
The evaluation of peer-review quality

Susan van Rooyen

Research Co-ordinator, BMJ, 1996–2000

© Susan van Rooyen 2001

ABSTRACT: This paper discusses the role of peer review in the decision-making processes of scientific journals. It outlines the objectives of peer review and some of the problems. It then goes on to describe research into the quality of peer review, in particular the BMJ's programme of research to date, the results obtained, and consequent changes in practice. It also looks briefly at future possibilities for research into the peer-review process.

Introduction

Peer review goes a very long way back in history; in fact its use can be traced back almost 300 years. When scientific societies were first started, in the 17th century, the prevailing attitude of scientists was one of secrecy, and wanting to keep their findings within their own small circle. Even after the advent of printing this attitude prevailed, and research findings were communicated by letter, rather than by being published in journals. One action taken to encourage a more open attitude was the external validation of the work by representatives of some prestigious organization, such as the Royal Society. This was the beginning of a system of peer review. By the middle of the 18th century this was becoming more formalized, with the then president of the Royal Society, the Earl of Macclesfield, reminding fellows that this was the only way to ensure that high standards were maintained, and he set up a committee for this purpose. This early peer-review system led to the initial rejection of one very well-known piece of work which, had that decision not later been reversed, could have changed the course of modern medicine. Jenner's paper on the first use of a vaccination against smallpox was sent for external review and, on the basis of that review, the Royal Society journal rejected the article. As a result, even though Jenner was a fellow of the Royal Society, that society was never associated with his most important discovery.¹

An account of this is given in an essay by Zuckerman and Merton,² which also quotes an often-cited letter written by Thomas Huxley concerning peer review. The editor of the *British Medical Journal* (BMJ) between 1868 and 1899, Ernest Hart, was also regularly sending articles for external review, while other journals were still making all decisions regarding publication themselves.³

Peer review is a fairly logical device, since one can argue that those who know and understand scientific research and its reporting best are those who have the greatest experience in its participation. However, despite its long and well-established use as a mechanism for editorial decision making, for much of its existence it has not itself been the subject of rigorous scientific investigation. It was over a quarter of a century ago that Ingelfinger made the following statement: 'that data on the performance of the reviewing system are lacking is all the more astounding in view of the momentous influence the system exerts on the lives of those who write biomedical articles'.⁴ However, it was not until the early 1990s that research on peer review became more formalized, and a substantial number of scientific research studies have now been reported in various journals.

A major thrust in the direction of the scientific study of the peer-review process was the first international congress on peer review which was held in Chicago in 1990 at which studies on peer review were presented, and subsequently reported in the *Journal of the American Medical Association*.⁵ Another key figure in the move towards a more rigorous investigation of peer review was Stephen Lock, editor of the *BMJ* between 1975 and 1991. His book, the first edition of which was published in 1991, shortly after the first peer-review congress, opened up the debate on peer review as to whether the system currently seen as the norm in most scientific journals is actually the fairest and most effective method of making editorial decisions.³

What is peer review?

The first task to examining the peer-review process is to define what exactly we mean by peer review, and what its purpose is. Journal editors clearly cannot be experts in every area of the subject represented by their journal. This is particularly true in the area of less specialized journals; for example, general medical journals like the *BMJ*.

Journal editors have a huge task in trying to ensure that the right research is published, and hence becomes part of the established

body of scientific literature. So, can peer review help? Clearly Sir Theodore Fox, one-time editor of the *Lancet*, was feeling very cynical about peer review when he wrote these words: 'When I divide the week's contributions into two piles – one that we are going to publish and the other that we are going to return – I wonder whether it would make any real difference to the journal or its readers if I exchanged one pile for another'.⁶

I would like to put forward the following definition of peer review as it applies to journal publication: peer review is a system by which the editor of a journal can be advised on the scientific content and reporting of a study and its appropriateness for publication by an independent individual who has particular expertise in the area reported by the study. The purpose of peer review is twofold: firstly, to distinguish between good and bad research with a view to seeing the former published and the latter rejected, and secondly, to inform the author as to how best (s)he can make improvements to the paper.

