

## The Future of Peer Review

The peer review process has been established as the standard method for selecting the highest quality scientific research since the Second World War. Especially in the US, the peer review process is now often also used in the allocation of funding, adding financial responsibility to the intellectual duty of the reviewer. In theory, the experts in a field are best qualified to judge and improve the work of their peers. In practice, the peer review system, with its shielding traditions of blind reviewing, can also lead to rejection or alterations based on the status, age, sex or race of the author, the conclusions that the paper presents, or even the advancement of the reviewer's own career or that of his colleagues. The peer review can also be lengthy, slowing down the already considerable waits for publication of important research.

To evaluate the quality of peer review, many studies have been carried out: peer reviewed journals have been measured against so called 'throwaway' journals, reviewers have been blinded to the identity of the author, or made to reveal their own identities, mistakes have been deliberately introduced into papers to test the awareness of the reviewers and the comments of several reviewers have been compared with one another. Most of these tests have suggested what the majority of scientists were already aware of, that the peer review process, while it can improve the quality of articles, is imbued with flaws and biases. With this in mind, discussion has centred on what can be done to improve the system, whether peer review should be more open, more voluntary. The rise of the Internet has led some to fear that poor quality, non-peer reviewed research will be misleading, while others hail it as an opportunity for freer systems of peer review, with those who wish to review the paper posting comments, which may be swiftly read and responded to by authors and others in the field.

**Evaluating peer review:** Several tests have been carried out to evaluate peer review. In 1997 the International Biomedical Congress on peer review heard a range of reports on the subject. One study (Godlee, Gale and Martyn 1997) introduced eight questionable areas into a previous published paper, which was sent to potential reviewers, 221 of whom responded. These reviewers were divided randomly into five groups. Two groups received papers bearing the authors name and two reviewed papers in which the author's identity was concealed. The reviewers in one group from each of the above groups were required to sign their papers, while those in the other groups remained anonymous. The final group, in which the reviewers were not aware that they were taking part in a study, was used as a control. This study found that the median

number of errors commented on by the reviews was only two of the eight inserted problem areas, with a range from zero to five. Despite this weak performance, the study found no significant variation between blinded and unblinded groups.



In contrast to the results of this study, an investigation by the General Accounting Office (Abate, 1995), part of the U.S. congress, tested a far larger sample of reviewers (1,400), from three American public institutions, who were reviewing reports to decide which applicants should receive funding. This report showed that reviewers gave higher scores to colleagues that they knew. Authors with less well-established reputations and those from less prestigious institutions, as well as those from ethnic minorities were given less favourable scores. Female authors received lower scores and thus less funding in two of three institutions, but were better funded in the remaining institution.

Other studies have produced still more evidence of deficiencies in the peer review process. Rothwell and Martin analysed the correlations between 14 independent reviews of a paper, concluding that the reviewer's remarks were so disparate as to be completely random (Horrobin, 2001). Another complaint often levelled at peer review is that the system is unfairly biased towards American values and styles of writing. Another study presented to the 1997 International Biomedical Congress (Link, 1997) showed that those authors submitting papers to the American journal *Gastroenterology* from overseas, a proportion which made up two-thirds of the total submissions, were judged more harshly by peer reviewers and had a 12% higher likelihood of rejection than American scientists submitting to the same journal. Surprisingly perhaps, non-American reviewers also rated American submissions more highly than those from elsewhere, although not by as wide a margin as American reviewers.

## Constructive Reviewing

**Acting as a peer reviewer can increase your reputation with journal editors as well as keeping you up-to-date with the latest research in your field. Here, First Author\* offers some tips on how to be a constructive peer reviewer.**

**General tips:** When refereeing, it is first important to make sure that the subject is one within the correct specialisation in order to be able to make an informed critique. If this is not the case, or if time shortage is a problem, return the manuscript as soon as possible to avoid delay. If an invitation to referee is accepted, the report should normally be complete within 4 to 6 weeks. When the report is complete it should be returned to the editorial office, normally by e-mail, with a covering letter stating the manuscript number and title with a summary opinion which should normally recommend; immediate or barely altered acceptance, acceptance with revisions, and the extent of the necessary changes, or rejection, possibly with suggestions for alternative publication strategies with reasons for the opinion given. If recommending that a paper be accepted, give a strong recommendation and leave details of its drawbacks for the report itself. Try to avoid always recommending rejection, or always acceptance, as both give the impression of lack of discrimination.

