

FROM THE EDITOR'S DESK

Standards in the Face of Uncertainty

Peer Review Is Flawed and Under-Researched, but the Best We Have

Stephan Mertens, Christopher Baethge

Robbie Fox was no friend of peer review. The former editor of the *Lancet* is quoted as saying that you might as well throw a pile of manuscripts down the stairs and publish the ones that reach the bottom. He found the review system subjective, hard to reproduce, and arbitrary.

Peer review—the evaluation of texts by the author's peers, i.e., other researchers—is a characteristic feature of science. Nowadays no journal and no selection committee or research-funding body functions without this form of scientific self-monitoring.

Although *Philosophical Transactions*, one of the earliest science periodicals—founded in London in 1665—had its manuscripts reviewed by experts, peer review has been standard in scientific journals for only around 60 years (1).

Today over a million articles a year are published in peer-reviewed journals (1). The reviewers work of their own volition and free of charge. They write reviews as an obligation to the scientific community, despite the fact that conscientious analysis of the merits of a manuscript generally takes several hours of their time. The Research Information Network calculated that fair remuneration of scientists for the time they invest in peer reviewing would cost £ 1.9 billion per annum worldwide. This corresponded to £ 1200 per published article (2).

The peer review process

Typically, for example at *Deutsches Ärzteblatt*, an editorial team first determines—on criteria such as topic, intelligibility, and originality—whether a given manuscript is suitable in principle for publication in their journal. If the decision is positive, they then select experts who are familiar with the material. Most manuscripts are evaluated by two or more reviewers, and on the basis of their assessments as well as editorial guidelines the editors decide whether or not to publish. The reviewers for *Deutsches Ärzteblatt*, and for other journals, begin their reports with an overall recommendation, such as “accept,” “accept after revision,” or “reject,” and go on to evaluate the manuscript in detail and make suggestions for improvement. Incidentally, evaluation by colleagues continues after publication, in the form of letters to the editor, blogs, and forums (“post-publication review”). A cohort study found that authors who reply to letters to the editor address only around half the substantial criticisms raised (3)—if, indeed, such correspondence takes place at all.

But does pre-publication review actually improve the manuscripts, or does it just favor entrenched, conventional ideas and findings? How precise are the reviewers? To what extent do their opinions coincide? Can reviewing and reviewers be improved by training? Given the importance of reviewing and the enormous effort invested in the process, these questions have to be asked. Our intention here is to answer them on the basis of the research carried out to date.

Researchers themselves seem to be in favor of peer review. In an electronic survey in 2009, 40 000 randomly selected scientists were asked about their experience and opinions with regard to the prevailing review system. According to their self-classifications, 616 (15 %) of the 4 000 respondents were working in the field of medicine (1). This subgroup basically evaluated peer review in the same way as scientists in other disciplines:

- 69% are satisfied with the system
- 84% fear that scientific communication would be uncontrollable without peer review
- 91% believe their most recent article was improved by peer review
- 86% are happy to review manuscripts themselves and will continue to do so.

However, 56% of the respondents complained of a lack of guidance, and 68% thought reviewing would be improved by formal training.

According to this survey it took 6 hours on average to review a paper, and 86% of the respondents stated that they completed their reviews within a month of receiving the manuscript. These findings are consistent with those from another investigation in which questionnaires were sent to 39 232 scientists and 3040 responded (4).

Blinding

Usually, the author does not know who has assessed his manuscript. Because the reviewer's identity is unknown, he or she does not need to worry about the author's reaction to a negative review. Shielded by anonymity, however, reviewers can act arbitrarily and unfairly: they can recommend rejection of a sound manuscript, drag the process out, or even steal ideas, since the best reviewers are ultimately the authors' rivals in the same specialized area of research. In the double-blind process the manuscript is also anonymized, so that neither author nor reviewer knows the identity of the other. This is not as simple as it sounds,

however, because reviewers can often identify the authors from the topic of research and the publications cited.

The few journals with open review systems, where both author and reviewer are known to the other, include the *British Medical Journal* and the periodicals published by BioMed Central. Open review increases transparency and may protect the author from arbitrary behaviour. It is unpopular with reviewers, however, because a critical evaluation can cause resentment on the part of the author. This is particularly problematic for reviewers who are still climbing the career ladder—a group who, experience shows, produce some of the best reviews, because they are still working on their qualifications and are thus closely involved in research. In the interests of authentic evaluation, reviewers at *Deutsches Ärzteblatt* enjoy the protection of anonymity.

Evaluating the evaluation

Although millions of articles undergo peer review every year, only a handful of studies have investigated the efficacy of the process.

