issue of the journal comprised only of papers that had failed peer review and see if anybody noticed. I wrote back ‘How do you know I haven’t already done it?’

DOES PEER REVIEW ‘WORK’ AND WHAT IS IT FOR?

But does peer review ‘work’ at all? A systematic review of all the available evidence on peer review concluded that

'the practice of peer review is based on faith in its effects, rather than on facts'.² But the answer to the question on whether peer review works depends on the question 'What is peer review for?'

One answer is that it is a method to select the best grant applications for funding and the best papers to publish in a journal. It is hard to test this aim because there is no agreed definition of what constitutes a good paper or a good research proposal. Plus what is peer review to be tested against? Chance? Or a much simpler process? Stephen Lock when editor of the *BMJ* conducted a study in which he alone decided which of a consecutive series of papers submitted to the journal he would publish. He then let the papers go through the usual process. There was little difference between the papers he chose and those selected after the full process of peer review.¹ This small study suggests that perhaps you do not need an elaborate process. Maybe a lone editor, thoroughly familiar with what the journal wants and knowledgeable about research methods, would be enough. But it would be a bold journal that stepped aside from the sacred path of peer review.

Another answer to the question of what is peer review for is that it is to improve the quality of papers published or research proposals that are funded. The systematic review found little evidence to support this, but again such studies are hampered by the lack of an agreed definition of a good study or a good research proposal.

Peer review might also be useful for detecting errors or fraud. At the *BMJ* we did several studies where we inserted major errors into papers that we then sent to many reviewers.^{3,4} Nobody ever spotted all of the errors. Some reviewers did not spot any, and most reviewers spotted only about a quarter. Peer review sometimes picks up fraud by chance, but generally it is not a reliable method for detecting fraud because it works on trust. A major question, which I will return to, is whether peer review and journals should cease to work on trust.

THE DEFECTS OF PEER REVIEW

So we have little evidence on the effectiveness of peer review, but we have considerable evidence on its defects. In addition to being poor at detecting gross defects and almost useless for detecting fraud it is slow, expensive, profligate of academic time, highly subjective, something of a lottery, prone to bias, and easily abused.

Slow and expensive

Many journals, even in the age of the internet, take more than a year to review and publish a paper. It is hard to get good data on the cost of peer review, particularly because reviewers are often not paid (the same, come to that, is true of many editors). Yet there is a substantial 'opportunity

cost', as economists call it, in that the time spent reviewing could be spent doing something more productive—like original research. I estimate that the average cost of peer review per paper for the *BMJ* (remembering that the journal rejected 60% without external review) was of the order of £100, whereas the cost of a paper that made it right though the system was closer to £1000.

The cost of peer review has become important because of the open access movement, which hopes to make research freely available to everybody. With the current publishing model peer review is usually 'free' to authors, and publishers make their money by charging institutions to access the material. One open access model is that authors will pay for peer review and the cost of posting their article on a website. So those offering or proposing this system have had to come up with a figure—which is currently between \$500–\$2500 per article. Those promoting the open access system calculate that at the moment the academic community pays about \$5000 for access to a peer reviewed paper. (The \$5000 is obviously paying for much more than peer review: it includes other editorial costs, distribution costs—expensive with paper—and a big chunk of profit for the publisher.) So there may be substantial financial gains to be had by academics if the model for publishing science changes.

There is an obvious irony in people charging for a process that is not proved to be effective, but that is how much the scientific community values its faith in peer review.

Inconsistent

People have a great many fantasies about peer review, and one of the most powerful is that it is a highly objective, reliable, and consistent process. I regularly received letters from authors who were upset that the *BMJ* rejected their paper and then published what they thought to be a much inferior paper on the same subject. Always they saw something underhand. They found it hard to accept that peer review is a subjective and, therefore, inconsistent process. But it is probably unreasonable to expect it to be objective and consistent. If I ask people to rank painters like Titian, Tintoretto, Bellini, Carpaccio, and Veronese, I would never expect them to come up with the same order. A scientific study submitted to a medical journal may not be as complex a work as a Tintoretto altarpiece, but it is complex. Inevitably people will take different views on its strengths, weaknesses, and importance.

So, the evidence is that if reviewers are asked to give an opinion on whether or not a paper should be published they agree only slightly more than they would be expected to agree by chance. (I am conscious that this evidence conflicts with the study of Stephen Lock showing that he alone and

the whole *BMJ* peer review process tended to reach the same decision on which papers should be published. The explanation may be that being the editor who had designed the *BMJ* process and appointed the editors and reviewers it was not surprising that they were fashioned in his image and made similar decisions.)

Sometimes the inconsistency can be laughable. Here is an example of two reviewers commenting on the same papers.

Reviewer A: 'I found this paper an extremely muddled paper with a large number of deficits'

Reviewer B: 'It is written in a clear style and would be understood by any reader'.

