

Peer Review: A Vital Ingredient

Jane Smith,
British Medical Journal

Sir Theodore Fox, a past great editor of the *Lancet*, Britain's other weekly general medical journal, once said:

"When I divide the week's contribution 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 whether I exchanged one pile for another."

Peer review - indeed editing in general - is about making sure that the answer to Fox's question is "Yes, it would make a difference" - though there is an element of truth in what Fox suggests. That element has been seized upon. Some librarians in the States, faced, like most librarians, with ever increasing demands on ever tighter budgets, have apparently started to challenge publishers of journals over their peer review practices. In fact - as always - they are really challenging them over their prices: but the challenge goes like this:

"You say that one of the reasons your journals cost so much is because of rigorous peer review. But we know that peer review is not infallible - look at all those cases of fraud there have been; you publish papers that are then knocked down in your own correspondence columns; you publish a paper one year and something refuting it the next. Peer review's not so hot, so why not drop it - it would cost you less and therefore cost us less."

That is, I hope, a travesty of their position - if only because it shows a misunderstanding of scientific and scholarly inquiry. But I want to defend peer review against accusations like that. Nevertheless, we do have to take account of our mythical librarian's observations, and there are many caveats in my defence of peer review.

But let me start by considering the context for peer review, and in particular the challenges to traditional publishing of scholarly journals - the challenges of technology, of the needs of the Third World, of the need to show that our activities are cost effective.

I'd like to start from the position that publishing as a whole is an intellectual activity, such that the medium of publishing, whether print on paper or digital characters on computer databases, is almost irrelevant. In STM publishing - and specifically

medical publishing - the term "publishing" encompasses notions of quality and expertise. Scholarly publishing is partly about members of scholarly communities communicating with each other, but it should be more than simply that. The communications will go on irrespective of the activities of publishers. They will go on via the traditional means of invisible colleges - at meetings, conferences, on visits, through preprints - and via some new ones as well such as electronic networks and bulletin boards.

Refereeing (or peer review) - the process of asking an expert for an opinion on a piece of work - has a longish if somewhat chequered

What publishers offer over and above that informal communication sounds remarkably like the original aims of peer review as expounded by the seventeenth century scientific societies that invented it. Publication after peer review, they claimed, offers the producer some validation of his work and protection from plagiarism, and for consumers it offers some assurance that the work is authentic. Also peer review can't be looked at in isolation. Although important it is only one element of the publishing process. You realise why it is not the only element if you think who referees are: they are also our authors, our editorial writers, and our sternest readers. In some small disciplines and journals the match between these three groups is almost absolute. Even in a very general journal like ours the match is fairly close.

Refereeing (or peer review) - the process of asking an expert for an opinion on a piece of work - has a longish if somewhat chequered history. Back in the 17th century the earliest scientific journals - the *Journal des Scavans* in Paris and the *Philosophical Transactions* of the Royal Society here in London - used the expertise and knowledge

of leading members of their scientific communities to assess whether communications should be published in their journals.

Since then peer review has had an erratic history, and the role of the strong individualist editor making decisions for himself has persisted well into this century. John Burnham, a United States historian, has shown that in medicine at least it is only since the second world war that peer review has gone from strength to strength - so much so that it is now entwined into the structure of scientific work and communication. Peer review is now central to getting grants to do the work in the first place, to getting jobs and promotion, and, of course, to getting published. Few serious scientific journals now operate without it. Possibly it is its status as a sacred cow that has led to it being criticised so bitterly.

So why has peer review become so important; and why do we as editors seek the help of referees? John Ziman, now professor of the history of science at Imperial College, restated the aims of peer review as follows: to prevent the publication of bad work; to improve scholarship by ensuring that the relevant literature is cited; to improve the language and presentation of data.

Equally important and rather closer to home is the fact that an editor can't be expected to be an expert in every aspect of his or her discipline. That is particularly so in these days of specialisation and subspecialisation. Inevitably therefore faced with a paper that strikes an editor as being interesting he or she will want to know whether the work is original, the question important, the methods scientifically reliable, and the results and interpretation reasonable. Who else to ask but an expert?

