Evidence on peer review—scientific quality control or smokescreen?

Sandra Goldbeck-Wood

Editorial by Smith
Papers p 23

BMJ, London WC1H 9JR
Sandra Goldbeck-Wood, *assistant editor*
10176.2501@compuserve.com

BMJ 1999;318:44–5

Peer review—the process by which experts advise editors on the value of scientific manuscripts submitted for publication—is traditionally surrounded by an almost religious mystique. Published papers are an important part of most assessment systems that decide how academic posts and research grants are distributed. Peer review confers legitimacy not only on scientific journals and the papers they publish but on the people who publish them. But if peer review is so central to the process by which scientific knowledge becomes canonised, it is ironic that science has little to say about whether it works.

Editors have described peer review as "indispensable for the progress of biomedical science."[1] They argue that peer review helps them distinguish between good and bad papers and between good and bad research, that it improves the presentation of what is being published, and even that it educates editors and authors.[2] When they ask reviewers to comment on a paper's scientific reliability, originality, relevance, appropriateness to the journal, and other matters, editors hope they are providing some kind of intellectual quality control, allowing the best science to be selected and improved. But is this belief more than just wishful thinking and self aggrandisement by editors and other beneficiaries of the peer review system? The question is all the more relevant because peer review is so time consuming, complex, expensive, and prone to abuse.[3]

Summary points

- Blinding reviewers to the author's identity does not usefully improve the quality of reviews
- Passing reviewers' comments to their co-reviewers has no effect on quality of review
- Reviewers aged under 40 and those trained in epidemiology or statistics wrote reviews of slightly better quality
- Appreciable bias and parochialism have been found in the peer review system
- Developing an instrument to measure manuscript quality is the greatest challenge

The evidence so far

In 1990, at the first international congress on biomedical peer review, some editors began to examine critically their own activities.[4 5] The most recent insights into what, if anything, is achieved by peer review and how it might be improved were presented at the third such congress in Prague last autumn and were brought together in the July 15 issue of *JAMA*.[6]

Before 1990, most articles on biomedical peer review reported descriptive or observational studies. Many were heavy on opinion and speculation and light on evidence. Some speculated that blinding reviewers to authors' identity ("blinding"), asking reviewers to sign their reviews ("signing"), or passing the comments of one reviewer to other reviewers ("unmasking") might improve the quality of reviews by increasing objectivity and eliminating prejudice and bias. Others said that there might be special characteristics associated with high quality reviews, such as age, seniority, holding an academic post, or having published widely. There were numerous anecdotes of biases and of abuses of the peer review system.[7 8]

Blinding, signing, and unmasking

McNutt et al were the first to use a randomised controlled trial to examine the issues of blinding and signing.[9] In their 1990 study of 127 consecutive manuscripts submitted to an American internal medicine journal, blinding increased the quality of reviews in a statistically significant way, though the improvement failed to reach their predefined threshold for administrative significance. The authors found no association between signing of reviews and review quality. Limitations of this trial were its small size, the specialist nature of the journal, the fact that reviewers were not randomly assigned to signing or not signing their reviews, the lack of a previously validated instrument for assessing review quality, and inability to exclude a "Hawthorn effect" (the possibility that reviewers' behaviour changed merely as a result of being studied).

In the most rigorous investigation of the question so far, published in the 15 July issue of *JAMA*, Van Rooyen et al were unable to confirm the effect of blinding on the quality of review.[10] They randomised 527 consecutive manuscripts submitted to the *BMJ* with regard to whether the reviewers were masked or unmasked to other reviewers, or unaware that a study was taking place. The latter group allowed a Hawthorn effect to be

excluded. Two reviewers for each manuscript were randomised to receive either a blinded or an unblinded version. But when review quality was measured with a validated instrument, neither blinding nor masking was found to make an important editorial difference to the quality of the review. A smaller trial conducted across a range of American journals found a similar lack of effect.[11] A third study conducted partly at smaller, specialist journals showed that masking, even were it to improve review quality, is not always possible, and that reviewers guess the authors' identity correctly in around 40% of cases.[12]

The most recent piece of research, published this week and from the *BMJ* (p 23), shows that making the reviewer's identity known to authors had no effect on quality.[13]

Reviewer characteristics

In a separate analysis of data from the same trial, Black et al found that the characteristics of reviewers, such as demographic factors, specialty, seniority, or academic appointments, had little association with the quality of the reviews they produced, explaining only 8% of review quality.[14] A logistic regression analysis found that training in epidemiology and statistics, and younger age, were the only characteristics significantly associated with higher quality ratings. Paradoxically, membership of an editorial board was associated with lower, not higher, review quality.