This—perhaps inevitable—inconsistency can make peer review something of a lottery. You submit a study to a journal. It enters a system that is effectively a black box, and then a more or less sensible answer comes out at the other end. The black box is like the roulette wheel, and the prizes and the losses can be big. For an academic, publication in a major journal like *Nature* or *Cell* is to win the jackpot.

Bias

The evidence on whether there is bias in peer review against certain sorts of authors is conflicting, but there is strong evidence of bias against women in the process of awarding grants.⁵ The most famous piece of evidence on bias against authors comes from a study by DP Peters and SJ Ceci.⁶ They took 12 studies that came from prestigious institutions that had already been published in psychology journals. They retyped the papers, made minor changes to the titles, abstracts, and introductions but changed the authors' names and institutions. They invented institutions with names like the Tri-Valley Center for Human Potential. The papers were then resubmitted to the journals that had first published them. In only three cases did the journals realize that they had already published the paper, and eight of the remaining nine were rejected—not because of lack of originality but because of poor quality. Peters and Ceci concluded that this was evidence of bias against authors from less prestigious institutions.

This is known as the Mathew effect: 'To those who have, shall be given; to those who have not shall be taken away even the little that they have'. I remember feeling the effect strongly when as a young editor I had to consider a paper submitted to the *BMJ* by Karl Popper.⁷ I was unimpressed and thought we should reject the paper. But we could not. The power of the name was too strong. So we published, and time has shown we were right to do so. The paper argued that we should pay much more attention

to error in medicine, about 20 years before many papers appeared arguing the same.

The editorial peer review process has been strongly biased against 'negative studies', i.e. studies that find an intervention does not work. It is also clear that authors often do not even bother to write up such studies. This matters because it biases the information base of medicine. It is easy to see why journals would be biased against negative studies. Journalistic values come into play. Who wants to read that a new treatment does not work? That's boring.

We became very conscious of this bias at the *BMJ*; we always tried to concentrate not on the results of a study we were considering but on the question it was asking. If the question is important and the answer valid, then it must not matter whether the answer is positive or negative. I fear, however, that bias is not so easily abolished and persists.

The *Lancet* has tried to get round the problem by agreeing to consider the protocols (plans) for studies yet to be done.⁸ If it thinks the protocol sound and if the protocol is followed, the *Lancet* will publish the final results regardless of whether they are positive or negative. Such a system also has the advantage of stopping resources being spent on poor studies. The main disadvantage is that it increases the sum of peer reviewing—because most protocols will need to be reviewed in order to get funding to perform the study.

Abuse of peer review

There are several ways to abuse the process of peer review. You can steal ideas and present them as your own, or produce an unjustly harsh review to block or at least slow down the publication of the ideas of a competitor. These have all happened. Drummond Rennie tells the story of a paper he sent, when deputy editor of the *New England Journal of Medicine*, for review to Vijay Soman.⁹ Having produced a critical review of the paper, Soman copied some of the paragraphs and submitted it to another journal, the *American Journal of Medicine*. This journal, by coincidence, sent it for review to the boss of the author of the plagiarized paper. She realized that she had been plagiarized and objected strongly. She threatened to denounce Soman but was advised against it. Eventually, however, Soman was discovered to have invented data and patients, and left the country. Rennie learnt a lesson that he never subsequently forgot but which medical authorities seem reluctant to accept: those who behave dishonestly in one way are likely to do so in other ways as well.

HOW TO IMPROVE PEER REVIEW?

The most important question with peer review is not whether to abandon it, but how to improve it. Many ideas

have been advanced to do so, and an increasing number have been tested experimentally. The options include: standardizing procedures; opening up the process; blinding reviewers to the identity of authors; reviewing protocols; training reviewers; being more rigorous in selecting and deselecting reviewers; using electronic review; rewarding reviewers; providing detailed feedback to reviewers; using more checklists; or creating professional review agencies. It might be, however, that the best response would be to adopt a very quick and light form of peer review—and then let the broader world critique the paper or even perhaps rank it in the way that Amazon asks users to rank books and CDs.

I hope that it will not seem too indulgent if I describe the far from finished journey of the *BMJ* to try and improve peer review. We tried as we went to conduct experiments rather than simply introduce changes.

The most important step on the journey was realizing that peer review could be studied just like anything else. This was the idea of Stephen Lock, my predecessor as editor, together with Drummond Rennie and John Bailar. At the time it was a radical idea, and still seems radical to some—rather like conducting experiments with God or love.

Blinding reviewers to the identity of authors

The next important step was hearing the results of a randomized trial that showed that blinding reviewers to the identity of authors improved the quality of reviews (as measured by a validated instrument).¹⁰ This trial, which was conducted by Bob McNutt, A T Evans, and Bob and Suzanne Fletcher, was important not only for its results but because it provided an experimental design for investigating peer review. Studies where you intervene and experiment allow more confident conclusions than studies where you observe without intervening.