This paper will concentrate mainly on research carried out into the quality of peer review, but there are two other aspects of peer review which I would also like to consider briefly, and which may be influenced by research into peer review and consequent changes in practice. Firstly, is peer review fair? What kinds of bias exist in peer review and are there any ways in which this bias can be minimized? Secondly, what is wrong with it? Are there flaws in the process and outcome of the peer-review process? I am indebted to Stephen Lock, Fiona Godlee, and Tom Jefferson for their excellent books on the subject of peer review, both of which I would recommend to those who would like to read more on the subject.^{3,7}

Bias

It is possible, indeed highly probable, that, despite peer review, good research is still being rejected from journals. Much of this may be due to bias on the part of the peer reviewer. In the normally used system of peer review, a reviewer will know the authors of a paper when writing a review,

good research
is still being
rejected from
journals

while the identity of the reviewer is kept concealed from the author. One of the reasons given against this method of undertaking peer review is that it allows bias on the part of reviewers to go unchecked.

One of the difficulties with bias is that it is hard to show whether the basis of reviewers' recommendations derives from what we might term 'good biases', such as importance, originality, good design, and good reporting, or not. For example, any evidence for age-related bias is confounded by experience. It would seem natural that more experienced researchers perform better research, write better manuscripts, and therefore are more likely to see their work published, simply because of their experience. This shows itself in an apparent age-related bias.

There is also evidence of gender bias. However, this is not easy to determine, since research is male-dominated and, in the bio-medical field at least, there are many more male peer reviewers than female.

Institutional bias is another potential source of bias, and there is disagreement as to whether or not this is a problem. Since highly qualified and experienced researchers tend to be at high status institutions, one argument is that the prestige of an institution is a valid criterion for editors to use in their decision-making processes.

Conflicts of interest may also contribute to bias shown by reviewers for or against particular papers or authors. It is for that reason that the majority of journals ask reviewers to declare any conflict of interest when submitting a review of a paper, and if the conflict of interest is sufficiently great, it is questionable whether the particular reviewer is appropriate for that paper. Examples of such conflicts of interest are: (a) a paper submitted by a research group, and a reviewer who is a member of another research group working within the same agency; (b) a reviewer asked to comment on the work of a rival institution, or whose explanation of the subject of the research differs from those of the authors; and (c) a reviewer is asked to review the work of an author with whom (s)he has worked on another, related study, and may be a named author on that related paper.

Other sources of bias are geographical (favouring papers originating from the same country as the journal), papers reporting positive findings, and innovative ideas. It is far from clear how much the editorial decision-making process is affected by any of these biases, although there has been some research. In an ideal world a peer reviewer's report will be an objective critique of a researcher's work. However, it is very difficult to see how any peer reviewer can completely avoid bringing with them into their work of peer review their own subjective biases for or against certain characteristics of authors or institutions.

Incorrect reviews

Another problem is that the reviewer, although an expert in their field, may misunderstand the objectives of the research being reported, and may therefore write an ill-informed report. This is a particular problem when highly innovative work is being reported for which there is no precedent in published studies. A reviewer may be giving his opinion on a very new piece of work from the perspective of the established paradigm, where the work being reported may represent a paradigm shift. The reviewer, consciously or unconsciously, may reflect a bias towards established understanding.

From time to time editors are contacted by the author of a paper which has been rejected, complaining that the peer reviewer has missed the point of the paper completely. Whilst it is clear that authors are likely to be biased concerning their own work, in many cases the complaint of the author may be justified. One of the difficulties we have encountered in our own research on peer review is that, in measuring the quality of a peer reviewer's report, it is not possible to assess the subjective opinions of a reviewer, although it is possible to define the aspects of the review which it is desirable and helpful for the reviewer to have commented upon. It is therefore possible to write a review which is very helpful to the editor in terms of the extent to which the reviewer has commented on these aspects, but which, nevertheless, is

the majority of journals ask reviewers to declare any conflict of interest

totally flawed in terms of its reporting of the research.

Peer-review quality

It is against this background that the *BMJ* started its research programme into peer review. The remainder of this paper will focus on this research, describing the programme to date, the results obtained, and consequent changes in peer-review practice.