Bias

Other contributors to *JAMA's* issue on peer review illustrate the worrying number of biases by which peer review is beset, including nationality bias,[15] language bias,[16] specialty bias,[17] and perhaps even gender bias,[18] as well as the recognised bias toward the publication of positive results.[19–21]

Major challenges

For all the progress that has been made since 1990, some of the most important questions remain unanswered. The greatest challenge for peer review researchers is perhaps the quest for an instrument capable of measuring the most interesting and least accessible outcome of all—manuscript quality. Up to now, researchers have had little choice but to study the intermediate outcome of review quality; but to discover whether peer review is an effective intervention, we want to be able to trace its effects on the manuscript.

The other major challenge is obtaining funding for this new area of research, which falls outside the sphere of interest of almost all grant giving bodies. Much of the research so far has been conducted "on a shoestring," using small, one-off grants and time borrowed from researchers' other, paid commitments.

Where does this leave us?

Researchers have an interest in knowing about the fairness of the systems by which their research is judged. If the peer review process should turn out to be worthless or, worse still, hopelessly corrupt, researchers would be better off committing their findings to the internet. Meanwhile, it may be some small comfort to those who conduct research and submit papers to journals that editors, forced to grapple with the challenges of designing their own trials, are now receiving a salutary taste of their own medicine.

1 Kassirer JP, Campion EW. Peer review: crude and understudied, but indispensable. *JAMA* 1994;272:96-7.
2 Goldbeck-Wood S. What makes a good reviewer of manuscripts? *BMJ* 1998;316:86.
3 Rennie D. Peer review in Prague. *JAMA* 1998;280:214-5.
4 Bailar JC, Patterson K. Journal peer review: the need for a research agenda. *N Engl J Med* 1985;312:645-57.
5 Guarding the guardians: research on editorial peer review. Selected proceedings from the first international congress on peer review in biomedical publication. *JAMA* 1990;263:1317-441.
6 Peer review theme issue. *JAMA* 1998;280:203-306.
7 Sharp DW. What can and should be done to reduce publication bias? *JAMA* 1990;263:1390-1.
8 Chalmers TC, Frank CS, Reitman D. Minimizing the three stages of publication bias. *JAMA* 1990;263:1392-3.
9 McNutt MD, Evans A, Fletcher R, Fletcher S. The effects of blinding on the quality of peer review. *JAMA* 1990;263:1371-6.
10 Van Rooyen S, Godlee F, Evans S, Smith R, Black N. Effect of blinding and unmasking on the quality of peer review. *JAMA* 1998;280:234-7.
11 Justice AC, Cho MK, Winker M, Berlin JA, Rennie D, and the PEER investigators. Does masking author identity improve peer review quality? A randomised controlled trial. *JAMA* 1998;280:240-2.
12 Cho MK, Justice AC, Winker MA, Berlin JA, Waekerle JF, Callaham ML,et al. Masking author identity in peer review: what factors influence masking success? *JAMA* 1998;280:243-5.
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* 1998;317:23-7.
14 Black N, Van Rooyen S, Godlee F, Smith R, Evans S. What makes a good reviewer and a good review for a general medical journal? *JAMA* 1998:280:231-3.
15 Link AM. US and non-US submissions, an analysis of reviewer bias. *JAMA* 1998;280:246-7.
16 Junker CA. Adherence to published standards of reporting, a comparison of placebo-controlled trials published in English or German. *JAMA* 1998;280:247-9.
17 Joyce J, Rabe-Hesketh S, Wessely S. Reviewing the reviews: the example of chronic fatigue syndrome. *JAMA* 1998;280:264-6.
18 Dickersin K, Fredman L, Flegal KM, Scott J, Crawley B. Is there a sex bias in choosing editors? Epidemiology journals as an example. *JAMA* 1998;280:260-3.
19 Misakinan AL, Bero LA. Publication bias and research on passive smoking, comparison of published and unpublished studies. *JAMA* 1998;280:250-3.
20 Callaham ML, Wears RL, Weber EJ, Barton C, Young G. Positive-outcome bias and other limitations in the outcome of research abstracts submitted to a scientific meeting. *JAMA* 1998;280:254-7.
21 Weber EJ, Callaham ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting: why investigators fail to publish. *JAMA* 1998;280:257-9.

One hundred years ago

A military doctor

Lord Wolseley has more than once said that medical officers are not soldiers. It may perhaps interest him to know that there is at least one doctor of medicine who is as much a soldier as himself. This is General J. Frederic Canonge, now commanding the 15th French Army Corps at Marseilles. This gallant officer, who has just been promoted to the grace of Commander in the Legion of Honour, studied medicine in the early part of his career, and received his doctor's degree on the same day as he was promoted to the rank of lieutenant. His thesis was entitled: "Considerations sur l'Hygiene de l'Infanterie a l'Interieur," a subject combining in itself both his special professional interests. The thesis, which was published in Paris in 1869, bore the appropriate motto: "Miles sum, militis nihil a me alienum puto." The general is said to be very proud of his medical degree, which cost him, in his own words, "so much labour and so many sacrifices to acquire." (*BMJ* 1899;i:171)