This trial was repeated on a larger scale by the *BMJ* and by a group in the USA who conducted the study in many different journals.^{11,12} Neither study found that blinding reviewers improved the quality of reviews. These studies also showed that such blinding is difficult to achieve (because many studies include internal clues on authorship), and that reviewers could identify the authors in about a quarter to a third of cases. But even when the results were analysed by looking at only those cases where blinding was successful there was no evidence of improved quality of the review.

Opening up peer review

At this point we at the *BMJ* thought that we would change direction dramatically and begin to open up the process. We hoped that increasing the accountability would improve the quality of review. We began by conducting a

randomized trial of open review (meaning that the authors but not readers knew the identity of the reviewers) against traditional review.¹³ It had no effect on the quality of reviewers' opinions. They were neither better nor worse. We went ahead and introduced the system routinely on ethical grounds: such important judgements should be open and accountable unless there were compelling reasons why they could not be—and there were not.

Our next step was to conduct a trial of our current open system against a system whereby every document associated with peer review, together with the names of everybody involved, was posted on the *BMJ*'s website when the paper was published. Once again this intervention had no effect on the quality of the opinion. We thus planned to make posting peer review documents the next stage in opening up our peer review process, but that has not yet happened—partly because the results of the trial have not yet been published and partly because this step required various technical developments.

The final step was, in my mind, to open up the whole process and conduct it in real time on the web in front of the eyes of anybody interested. Peer review would then be transformed from a black box into an open scientific discourse. Often I found the discourse around a study was a lot more interesting than the study itself. Now that I have left I am not sure if this system will be introduced.

Training reviewers

The *BMJ* also experimented with another possible way to improve peer review—by training reviewers.⁴ It is perhaps extraordinary that there has been no formal training for such an important job. Reviewers learnt either by trial and error (without, it has to be said, very good feedback), or by working with an experienced reviewer (who might unfortunately be experienced but not very good).

Our randomized trial of training reviewers had three arms: one group got nothing; one group had a day's face-to-face training plus a CD-rom of the training; and the third group got just the CD-rom. The overall result was that training made little difference.⁴ The groups that had training did show some evidence of improvement relative to those who had no training, but we did not think that the difference was big enough to be meaningful. We cannot conclude from this that longer or better training would not be helpful. A problem with our study was that most of the reviewers had been reviewing for a long time. 'Old dogs cannot be taught new tricks', but the possibility remains that younger ones could.

TRUST IN SCIENCE AND PEER REVIEW

One difficult question is whether peer review should continue to operate on trust. Some have made small steps

beyond into the world of audit. The Food and Drug Administration in the USA reserves the right to go and look at the records and raw data of those who produce studies that are used in applications for new drugs to receive licences. Sometimes it does so. Some journals, including the *BMJ*, make it a condition of submission that the editors can ask for the raw data behind a study. We did so once or twice, only to discover that reviewing raw data is difficult, expensive, and time consuming. I cannot see journals moving beyond trust in any major way unless the whole scientific enterprise moves in that direction.

CONCLUSION

So peer review is a flawed process, full of easily identified defects with little evidence that it works. Nevertheless, it is likely to remain central to science and journals because there is no obvious alternative, and scientists and editors have a continuing belief in peer review. How odd that science should be rooted in belief.

REFERENCES

- 1 Lock S. *A Difficult Balance: Editorial Peer Review In Medicine*. London: Nuffield Provincials Hospital Trust, 1985
- 2 Jefferson T, Alderson P, Wager E, Davidoff F. Effects of editorial peer review: a systematic review. *JAMA* 2002;**287**:2784–6
- 3 Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. *JAMA* 1998;**280**:237–40
- 4 Schroter S, Black N, Evans S, Carpenter J, Godlee F, Smith R. Effects of training on quality of peer review: randomised controlled trial. *BMJ* 2004;**328**:673
- 5 Wennerås C, Wold A. Sexism and nepotism in peer-review. *Nature* 1997;**387**:341–3
- 6 Peters D, Ceci S. Peer-review practices of psychological journals: the fate of submitted articles, submitted again. *Behav Brain Sci* 1982;**5**:187–255
- 7 McIntyre N, Popper K. The critical attitude in medicine: the need for a new ethics. *BMJ* 1983;**287**:1919–23
- 8 Horton R. Pardonable revisions and protocol reviews. *Lancet* 1997;**349**:6
- 9 Rennie D. Misconduct and journal peer review. In: Godlee F, Jefferson T, eds. *Peer Review In Health Sciences*, 2nd edn. London: BMJ Books, 2003:118–29
- 10 McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. A randomized trial. *JAMA* 1990;**263**:1371–6
- 11 Justice AC, Cho MK, Winker MA, Berlin JA, Rennie D, the PEER investigators. Does masking author identity improve peer review quality: a randomised controlled trial. *JAMA* 1998;**280**:240–2
- 12 van Rooyen S, Godlee F, Evans S, Smith R, Black N. Effect of blinding and unmasking on the quality of peer review: a randomised trial. *JAMA* 1998;**280**:234–7
- 13 van Rooyen S, Godlee F, Evans S, Black N, Smith R. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *BMJ* 1999;**318**:23–7