In 1996, following the First and Second International Congresses on Peer Review in Biomedical Publication, held in Chicago in 1989 and 1993, the *BMJ* obtained a research grant to conduct a randomized controlled trial (RCT) on the quality of peer review. Evaluating the quality of the manuscript itself is a huge task, and instruments for this purpose are only in the early stages of development. Of necessity each instrument can only be used on papers reporting one type of study, whereas there are many different types of study reported in papers submitted to scientific journals. For this reason, much research on peer review has concentrated on defining and evaluating review quality, and examining the effect of particular interventions on quality.

The Review Quality Instrument

In order to evaluate the quality of reviews we went about the task of developing an instrument for the purpose and validating it. A good review will assist editors in their decision making and will also assist authors in improving their paper. This dynamic process should result in better quality papers being published. In order to develop our instrument, we needed first to define what a good quality review should contain, and then to summarize these in a number of clearly defined questions. The result was the Review Quality Instrument (RQI).⁸ It can be used by either editors or authors to evaluate the quality of seven aspects of a peer reviewer's report on a Likert scale of 1–5 (Table 1). An overall measure of the quality of a review is then obtained by taking the mean of the scores for these seven questions.

The RQI does have some limitations; in particular, although it can measure the quality of the comments the reviewer has

Table 1 Aspects of a review evaluated by the RQI

Question 1	Did the reviewer discuss the importance of the research question?
Question 2	Did the reviewer discuss the originality of the paper?
Question 3	Did the reviewer clearly identify the strengths and weaknesses of the method (study design, data collection, and data analysis)?
Question 4	Did the reviewer make specific useful comments on the writing, organization, tables, and figures of the manuscript?
Question 5	Were the reviewer's comments constructive?
Question 6	Did the reviewer supply appropriate evidence using examples from the paper to substantiate their comments?
Question 7	Did the reviewer comment on the author's interpretation of the results?

Each question is scored on a scale of 1 (poor) to 5 (excellent).

made, it cannot assess the accuracy of those comments. It can therefore do nothing to eradicate this problem to which I have already referred. However, by using the RQI to evaluate review quality, it has been possible to study different methods of undertaking peer review and the effect of these interventions on the quality of reviews.

Blinding and unmasking

Blinding (removal of authors' details from a paper) is seen as a way of removing or reducing author-related and institution-related bias. The first study the *BMJ* carried out was a large RCT to examine the effects of both blinding and unmasking (the revealing of the identity of a reviewer, in this study to a co-reviewer).⁹ The objectives were two-fold: firstly, was blinding feasible? Previous work had indicated it was both desirable¹⁰ and feasible¹¹, and we wanted to see if we found the same result. Secondly, did either blinding or unmasking make any difference to review quality?

Papers were randomly allocated to be reviewed masked (pairs of reviewers had their reviews exchanged anonymously), unmasked (pairs of reviewers had their reviews exchanged with their names on), and uninformed (reviewers did not know they were the subject of a study). Each paper

*A good review
will assist
editors in their
decision making*

was sent to two reviewers. In the intervention groups (masked and unmasked), one reviewer was blinded to the identity of the authors and the other was not. Reviewers who were blinded were asked if they knew who wrote the paper and, if so, to identify them (name and/or institution). Reviewers were also asked how long it took them to write their review and what their recommendation was regarding publication.

We found that blinding was feasible, though not universally possible. We had a success rate of 58% in blinding reviewers, much the same as other studies have found. There was no difference in quality between blinded and unblinded reviewers. Unmasked reviewers produced slightly better quality reviews than masked reviewers, which tended to indicate that the removal of the cloak of anonymity from reviewers could lead to better quality reviewing. The results of this study did not indicate any change of policy was desirable. A multi-journal study, carried out at about the same time in the USA, obtained similar results.¹² We also looked at demographic information about reviewers to see if there were any obvious characteristics of good reviewers.¹³ However, we were only able to predict 8% of the possible variation in review quality, with younger reviewers producing better reviews and also reviewers who have postgraduate training in epidemiology or statistics. However, gender was not a predictor of better quality, and those on editorial boards of journals tended to produce poorer quality reviews.

However, there was uneasiness over the status quo as it related to *BMJ* procedures for peer review. Reviewing was still carried out in a somewhat lopsided way – the reviewer knew the identity of the author, but the author did not know the identity of the reviewer. Fairness seemed to indicate that peer review should either be totally blind or totally open.