**Taking a positive approach:** It is usually best to adopt a positive approach when reviewing; focus on the merits of the paper before its problems. Comments should be made on the basis of the manuscript's; clarity, breath, relationship to existing literature, practical relevance and the indications it gives for possible directions that further research should take. Suggestions should be made about the papers' organisation of information as transmitted in the text, tables and figures as well as commenting on any errors in figures, spelling or grammar. The main criteria when judging whether a paper should be accepted or rejected should be whether it contains an original idea or discovery that is relevant to current research. If this is the case, most other faults are resolvable, and it is up to the referee to give practical, specific and constructive advice about how to do so.

Others complain that innovative work is blocked by conservative peer reviews who, as established experts in their field, do not wish their orthodox ideas to be challenged; 'normally, those who review essays for inclusion in scholarly journals know what they are supposed to do. Their function is to take exciting, innovative, and challenging work by younger

scholars and find reasons to reject it' (Readings, 1994).

**Peer Reviewed vs. Throwing Journals:** In the face of all its evident defects, the first question that needs to be addressed when examining peer review is whether the process is in fact beneficial, or whether it should be abandoned altogether. One way to test the effectiveness of the system is to compare peer reviewed journals with so-called 'throwaway', non-peer-reviewed, publications. A study by Rochan et al (2002) investigated review articles published in the top five general medicine throwaway journals, compared with those in the five most widely read peer-reviewed journals on the same topic. This study found that peer reviewed articles were considered to be of a methodologically higher quality, with important issues better reported, and contained more references. However, the throwaway journals were judged more readable, both in layout and style of writing, whereas some peer-reviewed journal articles being rated as too difficult to read. Throwing journals were also considered more relevant to clinical practice, and were therefore more likely to be read. The indication from this study is therefore that peer review does help to ensure methodological and precision of reporting, but does not aid accessibility. Clearly, peer reviewed and throwing journals aim at different audiences, but a combination of the two styles of writing in addition to the existing journals could help to communicate important issues to practicing scientists in a more readable and relevant manner.



Although this is a valuable conclusion, comparison with non-peer reviewed journals cannot provide an accurate assessment of what peer review does to improve the quality of papers, as differences could reasonably be attributed to the original calibre of manuscripts submitted to or accepted for inclusion in peer-reviewed as opposed to non-peer-reviewed journals. To provide an accurate assessment of the true contribution of peer review, it would be necessary to carry out large-scale tests across several journals, with controls randomly selected for publication rather than chosen through the peer review process. It is likely that it would be difficult to secure the consent of authors and editors for such an undertaking, as this could (if peer review does indeed function as it should) compromise the quality

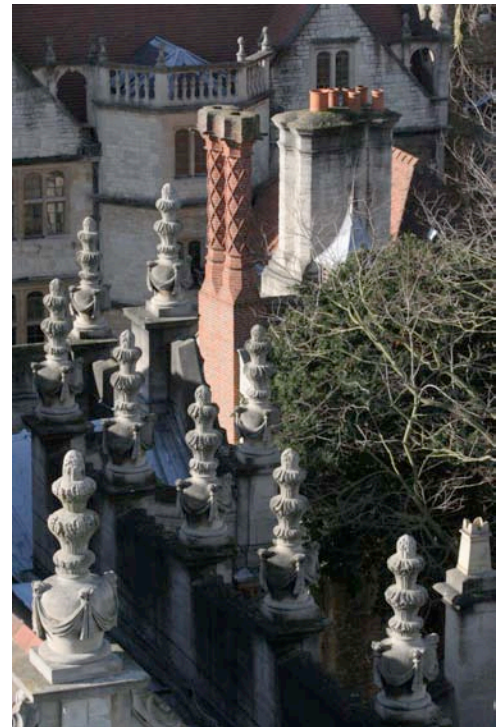
of those articles published, as well as representing an unfair selection process (Jefferson, Wager and Davidoff, 2002). Another problem with such a wide sample would be that standard criteria would be necessary to judge the effects of peer review. This is hampered both by the diversity of scientific writing and by the variable nature of individual peer reviewers. For example, it would be difficult to compare the effect of a reviewer's suggestions on statistical analysis on one hand, with another's on the research that should be included in a review of past literature on the other.