Four randomized controlled trials showed that the quality of the review is affected neither by the identity of the reviewers being known nor by anonymization of the authors (5–8). Thus the review process seemed not to be influenced by social issues.

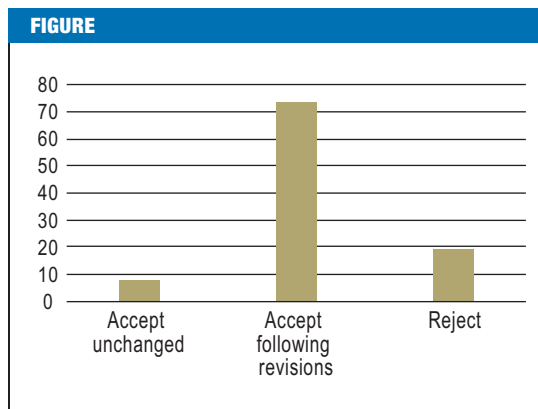
Two studies were carried out to determine how many deliberately inserted mistakes would be uncovered. Overall, 700 reviewers spotted 25% to 30% of eight or nine gross errors. The detection rate was not improved by any of various training measures (5, 9).

The editors of the *British Medical Journal* examined whether training could improve the quality of reviewers' work. While one study (10) showed that a workshop improved the quality of reviews in comparison with a control group, at least initially, a similar investigation found no effect at all, even in the short term (11). Further attempts to train reviewers, e.g., by means of workshops or feedback from the editors, proved fruitless (12). Other studies showed that reviewers well versed in methodology may improve manuscript quality, but here too the results were contradictory (12). Overall, a wide-reaching systematic Cochrane Review concluded there is little to indicate that peer review guarantees high quality of the articles that go on to be published (12).

Agreement between reviewers

In a retrospective study, Kravitz et al. examined the agreement between the reviewers of 2264 manuscripts submitted to the *Journal of General Internal Medicine* (13). They analyzed a total of 5881 reviews and found that 28% voted for acceptance, 28% for rejection, and around 45% for revision. Agreement between the reviewers' recommendations was found in 55% of cases. A meta-analysis of 52 studies found only slight agreement, however, with mean correlation of 0.3 (14).

The editors often followed the reviewers' recommendations: They rejected only 20% of the manuscripts where all reviewers had voted for acceptance. If all



The publication recommendations made in 554 reviewers of 206 manuscripts which cleared preliminary editorial screening at *Deutsches Ärzteblatt* between 1 July 2008 and 31 December 2009 (expressed as a percentage)

reviewers had been in favor of rejection, the editors followed their recommendation in 89% of cases. The slight discrepancy between recommendation and editorial decision can presumably be explained by reasons of space: Because only a certain volume of articles can be accommodated, some positively judged manuscripts have to be turned down. A second aspect is the difference between reviewers' recommendations and their detailed comments on the manuscript: At *Deutsches Ärzteblatt* it is not uncommon to find a discrepancy between these two components of the review, e.g., when a recommendation to reject the manuscript is not adequately substantiated by the reviewer's comments. Sometimes it seems that when reviewers write "reject," what they really mean is "reject in present form."

Between 1 July 2008 and 31 December 2009, *Deutsches Ärzteblatt* received 554 reviews of 206 manuscripts (Figure)—after a large number of manuscripts had been immediately rejected for editorial reasons and not even sent out for review. The mean number of reviews per manuscript was therefore 2.7. Only a small proportion of reviews (7.0%; n = 39) recommended acceptance of the manuscript without revision. Around three fourths (73.6%; n = 408) voted for revision, and a fifth (19.3%; n = 107) were in favor of rejection. As in the case of the *Journal of General Internal Medicine*, pairs of reviewers of a given manuscript were frequently of the same opinion: In six out of ten cases (61.2%), they agreed in recommending acceptance (0.4%), revision (55%), or rejection (5.8%). In almost one third of manuscripts (30.5%), however, one of the reviewers had recommended rejection.

Do original ideas have a chance?

Conservatively couched manuscripts that follow the mainstream have a higher chance of receiving a positive review (1). Unconventional, original ideas often meet with mistrust. At least this was the finding of a randomized trial in which a manuscript about a novel active substance was recommended for rejection more often than a manuscript about a known drug, although the two fictive papers were identical but for the name of the agent (15).

Ophthof et al. investigated whether the assessment of a manuscript by the reviewers and editors is in accordance with the reception of the published article by the scientific community. They compared the overall evaluation that led to acceptance for publication with the number of citations of the article—as a surrogate for importance—and found that the articles most positively evaluated at the manuscript stage were those most frequently cited (16).