In practice I think that editors often underestimate their ability to assess originality - as editors they see a broader range of material than the average referee and are often well able to judge originality and indeed importance. Where they do need referees is to tell them whether a paper is scientifically reliable: are the implicit assumptions correct, do the methods stand up? That is the validation they need.

But how effective is peer review as a means of validating work? After all we know that bad science does get published and that much more sinks without trace. There are many criticisms of peer review to add to those of our mythical librarian.

Firstly, it introduces delay, though this is an overstated criticism: getting it right is better than getting it fast. Secondly, referees do not agree among themselves so how can they validate work? Thirdly, referees are accused of being biased against unknown authors and towards the well

known, and, perhaps more damagingly, of being biased against new ideas that challenge the prevailing orthodoxy. Fourthly, in disciplines like medicine most full papers will be published somewhere eventually, so what function is a referee performing? And, finally, there are still errors in published work despite papers having been seen by a referee - ranging from duplicate publication to statistical errors and completely fraudulent work.

The problem with many of these criticisms is that there is not much solid data. We can all point to anecdotal cases but we do not know what the denominator is. And the problems may well differ between disciplines.

The main problem is that for such a central activity we don't know much about it. Peer review has gained its sacred cow status on the basis of little evidence. This is probably partly because of its seeming psychological inevitability, but editors are now starting to look more closely into the process and outcome of peer review. One example of this was the first of a series of international conferences into peer review in biomedicine hosted by the Journal of the American Medical Association in 1989, which aimed at getting editors and others to do some research into their practices. The US conference raised some interesting research issues, but firstly, I want to outline what we do at the British Medical Journal. Because we are a general medical journal publishing a very broad variety of material we start from the premise that we couldn't cope without expert input from referees. The task then is to get the best out of our referees and out of the relationship between editor and referee. We owe it to our readers to do the best we can and not to waste their time on papers that don't deserve it. And we'd be doing this sort of thing whether the BMJ is published on paper or in electronic form.

We have some data on the worth of what we do - much of it from a study started in 1979 by Stephen Lock, the then editor. This was designed to assess the role played by referees, to measure consensus between the people who read the papers - editors and referees - to see what happened to papers we rejected, and to see what happened to those we accepted. We analysed a consecutive series of 1551 papers received by the journal over four or five months.

Like many general medical journals we have a high rejection rate: it was 80% at the time of that study and it has now risen to nearer 90%. Papers are initially screened in the office by our full time editors, and only half are sent out to a referee. The other half are rejected in the office after being read by at least two editors. Once the paper is back from a referee - assuming the referee's report is favourable the final decision is made by an editorial committee.

The problem in studies of peer review is what to use as a gold standard: how do you assess whether the referee's recommendation is right? We used two measures: one was consensus between the various people who read a paper. Previous writers had argued that consensus among referees in medical journals was no better than chance, suggesting that the whole process is fairly arbitrary. But it was not what we found; in general our rates of agreement between editors and referees were much better than chance.

The other major criticism of outcome was citation analysis - both for journals, as measured by their impact factor (a measure of the average number of times papers published in a particular journal are cited), and by the citation for individual articles rejected and accepted by the BMJ.

The BMJ itself has a fairly high impact factor. If we look at the original 1551 articles and ask how many of them were eventually published in a high impact factor journal - whether the BMJ or any other - then we find that 458 were. 328 of these were in the BMJ itself, so we "captured" 4/5ths of these high quality articles.

Moreover, as the articles went through the stages of judgement - in the office, by a referee, by the editorial committee - we filtered out the worse papers and got a better and better subset as we went along. And it is here I think that there is an element of truth in Fox's remark that I cited at the beginning. By the time they have reached the editorial committee the papers have been thoroughly reviewed and it may well be that swapping the piles of accepted and rejected papers at that stage may not make that much difference - because it is here that editorial judgements come into play. Nor, for the same reason, might it matter that the Lancet ends up publishing a paper we reject - or vice versa - as happens most weeks.

The other aspect that emerged very clearly was the amount of change that authors produced in their work before we published it. Eighty percent of the articles the BMJ accepted were revised before publication, over half of them on scientific grounds. That proportion is probably higher now since we rarely accept a paper without asking for some changes.