**Ideas for reform:** The effect of peer review is undoubtedly disparate between journals and between reviewers and is thus difficult to quantify. However, few are willing to abandon it although, even the most vociferous critic of the current system, David Horrobin (2001), advocating change rather than abandoning the entire system. The direction of reform therefore needs to be examined. The most obvious method of improving peer review is to train those who take part to improve their performances. A study reported in the BMJ (Schroter et al 2004) that examined the effects of training peer reviewers found that training, either by attending a course or receiving an instruction pack, showed a short-term benefit in that trained reviewers identified more errors in papers than a control group. However, this improvement was not sustained in the long term: no significant difference between the trained reviewers and the control group was evident six months after the original study. Therefore, more sustained training programs would have to be implemented to achieve sustained improvement, something which might meet with resistance from already over-worked academics engaged in voluntary peer review.

Another approach is to make changes in the selection of peer reviewers and the process of peer review. The GAO report discussed above suggested the institutions should make more effort to use female reviewers as well as those who are younger and less well established and those from under-represented ethnic groups. Abate (1995) also discusses the journal PLMA whose reviewers have been blinded to the identities of authors since 1980, with an apparently beneficial effect for younger and female scholars.

Another suggestion is that the peer review system, rather than being limited, is extended. Altman (2002) reviews studies carried out within various disciplines, one of which found that 38% of papers published in a peer reviewed psychiatry journal contained serious statistical errors, while another survey, this one referring to cancer trials, found that 81% did not report any systematic statistical design. Altman's solution is to implement standard procedures for methodological review. The CONSORT group (2001) provides detailed guidelines on reporting RCTs in the form of a

flow chart, a policy which has been adopted by major journals including Lancet and JAMA as a guide for authors, showing improvements in the quality of reporting when it has been used (Moher, Jones and Lepage, 2001).



**Letters to the Editor:** Enforcing the use of such guidelines for those refereeing work, in an attempt to force them to correct statistical errors, would require a change in the nature of peer review, making the now voluntary and unpaid task a much more formally regulated system. A different approach, although not mutually exclusive with the idea of the extending the duties of peer reviewers, is the concept that post publication criticism should be more central to the process of revising work. Several, but not all journals have a correspondence column with a facility for authors to respond to questions or criticisms about the paper. Horton (2002) reviewed the working of this process, looking at three research papers published in Lancet, the issues raised by readers, and the responses of the authors. In all three cases, the letters highlighted serious problem areas in the studies, several of which were acknowledged by the authors who replied to the queries. In none of the studies, however, did the authors respond to all of the queries, with under half addressed in two cases. As this study demonstrates, letters to the editor are a valuable, though undervalued, method of post-publication criticism, and their use should be extended, with correspondence columns in all journals, and authors encouraged to address readers' comment. This is an area in which online content of journals is particularly valuable, as it is easier to maintain a complete debate, including all previous comments and answers online.

**Open Peer Commentary:** 'Open peer commentary' has been pioneered by a few journals

including *The American Journal of Bioethics* and *Behavioural and Brain Sciences* (BBS). In BBS's system, after conventional refereeing and acceptance for publication, the articles are then circulated to up to 100 suitably qualified international commentators. The article is ultimately published with around 20-30 of those commentaries judged most relevant, with the authors' responses to them alongside. Although this approach has many advantages, it nevertheless suffers from the traditional problem of slow peer review, compounded by the volume of commentary gathered. *The American Journal of Bioethics*, rather than appointing peer commentators, invites submissions which, if approved by the editorial board, are included in the journal's correspondence page.