Without doubt new concepts of quality assurance could be tested. The internet opens up new possibilities, particularly for more specialized journals: the evaluation process could be made interactive, allowing comments and cross-references. The *EMBO Journal* adopted the policy of publishing reviewers' and editors' comments and authors' replies in 2009. In this way the reader has access not only to the article but also to background information that may point to weaknesses or new topics for research.

Conclusion

The prevailing peer review process is under-researched, and the available evidence indicates that the faith in its efficacy is greater than the effect that can actually be measured. Nevertheless, peer review remains indispensable at the present time, particularly for general medical journals. At *Deutsches Ärzteblatt*, for example, the editorial team alone would not be able to evaluate the scientific quality of manuscripts from all the many fields of medicine. We are deeply indebted to our reviewers for their support.

Acknowledgment

We are grateful to Melanie Engels for her help in data acquisition.

Conflict of interest statement

Prof. Baethge is Chief Scientific Editor of *Deutsches Ärzteblatt* and *Deutsches Ärzteblatt International*. Dr. Mertens is Managing Editor for the Science and Medical Section of *Deutsches Ärzteblatt*.

Translated from the original German by David Roseveare.

REFERENCES

1. Peer review survey 2009: full report. www.senseaboutscience.org/data/files/Peer_Review_Survey_Final_3.pdf (last accessed June 2012)
2. Research Information Network: Activities, costs and funding flows in the scholarly communications system in the UK. <http://rinarchive.jisc-collections.ac.uk/our-work/communicating-and-disseminating-research/activities-costs-and-funding-flows-scholarly-commu> 2008 (last accessed August 2012).

3. Gotzsche PC, Delamothe T, Godlee F, Lundh A: Adequacy of authors' replies to criticism raised in electronic letters to the editor: cohort study. *BMJ* 2010; 341: c3926.
4. Mark Ware Consulting: Peer review in scholarly journals: perspective of the scholarly community—an international study. www.publishingresearch.net/documents/PeerReviewFullPRCReport-final.pdf 2008 (last accessed August 2012).
5. Godlee F, Gale CR, Martyn CN: Effect on the quality of peer review of blinding reviewers and asking them to sign their reports. *JAMA* 1998; 280: 237–40.
6. 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.
7. Justice AC, Cho MK, Winker MA, Berlin JA, Rennie D, and the peer investigators: Does masking author identity improve peer review quality? *JAMA* 1998; 280: 240–2.
8. 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* 1999; 318: 23–7.
9. Schroter S, Godlee F, Black N, Osorio L, Evans S, Smith R: What errors do peer reviewers detect, and does training improve their ability to detect them? *J R Soc Med* 2008; 101: 507–14.
10. Schroter S, Black N, Evans S, Carpenter J, Godlee F, Smith R: Effects of training on quality of peer review: randomised controlled trial. *BMJ* 2004; 328: 673–5.
11. Callaham ML, Wears RL, Waeckerle JF: Effect of attendance at a training session on peer reviewer quality and performance. *Ann Emerg Med* 1998; 32: 318–22.
12. Jefferson T, Rudin M, Brodny Folse S, Davidoff F: Editorial peer review for improving the quality of reports of biomedical studies (review) *Cochrane Database of Systematic Reviews* 2007, issue 2.
13. Kravitz RL, Franks P, Feldman MD, Gerrity M, Byrne C, Tierney WM: Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? *PLoS One*. 2010; 5: e10072.
14. Bornmann L, Mutz R, Daniel HD: A reliability-generalization study of journal peer reviews: A multilevel meta-analysis of inter-rater reliability and its determinants. *PLoS ONE* 2010; 5: e14331.
15. Resch KI, Ernst E, Garrow J: A randomized controlled study of reviewer bias against an unconventional therapy. *J R Soc Med* 2000; 93: 164–7.
16. Ophthof T, Coronel R, Janse MJ: The significance of the peer review process against the background of bias: priority ratings of reviewers and editors and the prediction of citation, the role of geographic bias. *Cardiovasc Res* 2002; 56: 339–46.

Corresponding author

Dr. sc. nat. Stephan Mertens
Deutsches Ärzteblatt,
Ottostr. 12, 50859 Köln, Germany
Mertens@aerzteblatt.de

Cite this as:

Mertens S, Baethge C: Standards in the face of uncertainty—peer review is flawed and under-researched, but the best we have. *Dtsch Arztebl Int* 2012; 109(51–52): 900–2. DOI: 10.3238/arztebl.2012.0900