Since then too we have tried to be more explicit in what we ask of our referees, prompting them to hone in on the key issues that we as editors need to know. With our statistical referees we now have a very structured checklist of items that tests whether a study is well designed and well analysed. There are also other things we can do that indirectly increase our chances of getting a better referee's judgement and a better paper. There has been a move in clinical medicine towards using structured abstracts. Instead of leaving the quality more or less

entirely to the skill of the author and the subeditor a structured abstract requires authors to provide certain key data about the design of the study and its key results. Not only does it force authors to be more explicit about what they did; it also helps editors and referees in evaluating the paper.

Let me return now to the Chicago conference. Much of the research it generated was small scale and descriptive rather than analytical but it did raise the sorts of issues that we as editors need to address. For example, one very relevant study was performed by the BMJ's statistical adviser, Martin Gardner, showing the effectiveness of statistical review. He got an independent assessor to look at 45 published papers that had been seen initially by our statistical referee; the assessor scored them against the checklist and then compared them with the original submissions. On submission our referees had found only 5 of the 45 papers statistically acceptable for publication. By the time they were published, however, 38 were thought to be acceptable. Something clearly went wrong with the remaining seven, but these findings do seem to show that statistical refereeing was having some effect.

Secondly, the problem of publication bias was discussed. This is really the issue of newsworthiness. If you publish a well constructed trial that shows that the pill doesn't increase the risk of breast cancer that's not news. There are some data that show that authors are the biggest influence on this in that they don't bother to write up and submit their negative results. But this is an important issue because if the question is worth asking then the answer is valuable, if only to ensure that others don't waste their time repeating research. One way that has been suggested of getting round this problem is for editors and referees to assess papers on the basis of their protocols or methods section - before they know the results. This issue of negative data also illustrates more broadly my theme about quality: of adding knowledge to information.

Iain Chalmers is an academic obstetrician in Oxford, who has recently published the Oxford Database of Perinatal Trials. He and his colleagues put an enormous amount of effort into collecting both published and unpublished trials in perinatal medicine and then did a series of meta-analyses on them. Meta-analysis is a technique for pooling data from clinical trials that may otherwise be too statistically weak to show a result one way or the other. That isn't refereeing as such - but it shows the role of expert analysis and effort in the publishing process. Editors will probably have to start commissioning meta-analyses and structured reviews in the same way that we now publish much looser reviews; the gainer from this should be the

reader - in our case clinical readers who want to know how they should be altering their practice. What's interesting about the Oxford database is not simply or even mainly that it's an early example of database publishing in medicine but that it's an early example of a way of extracting valuable information from a mass of unwieldy data.

I'd like to conclude by reinforcing that thought in relation to the theme of this conference. The technology seems to be going in the direction of the single article: in future a paper can go in a seamless web from the author's word processor to the reader's laser printer, and then, people will ask, why have journals? I think this probably will happen at the margins, but it won't happen in a large scale because we haven't really worked out the economics of single articles or their value system. Journals are actually quite an economical way of producing a bundle of articles. And they are

also more than bundles of articles: sometimes they convey a set of values and express a sense of community. Nevertheless, if the day does ever come when the journal dissolves and the single article reigns supreme then the buyer of that single article is going to want to be very sure that she's not buying a pig in a poke because she'll be paying a lot for it. So quality will be ever more important, which means that we editors and publishers are going to demand more of our authors, referees, review writers - and that won't come cheap. □

Editor's Note:

This paper was given at the Seventh International Learned Journals Seminar held on 5th April 1991 and is published with thanks to the author and the editor of "Learned Publishing" who released it for publication in "Serials".

Serials Administration and Management

A one-day course for junior library staff
and librarians new to working with serials

Birmingham and Midland Institute

Monday, 14th October 1991

This is a repeat of a very popular event held in June. Some places are still available. For more information please contact:

Mrs. Jill Tolson,
UKSG Administrator,
114 Woodstock Road,
Witney, Oxon OX8 6DY
Tel. 0993 703466
Fax 0993 778879

Comments from some of the delegates in June:

"The talk about publishing was excellent — much of the ground was new to me."

"Automation — good coverage. Roles of subscription agent / publisher / librarian well defined. Helpful course overall."

"It isn't often junior staff have the chance to attend such seminars. I appreciated the opportunity."