

EDITORIAL

What do we know about peer review?¹

The work of a scientist should be viewed by disinterested, objective standards. The most appropriate person to judge those standards is an appropriate peer – another scientist. Extraneous and subjective issues such as rank, jealousy, affiliation and rivalry should play no place in this process, neither should lack of expertise. This is the theory behind the practice of peer review, the process by which papers submitted to professional journals or grant applications made to funding bodies are usually judged.

For many years a curious double standard has existed about the peer review process. The purpose of peer review is to evaluate research, but no thought was given to research on peer review itself (Bailar & Patterson, 1985; Smith, 1994). There were many possible reasons for this lacuna – peer review was so clearly the best method it did not require evaluation. There was no point because no better system existed. Alternatively, like the jury system, it was felt better to maintain some mystique rather than expose the working of what might prove to be a fallible system, dependent on the goodwill of unpaid volunteers.

This has started to change. Particular landmarks have been the seminal review by Cicchetti (1991), the First and Second International Congresses on Peer Review, both published in themed issues of the *Journal of the American Medical Association*, and the founding of Locknet (Smith, 1994), an informal network of researchers interested in the process of peer review, which takes its name from Stephen Lock, former editor of the *British Medical Journal* and an earlier writer on peer review (Lock, 1985). In their recent systematic review of peer review research Pierie & Overbeke were able to identify 80 studies reporting original research, albeit nearly all retrospective (Pierie & Overbeke, 1996).

What has been established? The founder of this journal, with his tongue as usual firmly in his cheek, once quoted Adlai Stevenson on the task of the editor as being to ‘separate out the wheat from the chaff, and publish the chaff’ (Shepherd, 1986). Most of us who have experienced the pain of having their papers rejected will have blamed the reviewers for assisting the editor in this ignoble purpose. Not so, but recent research has identified several shortcomings in the peer review process.

The first is the problem of referee reliability. Several studies have looked at what reviewers consider to be important in a manuscript (Cicchetti, 1991). The answers were the importance of the contribution to the field under investigation, and the perceived adequacy of the methods used. Such findings are unsurprising, and reassuring. However, less reassuring are those studies that attempt to judge how reliably reviewers make these assessments.

Unstructured peer review in which two or more referees comment on the same manuscript has relatively low reliability. Agreement on global measures such as recommendation to accept or reject is satisfactory (although this strictly is an editorial, and not a peer review, judgement), however, assessments of study importance and methodological quality are not reliable (Ernst *et al.* 1993). Measures of agreement vary across studies, but the usual finding in the medical and behavioural science journals is of reliabilities of between 0.2 and 0.4 (Cicchetti, 1991; Goldman, 1994). Defenders of the status quo have replied that referees are not chosen randomly, and that editors are not looking for reliability from their referees (Bailar, 1991; Kiesler, 1991), but the arguments are unconvincing.

Peer review is fallible. The study by Peters & Ceci in which they substituted authors’ names and

¹ Address for correspondence: Dr Simon Wessely, Department of Psychological Medicine, King’s College School of Medicine and Dentistry, Denmark Hill, London SE5 9RS.

institutions on previously accepted papers, replacing prestigious institutions with a mythical address, is well known. Most journals rejected the resubmission from the unknown institution, and failed even to recognize that the paper had already been reviewed and accepted (Peters & Ceci, 1982). Critics have responded by pointing out that some questions from this notorious study remain unanswered, in a manner reminiscent of the unanswered criticisms of another 'revolutionary' study that also depended upon deception (Spitzer, 1975). Those working in less prestigious institutions may also be reassured by the failure of others to replicate this finding (Garfunkel *et al.* 1994).

Peer review is also onerous (Anon, 1995). Reviewers in both Britain and Scandinavia usually work for more than one journal (three to four is the median), review approximately one paper per month, and spend between 1 and 2 hours doing so (Lock & Smith, 1990; Nylenna *et al.* 1994). Reviewing is a generally unsung task – although it is certainly not unappreciated, not least by the editors. The finding that nearly half of reviewers for the *British Medical Journal* themselves were editors or served on editorial boards may explain why reviewers persist at all (Lock & Smith, 1990).