**Peer Review as Editorial Censorship?** Reform of the peer review system has been discussed numerous times without much significant change. In his editorial to the reports of Fourth International Congress on Peer Review published in *JAMA*, Rennie (2002) admits that 'in this issue of *THE JOURNAL* we publish studies that fail to show any dramatic effect, let alone improvement, brought about by editorial peer review'. His defense for the continuing use of peer review is that it represents a 'democratisation of the editorial process'. Some would argue that there are other motives behind the major journals' dedication to the retention of the system of peer review. Critics point to *JAMA*'s ban on authors speaking to the media prior to the publication of their articles as a deliberate 'gagging' of the scientific community for financial interests and claim that in order to secure grants, *JAMA* and others use a selective peer review process to ensure that articles 'toe the line'. Claims such as this, however, are largely unsubstantiated and form part of the wider conflict between advocates of freedom of information and publishers who seek to retain traditional barriers. The purpose of journals such as *JAMA* is undoubtedly not to prejudice and restricted access to information: however, it cannot be denied that journals themselves are a part of the competitive academic world, and that editors have at least some influence over the verdict of peer review, if only in their choice of a certain reviewer for a particular paper. Therefore, the extension of peer review, both pre and post-publication may remove some of the internal politics involved in publishing, if journals themselves are willing to take such an initiative.

**Electronic media and peer review:** The swiftest form of debate now takes place on the Internet, unconstrained by the traditional space limitations of print journals. The Internet therefore appears the most natural place for the reform of reviewing, as well as Horton's suggestion for the extension of post-publication criticism, to be adopted. Some journals have begun to publish articles in the form of AOP or electronic preprints (see Issue 1 for a glossary

of terms). Other suggestions that are beginning to be adopted in some areas are the selection of referees using electronic surveys of the literature, citation analysis, even posting calls for reviewers to pertinent professional experts' bulletin boards and allowing those who happen to have the time to volunteer themselves. A few of the electronic journals also run debate threads with answers from authors alongside articles, which may be revised in line with comments received. Some also publish online peer commentary after the fashion of BBS, (for example the APA-sponsored *PSYCOLOQUY*). The question of speed is of less importance in electronic publishing, and software such as LinkOut, allows external parties to link out from the PubMed records they think most relevant to their resource. Such initiatives have not been universally successful, however, authors are often skeptical about the safety of posting work and are often discouraged from doing so because of embargoes such as *JAMA*'s on pre-publication, and the fear that work may be plagiarised. Among the target audiences for the extension of Internet peer-review, many remain unaware, or suspicious of the validity, of work on the Internet, or fear that improper use of material could lead to harm to those who misunderstand information.



**Distrust of the Internet:** Although online counterparts to the traditional print journals are widely used, the new e-journals, despite optimistic statements, are still a minority, largely ignored by 'serious' academics, uncited and omitted from impact factor ratings. A survey in *JAMA* concluded that e-journals without a print counterpart published 'significantly fewer indexed complex types of publications, editorials, letters and case reports' than print journals (Weller, 2002). Even a more positive report, celebrating the numbers of e-journals

concludes that, in terms of the influence of electronic journals, 'for better and worse, commercial journals (as print products and as electronic aggregations) seem likely to dominate the field of refereed scholarly journals for some time to come' (Crawford, 2002). Given that electronic journals have a higher, faster circulation rates, opportunities for commentary and feedback, and, even by the most conservative estimates, production and distribution costs of under 75% (other estimates place the figure as low as 25%) the cost of print media, ease and limited (or non-existent) access costs for researchers, the unwillingness of scholars to participate cannot be attributed solely to the jealous guarding of territory by the traditional media. Harnad (1996) argues that it is the peer review process itself that authors feel is missing from electronic journals, and therefore suggests that a structured peer review system, limited by a system of write privileges for specialists, while others are allowed read-only access, must be implemented in order to persuade scholars and potential peer-reviewers that the Internet need not be an 'anarchic' medium, with writing unprotected from corruption or theft.

