What Makes a Good Reviewer and a Good Review for a General Medical Journal?

Nick Black, MD; Susan van Rooyen, BSc; Fiona Godlee, MRCP; Richard Smith, FRCP; Stephen Evans, MSc

Context.—Selecting peer reviewers who will provide high-quality reviews is a central task of editors of biomedical journals.

Objectives.—To determine the characteristics of reviewers for a general medical journal who produce high-quality reviews and to describe the characteristics of a good review, particularly in terms of the time spent reviewing and turnaround time.


Main Outcome Measures.—Review quality was assessed independently by 2 editors and by the corresponding author using a newly developed 7-item review quality instrument.

Results.—Of the 420 manuscripts, 345 (82%) had 2 reviews completed, for a total of 690 reviews. Authors’ assessments of review quality were available for 507 reviews. The characteristics of reviewers had little association with the quality of the review. Review quality increased with time spent on a review, up to 3 hours but not beyond.

Conclusions.—The characteristics of reviewers we studied did not identify those who performed high-quality reviews. Reviewers might be advised that spending longer than 3 hours on a review on average did not appear to increase review quality.

ALTHOUGH all editors would like to know how to select good reviewers, there have been only 3 attempts to identify their characteristics. Two of these studies found that the best-quality reports were provided by reviewers who were younger and, therefore, of junior academic status, particularly if they were working at a top academic institution or were known to the editors. The third study demonstrated that younger reviewers with considerable refereeing experience provided stricter assessments of manuscripts than other reviewers. None of the other characteristics examined (such as research training and postgraduate qualifications) were associated with review quality.

The process of peer reviewing has also received little attention. While 3 studies have reported on the time spent by reviewers on the task, none examined the relationship between time spent and review quality.

Our principal objective was to determine the characteristics of reviewers who produce high-quality reviews. In addition, we considered the characteristics of good reviews in terms of the time spent by the reviewer and time taken to deliver it to the journal.

METHODS

Consecutive manuscripts (research papers) submitted to BMJ and sent for review between January and June 1997 were eligible for inclusion. Each manuscript was sent to 2 reviewers (selected from the existing reviewer database as having an interest in and knowledge of the subject matter of the manuscript) as part of a randomized trial of blinding and unmasking to coreviewer. Reviewers were supplied with the journal's standard advice and asked to return their reviews for the Journal of the American Academy of Child and Adolescent Psychiatry, Am J Psychiatry. The characteristics of peer reviewers who produce good-quality reviews. J Gen Intern Med. 1993;8:422-428.


Characteristics of a Good Reviewer

**Editors’ Assessments.**—Four of the 16 characteristics of reviewers were significantly associated (P<.05) with the editors’ assessment of review quality: age (r² = 0.03), resident in North America (mean score, 3.22 vs 2.88; r² = 0.02), training in epidemiology or statistics (3.03 vs 2.74; r² = 0.04), and current research investigator (2.92 vs 2.74; r² = 0.006). Age showed a quadratic relationship in which younger reviewers up to about 60 years were more likely to produce higher-quality reviews (Figure), beyond which there was no statistically significant change in review quality.

When entered in a multiple regression model, only 2 characteristics remained significantly associated (P<.01) with higher-quality reviews: younger age and having training in epidemiology or statistics (Table). Together with the characteristic “resident in North America,” these 3 were, however, of very limited predictive power (r² = 0.08).

**Author’s Assessments.**—Four of the 16 characteristics of reviewers were significantly associated (P<.05) with the author’s assessment of review quality: age (r² = 0.008), training in epidemiology or statistics (2.94 vs 2.77; r² = 0.009), not a member of a journal editorial board (2.96 vs 2.81; r² = 0.011), and resident in North America (2.89 vs 2.80; r² = 0.006). The variables of very limited predictive power were: sex (r² = 0.003), postgraduate training in epidemiology or statistics (3.03 vs 2.88; r² = 0.02), and current research investigator (2.92 vs 2.88; r² = 0.006).