What factors influence the quality of peer review? Who makes a good reviewer? Coming from a prestigious academic institution is associated with a better quality review (Evans *et al.* 1993), but straightforward seniority is not. Indeed, there is empirical support for this author's own prejudice – that review quality is inversely proportional to the age and seniority of the reviewer (Evans *et al.* 1993). Time spent on the review is also inversely proportional to the seniority of the reviewer (Stossel, 1985; Nylenna *et al.* 1994) – those in the upper echelons of medical science are often too busy to devote proper attention to the process, or perhaps have a greater personal reputation to defend. Perhaps as psychiatrists or psychologists we should not be surprised by one further finding – referees' judgements about the scientific merits of a paper are influenced by their own preconceptions and stated positions on a topic (Ernst & Resch, 1994).

Can the quality of reviewers be increased? Structured review forms of acceptable reliability have now been developed (Feurer *et al.* 1994). Whether or not they lead to actual improvements in the eventual quality of the published manuscript is currently the subject of research. This journal uses a check list, which we have recently revised. Feedback to reviewers of the other reviewer's report acts as an internal quality audit, and is valued by reviewers. This has been our practice from the beginning of last year.

Another option is increasing the number of referees. The chances of detecting a significant flaw increase, while reliability also increases with the number of opinions (Cicchetti, 1991). The limiting step is the time and good-will of referees. *Psychological Medicine* has since its inception usually used two, and often three, reviewers, and will continue to do so.

Another area of controversy is whether or not the reviewer should be aware of the authors identity. Are reviews of higher quality when the author of the paper is anonymous? The answer is a qualified yes (McNutt *et al.* 1990; Evans *et al.* 1993). Reviewers for the *American Journal of Public Health*, where this is practiced, themselves favoured author anonymity (Yankauer, 1991). However, the technical difficulties in making most papers anonymous to an informed reviewer are considerable (Yankauer, 1991; Fisher *et al.* 1994). Simply removing the title page is rarely sufficient. The *Canadian Medical Association Journal* used blinded reviewing during the 1980s, but gave up the practice because it proved time consuming and often impossible (Squires, 1990).

Psychological Medicine allows referees to remain anonymous. This practice is almost universal, although there is little or no empirical evidence as to its benefits. The most frequently stated reason, which we share, is the need to permit more junior reviewers to criticize papers submitted by their seniors, but not necessarily better. Allowing referees to sign their reports if they so wish has been suggested (Cicchetti, 1991). However, the natural desire of referees to sign positive but not negative reviews makes this practice suspect. We have no plans to alter our current practice of maintaining referee anonymity.

The pros and cons of referee and reviewers anonymity will continue to be debated, and are currently being examined in several controlled studies. However, to our knowledge no one has addressed a simple question – can authors deduce the identity of the anonymous referee anyway? Anyone who has ever submitted a scientific paper will be familiar with the process of guesswork,

intuition, rumour and comparing type fonts in order to identify the anonymous referee. To answer this question during the first part of 1995 all those who submitted a paper to *Psychological Medicine* were asked to make a guess at the identity of the referees who commented on their paper. We thank all those who took part. The results showed that in 80% of cases the authors had no idea as to the identity of the referee, 14% did guess, but incorrectly, and only 6% were correctly identified (Wessely *et al.* 1996). Overall, referees, even in a specialist journal, remain anonymous.