**Conclusions:** The evidence shows that peer review, despite its many faults, is still regarded by authors themselves as a necessary valedictory stamp for their work. The way forward, therefore, rather than placing more responsibility and pressure on the shoulders of existing peer reviewers, who are after all, performing a voluntary service, is to expand the area of peer review with more representative choices of referees including younger scholars, women, those from ethnic minorities and from overseas.

The types of reviewing carried out should be more extensive, with checking of procedures and statistical analyses, where applicable, as standard requirements. To close the gap between academic and laboratory scientists to some extent, reviewing should also ensure that articles are readable by and relevant to, those practising in the field. Blinding the reviewer to the identity of an author is likely to reduce the number of cases of prejudice in reviewing, difficult to entirely avoid, even for the most conscientious of reviewers. The selection process for reviewers should be opened up so that potential referees who have time and an interest in the subject, may volunteer themselves, thus reducing the chances of a hurried and indifferent or delayed review. Post-review commentary should also be encouraged, as should feedback from authors, both in print and electronic media.

Peer review can be adapted for the Internet, as a counterpart to a print journal as well as in solely Internet-based publications, and indeed, many of the improvements discussed here are more applicable to the faster and wider dissemination afforded by the Internet. Authors must be assured, however, that Internet mate-

rial may be reliably peer reviewed by respected contemporaries, as well as exposed to wider discourse. A two stage process of peer review consisting of an initial formal and structured system of evaluation by experts in the field, followed by a more interactive and open discussion between authors, experts and readers, is a possible solution.

## References:

- Abate, T. Ethics in Research. (Columbia University, 1995).
- Altman, D. Poor Quality Medical Research - What Can Journals Do? *JAMA* **287**, 21 (June 2002).
- CONSORT statement  
<http://www.consort-statement.org/> (accessed 11/05/2005) copyright David Moher, MSc,
- Crawford, W. Free Refereed Electronic Journals. *Learned Publishing*. **15**, No. 2 (2002)
- Schulz, KF and Altman, DG for the CONSORT group (2001).
- Godlee, F, Gale, CR and Martyn, CN., *University of Southampton, Report Presented to International Biomedical Congress September 1997*. (American Medical Association, 1997).
- Harnad, S. *Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals*. In: Peek, R. & Newby, G. (Eds.) *Scholarly Publication: The Electronic Frontier*. (Cambridge MA: MIT Press, 1996). pp. 103-108.
- Horrobin, David, F. Something Rotten at the Core of Science? *Trends in Pharmacological Sciences*, **22**, 2, February 2001.
- Horton, Richard, Postpublication Criticism and Clinical Knowledge. *JAMA*. **287**, 21 (June 2002).
- Jefferson, T, Wager, E and Davidoff, F. Measuring the Quality of Editorial Peer Review. *JAMA* **287**, 21 (June 2002).
- Link, Ann M. *Papers submitted from Inside and Outside the United States. Gastroenterology, American Gastroenterology Association, paper presented to International Biomedical Congress September 1997*. (American Medical Association, 1997).
- Moher, D, Jones, A, Lepage, L, Use of the CONSORT Statement and Quality of Reports of Randomised Trials. *JAMA* **285**,15 (April 2001).
- Readings, Bill. *Caught in the Net: Notes from the Electronic Underground*. (University of Montreal, 1994).
- Rennie, Drummond, Fourth International Congress on Peer Review in Biomedical Publication. *JAMA* **287**, 21 (June 2002).
- Rochan, PA, Bero, LA., Bay, AM, Gold, JL, Dergal, JM, Binns, MA, Streiner, DL. and Gurwitz, JH. Comparison of Review Articles Published in Peer Reviewed and Throwaway Journals. *JAMA*, **287**, 21 (2002).
- Schroter, S, Black, N, Evans, S, Carpenter, J, Godlee, F and Smith, R . Effects of training on quality of peer review: randomised controlled trial. *BMJ* 2004 328:673 (20 March), doi:10.1136/bmj.38023.700775.AE (published 2 March 2004).
- Weller, AC. *JAMA*. **287**, 21, (June 2002).