**RESULTS

**Recruitment and Response Rate**

An estimated 420 eligible manuscripts were submitted to the journal during the recruitment period of which 2 reviews were obtained for 345 (82%). Of these 690 reviews, information on the characteristics of reviewers was available for 670. The analyses presented here are based on all 670 reviews for the editors’ assessment of quality and on the first 507 reviews for which the corresponding author’s assessment was available. The only exception is the analysis of time spent carrying out the review, for which the sample was restricted to the 438 (editors’ assessment) and 342 (author’s assessment) reviewers who answered that question.
North America (3.13 vs 2.85; \( r^2 = 0.014 \)). When entered in a multiple regression model, 2 characteristics remained significantly associated with higher-quality reviews (Table). As above, these were, however, of very limited predictive power (\( r^2 = 0.02 \)).

**Characteristics of a Good Review**

There was no association between the editors’ assessment of review quality and the time taken by reviewers to return their reviews. There was, in contrast, a clear nonlinear relationship with the time spent by reviewers on their reviews. Review quality improved with increasing time up to about 3 hours, but not beyond.

**COMMENT**

The characteristics of reviewers considered in this study had little association with the quality of the reviews they produced. This was true regardless of whether editors or authors defined the quality of the review. The only consistent finding was that reviewers trained in epidemiology or statistics were more likely to produce good reviews. While age was associated with review quality according to editors’ assessments (consistent with previous studies\(^2\)), it was not when reviews were assessed by authors. The converse was true for the characteristic “not a member of an editorial board.” In regard to what makes a good review, the longer time spent on the task (up to about 3 hours), the better the review.

The lack of association between review quality and certain reviewer characteristics was surprising. We had expected that those actively involved in research, those occupying academic positions, and members of research funding bodies would have made better reviewers than others. This was not so. We did not seek to categorize the prestige of academic institutions so were unable to investigate the previously reported association between high-quality reviews and highly prestigious institutions. The association with North American residency was probably confounded given that such reviewers are more highly selected by a British journal than their British counterparts.

Before discussing the implications of these findings, 2 potential methodological limitations need to be considered. First, two thirds of the reviewers knew they were participating in a study, which may have affected the quality of their review, and the time spent carrying it out. Concern about a Hawthorne effect, however, appears to be unfounded as data presented elsewhere demonstrate.\(^7\) The mean total score for the unblinded, masked reviewers was 2.79 and for the uninformmed reviewers was 2.87 (difference, 0.08; 95% confidence interval, −0.06 to 0.22). Second, our findings depend crucially on the review-quality instrument. Full details of its development and validation are available from the authors. It has good internal consistency and interrater reliability, and we believe it was sufficiently accurate and robust for the purposes of this study. However, it should be noted that the instrument can only assess review quality in terms of content and completeness, not in terms of whether the reviewer’s judgment was correct.

So, what makes a good reviewer and a good review? Our failure to explain more than 8% of the characteristics of a good reviewer is either because we did not measure the relevant factors or no consistent pattern exists. In other words, there are almost as many types of good reviewers as there are good reviews. If true, the implication for editors is that they simply have to try new reviewers, assess their performance, and decide whether to continue to use them. This course of action raises the question of how new reviewers might learn their trade. It may be time for journals to start training their reviewers—though this assumes peer review is worthwhile and that people can be trained. Meanwhile, one suggestion that we can offer editors is to recruit reviewers with training in epidemiology or statistics, and probably to enlist people nearer 40 than 60 years of age. Reviewers might also be advised to spend no longer than 4 hours on their task.

Finally, it is unclear whether these findings are applicable to the vast majority of biomedical journals that have a specialized rather than a general focus and some of which may not provide guidance to their reviewers. Further research in this area might usefully include both types of journal.

We thank all the authors, reviewers, and editors who participated so willingly; Sue Minns and Marita Batten for any disruption we caused; and the National Health Service Executive North Thames Research & Development Responsive Funding Group, London, England, for their vision in supporting peer review research.

**References**