Many areas remain opaque. Perhaps the greatest gap lies in the process of peer review of grant applications (Anon, 1994). It has been frequently observed that once a piece of research has been undertaken, it will eventually see the light of day if its authors are so inclined – ‘a determined author can get any rubbish printed’ (Tyrer, 1991). The only limitation to publication has been described as a lack of stamps (although this applies to the biomedical and social sciences but not, intriguingly, to physics and astronomy (Cicchetti, 1991)). When completed research projects fail to see the light of day this is usually due to lack of enthusiasm of the authors, and not persistent editorial refusals (Dickersin *et al.* 1987, 1992; Easterbrook *et al.* 1991). Peer review may prevent flawed research appearing in the better journals, including, we hope *Psychological Medicine*, but the supply of journals is inexhaustible, and if the authors have sufficient patience and perhaps an absence of pride, their efforts will eventually be rewarded somewhere. However, research that is not funded may never take place at all. The process by which grant applications are reviewed and the decisions on funding taken is of more fundamental importance to the scientific process, and is even more obscure than those of journal peer review (Horrobin, 1974; Anon, 1994). Little more is known today, but there is nothing to suggest that peer review of grant applications avoids any of the pitfalls noted in the journal peer review literature (Cole *et al.* 1981; Cicchetti, 1991). The decisions of both the National Institutes of Health and the NHS Research and Development programme to carry out original research on peer review and grant funding are timely.

The future of peer review is once again in question. The growth in electronic publication has led to suggestions that peer review prior to publication is obsolete – instead papers will attract their own peer review and commentary after publication, which will be electronically linked to the article, and may also enable the article itself to be continuously revised and improved (LaPorte *et al.* 1995). Publication delay, inevitable with conventional journals, will disappear, as will the irritating comment ‘data not shown’. The editorial mantra ‘please shorten’ may also disappear. Whether or not the resulting publications will benefit is a moot point. Furthermore, as Judson has pointed out, limitless space will not reduce the more pertinent restriction of ‘Data not read’ (Judson, 1994).

Prognostications such as the death of the journal have been equally energetically refuted by others who see that conventional paper publication in leading journals will remain the preferred method of scientific communication (Cicchetti, 1993; Gellert *et al.* 1995). We continue to see a major role for the printed academic journal, not least because of the difficulties in reading other forms of publication on the plane or in the bath. More seriously, we remain committed to the role of peer scrutiny in maintaining the quality of science. Opponents of peer review often point to those instances in which the conservative nature of peer review has stifled innovation (Horrobin, 1990). While such instances happen, they should be balanced against the many examples in which peer review has protected the public from the malign effects of bad science.

It is tempting to conclude with Churchill’s famous remark on democracy. Both are flawed and inefficient systems – and both are better than any alternatives.

SIMON WESSELY

I am grateful to Iain Chalmers, Roger Jones and Eugene Paykel for helpful comments.