Open peer review

We therefore set out to do a second study, this time into the effect on review quality of revealing the identity of reviewers to authors (so-called open peer review).¹⁴

Small questionnaire studies had indicated to us that we might encounter an unacceptable level of reviewers declining to review under such a system, but in reality we found otherwise. It seemed that people were more prepared to declare their unwillingness to review under particular conditions when they were not actually required to do so at the time. When faced with the actual event, many fewer declined than we expected. Twenty-three per cent of anonymous reviewers declined, as did 35% of identified reviewers. The majority declining in both groups stated that they did so because they were too busy to review on that occasion. The difference of 12% was statistically significant, though only marginally so. More identified reviewers than anonymous reviewers recommended publication (with or without revision).

We also sent out a questionnaire to find out whether authors would be more or less likely to submit to our journal if we moved to a policy of open peer review. We found that just over one-half were in favour of open peer review against just over one-quarter who were against. Two-thirds said it would make no difference to where they submitted their manuscripts, with almost one-third indicating they would be more likely to submit to the *BMJ* under such a policy. Those who would be less likely to submit in future were very few (4%).

The results showed that open peer review made no difference to review quality, and in particular, there was no reduction in quality. Open reviewers were more likely to decline to review. The implications of this might be more serious in a small journal with a much smaller pool of potential reviewers. Open reviewers were also more likely to recommend publication. One finding which was hardly surprising was that reviews which recommended publication were judged to be of higher quality by authors, but not by editors.

Change of policy

Since open peer review had no effect on the quality of peer review (and in particular there was no reduction in review quality), and the number of reviewers declining to

open peer review made no difference to review quality

review was not increased to an unacceptably high level, the way was open to look at open peer review as a realistic option.

A decision was made, and as of 1 January 1999, the *BMJ* has had a routine policy of open peer review. Reviewers are free to decline to review individual papers, and on occasions they exercise this right. However, the number of reviewers declining to review openly is very small, and most reviewers who decline to review still do so because they are too busy, or unavailable at the time. Peer reviewers have many other calls on their time in addition to reviewing (generally without payment), and it is to be expected that many will have to decline on occasion.

On the whole this change of policy has been well received by reviewers, with some being highly enthusiastic. There has been much debate, both inside the journal and outside, in scientific meetings and within such organizations as Locknet (the international network for research into preparation, publication, and dissemination of health research) to which I have referred in a previous article in this journal.¹⁵ Some reviewers are vehemently against the open peer-review policy. They fear that it may bring about acrimonious exchanges between authors and reviewers, and destroy good relationships between those working in the same field of study. Some concern was expressed about such 'adverse events'. As a result, a system of reporting adverse events has been introduced in the *BMJ*. Reviewers and authors are asked to notify the editor, in confidence, of any adverse event occurring as a direct result of the open peer-review policy. To date we have had only one reported to us.

It is hoped that the existence of such a system will allay the fears of some. However, peer reviewers are not always consistent in their views. Many reviewers who are also researchers themselves, and as such understand the need for evidence-based science, nevertheless decline to take part in a study on the peer review process on the grounds that they don't think the intervention being studied is 'a good idea'. I find it surprising that researchers who are vitally interested in objective scientific study should base their consent on such a subjective viewpoint.

Looking to the future

The *BMJ* research programme is ongoing, we have already embarked on our next study, which examines the effect on review quality of posting reviewers' reports onto the *BMJ* website, and we have plans for further studies, depending on the results we find. However, the more we research into peer review, the less confident we feel that it is a truly objective process, and the more potential flaws become evident. It might be the best we have at the moment, but is it really the best and most objective way to make the editorial decisions that journals face every day? If the answer to that question is 'no', what other methods can we use?

Despite the interventions we have made during the two studies completed to date, we have failed to find any resulting improvement in review quality. It would be reasonable to expect that, if a reviewer knows other people are going to read his or her review and that they will be identified, to whomever they are identified, the reviewer would put more effort into writing a good quality review. The fact that we have failed to find any difference could have one very important interpretation – reviewers are already putting as much effort as they can in producing the best quality review they can write. Perhaps they are not able to produce a better one.