REFERENCES

- Anon (1994). *Peer Review: Reforms Needed to Ensure Fairness in Federal Agency Grant Selection*. United States General Accounting Office: Washington, DC.
- Anon (1995). *Peer Review: An Assessment of Recent Developments*. Royal Society of London: London.
- Bailar, J. (1991). Reliability, fairness, objectivity and other inappropriate goals in peer-review. *Behavioral and Brain Sciences* **14**, 137.
- Bailar, J. & Patterson, K. (1985). Journal peer review; the need for a research agenda. *New England Journal of Medicine* **312**, 654–657.
- Cicchetti, D. (1991). The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation. *Behavioral and Brain Sciences* **14**, 119–186.
- Cicchetti, D. (1993). The reliability of peer-review for manuscript and grant submissions – its like *déjà-vu* all over again – authors response. *Behavioral and Brain Sciences* **16**, 401–403.
- Cole, S., Cole, J. & Simon, G. (1981). Chances and consensus in peer review. *Science* **214**, 881–886.
- Dickersin, K., Chann, S., Chalmers, T., Sacks, H. & Smith, H. (1987). Publication bias in randomised controlled trials. *Controlled Clinical Trials* **8**, 343–353.
- Dickersin, K., Min, Y. & Meinert, C. (1992). Factors influencing publication of research results. *Journal of the American Medical Association* **267**, 374–378.
- Easterbrook, P., Berlin, J., Gopalan, R. & Matthews, D. (1991). Publication bias in clinical research. *Lancet* **337**, 867–872.
- Ernst, E. & Resch, K. (1994). Reviewer bias – a blinded experimental study. *Journal of Laboratory and Clinical Medicine* **124**, 178–182.
- Ernst, E., Saradeth, T. & Resch, K. (1993). Drawbacks of peer review. *Nature* **363**, 296.
- Evans, A., McNutt, R., Fletcher, S. & Fletcher, R. (1993). The characteristics of peer reviewers who produce good-quality reviews. *Journal of General Internal Medicine* **8**, 422–448.
- Feurer, I., Becker, G., Picus, D., Ramirez, E., Darcy, M. & Hicks, M. (1994). Evaluating peer reviews: pilot testing of a grading instrument. *Journal of the American Medical Association* **272**, 98–100.
- Fisher, M., Friedman, S. & Strauss, B. (1994). The effects of blinding on acceptance of research papers by peer review. *Journal of the American Medical Association* **272**, 143–146.
- Garfunkel, J., Ulshen, M., Hamrick, H. & Lawson, E. (1994). Effect of institutional prestige on reviewers' recommendations and editorial decisions. *Journal of the American Medical Association* **272**, 137–138.
- Gellert, G., Pitkin, R., Longmore, M. & Cowie, J. (1995). The death of biomedical journals. *British Medical Journal* **311**, 507–508.
- Goldman, R. (1994). The reliability of peer assessments: a meta-analysis. *Evaluation and the Health Professions* **17**, 3–21.
- Horrobin, D. (1974). Referees and research administrators: barriers to scientific research? *British Medical Journal* **ii**, 216–218.
- Horrobin, D. (1990). The philosophical basis of peer review and the suppression of innovation. *Journal of the American Medical Association* **263**, 1438–1441.
- Judson, H. (1994). Structural transformations of the sciences and the end of peer review. *Journal of the American Medical Association* **272**, 92–94.
- Kiesler, C. (1991). Confusion between reviewer reliability and wise editorial and funding decisions. *Behavioral and Brain Sciences* **14**, 151–152.
- LaPorte, R., Marler, E., Akawaza, S., Sauer, F., Gamboa, C., Shenton, C., Glosser, C., Villasenor, A. & Maclure, M. (1995). The death of the biomedical journals. *British Medical Journal* **310**, 1387–1390.
- Lock, D. & Smith, J. (1990). What do peer reviewers do? *Journal of the American Medical Association* **263**, 1341–1343.
- Lock, S. (1985). *A Difficult Balance: Editorial Peer Review in Medicine*. Nuffield Provincial Hospitals Trust: London.
- McNutt, R., Evans, A., Fletcher, R. & Fletcher, S. (1990). The effects of blinding on the quality of peer review: a randomized trial. *Journal of the American Medical Association* **263**, 1371–1376.
- Nylenna, M., Riis, P. & Karlsson, Y. (1994). Multiple blinded reviews of the same two manuscripts: effects of referee characteristics and publication language. *Journal of the American Medical Association* **272**, 149–151.
- Peters, D. & Ceci, S. (1982). Peer-review practices of psychological journals; the fate of published articles submitted again. *Behavioral and Brain Sciences* **5**, 187–255.
- Pierie, J. & Overbeke, J. (1996). Has any research been done on editorial peer review? A structured review of the literature. Submitted.
- Shepherd, M. (1986). Psychological medicine redivivus: concepts and communication. *Journal of the Royal Society of Medicine* **79**, 639–645.
- Smith, R. (1994). Promoting research into peer review. *British Medical Journal* **309**, 143–144.
- Spitzer, R. (1975). On pseudoscience in science: logic in remission and psychiatric diagnoses: a critique of Rosenhan's 'on being sane in insane places'. *Journal of Abnormal Psychology* **84**, 442–452.
- Squires, B. (1990). Editor's page; blinding the reviewers. *Canadian Medical Association Journal* **142**, 279.
- Stossel, T. (1985). Reviewer status and review quality: experience of the *Journal of Clinical Investigation*. *New England Journal of Medicine* **312**, 658–659.
- Tyrer, P. (1991). Chairman's action; the importance of executive decisions in peer review. *Behavioral and Brain Sciences* **14**, 164–165.
- Wessely, S., Brugha, T., Cowen, P., Smith, L. & Paykel, E. (1996). Do authors know who refereed their paper? *British Medical Journal* (in the press).
- Yankauer, A. (1991). How blind is blind review? *American Journal of Public Health* **81**, 843–845.