We have had plans to conduct some research on training peer reviewers, and this viewpoint makes the need for such research even more pressing. We need to find ways of procuring a true improvement in review quality. Electronic advancements and the World Wide Web make it possible for there to be a dialogue between authors, reviewers, and editors on a real-time basis. Could a combination of these two methods into a more dynamic process produce a true improvement in review quality, and thence in manuscript quality? An answer to such questions requires the means to evaluate manuscript quality, and to record the changes made in a manuscript during revision. Electronic storage of manuscripts makes the latter feasible. It will require considerable research before we are able to evaluate the quality of all types of

*We need to
find ways of
procuring a
true
improvement
in review
quality*

manuscripts, since the criteria for a good manuscript will vary with different study designs.

Peer-review research has, in the main, been conducted in larger journals, many of a more general nature. In particular, our own studies have been conducted in a large general medical journal, and only in the UK. Other large studies have been conducted in the USA, but again in larger journals. It is important to know if the results obtained can be generalized to smaller or more specialist journals. It is also important to know if the RQI can be used in other disciplines besides biomedicine. For that reason, we would be very pleased to see others carrying out research in their journals, either individual journals or groups of smaller journals together. This will improve confidence in the RQI and extend the body of peer-review research outside the bounds in which it exists at present. Those interested in research into peer review and editorial decision making, whether actively involved or not, are invited to join Lockett. This network is not confined to the biomedical field, but is open to anyone who is interested in this kind of research. Information relating to this and other relevant matters can be found on the website of WAME (the World Association of Medical Editors) at www.wame.org.

It appears that the total demise of peer review is not in sight, even though changes in methods of undertaking peer review are likely as a result of more research in the area. At least any changes are more likely to be based on an objective scientific evaluation of the process, rather than because of the particular preferences (or dare one say bias?) of the journal editor. This seems to me to be an entirely good thing.

However, research into peer review is not common knowledge among scientific editors. Until this research is more widely known, any changes in process as a consequence of research is likely to have an

extremely limited impact on the world of scientific journalism at large.

References

1. Booth, C. C. Medical communication: the old and the new. The development of medical journals in Britain. *BMJ* 1982;285,105-8.
2. Zuckerman, H. and Merton, R. K. Patterns of evaluation in science: institutionalism, structure and functions of the referee system. *Minerva* 1971;9, 66-100.
3. Lock, S. A *Difficult Balance – Editorial Peer Review in Medicine*. London: BMJ, 1991.
4. Ingelfinger, F. J. Peer review in biomedical publication. *American Journal of Medicine* 1974;56, 686-92.
5. Guarding the guardians: research on editorial peer review: selected proceedings from the First International Congress on Peer Review in Biomedical Publication. *Journal of the American Medical Association* 1997;278, 1317-1441.
6. Fox, T. F. *Crisis in Communication*. London: Athlone Press, 1965.
7. Godlee, F. and Jefferson, T. *Peer Review in Health Sciences*. London: BMJ, 1999.
8. van Rooyen, S., Black, N. and Godlee, F. Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. *Journal of Clinical Epidemiology* 1999;52, 625-9.
9. van Rooyen, S., Godlee, F., Smith, R., Evans, S. and Black, N. The effect of blinding and unmasking on the quality of peer review: a randomized trial. *Journal of the American Medical Association* 1998;280, 234-7.
10. Fletcher, R. H. and Fletcher, S. W. Evidence for the effectiveness of peer review. *Science and Engineering Ethics* 1997;3, 35-43.
11. McNutt, R. A., Evans, A. T., Fletcher, R. H. and Fletcher, S. W. The effects of blinding on the quality of peer review. *Journal of the American Medical Association* 1990;263, 371-6.
12. Justice, A. C. *et al.* Does masking author identity improve peer review quality? A randomized controlled trial. *Journal of the American Medical Association* 1998;280, 240-2.
13. Black, N., van Rooyen, S., Godlee, F., Smith, R. and Evans, S. What makes a good reviewer and a good review for a general medical journal? *Journal of the American Medical Association* 1998;280, 231-4.
14. van Rooyen, S., Godlee, F., Evans, S., Black, N. and Smith, R. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *BMJ* 1999;318, 23-7.
15. van Rooyen, S. A critical examination of the peer review process. *Learned Publishing* 1998;11, 85-91.

Susan van Rooyen,
6732 SW BriarMeade Lane,
Topeka, KS 66619, USA
Email: svanrooyen@kscable.com

research into
peer review is
not common
knowledge
among
scientific editors