

3. Wilkes MS, Kravitz RL. Policies, practices, and attitudes of North American medical journal editors. *J Gen Intern Med.* 1995;10:443-450.

4. Schulman K, Sulmasy DP, Roney D. Ethics, economics, and the publication policies of major medical journals. *JAMA.* 1994;272:154-156.

5. Feurer ID, Becker GJ, Picus D, Ramirez E, Darcy MD, Hicks ME. Evaluating peer reviews: pilot testing of a grading instrument. *JAMA.* 1994;272:98-100.

6. Evans AT, McNutt RA, Fletcher SW, Fletcher

RH. The characteristics of peer reviewers who produce good-quality reviews. *J Gen Intern Med.* 1993;8:422-428.

7. Oxman AD, Guyatt GH. Validation of an index of the quality of review articles. *J Clin Epidemiol.* 1991;44:1271-1278.

8. Oxman AD, Guyatt GH, Singer J, et al. Agreement among reviewers of review articles. *J Clin Epidemiol.* 1991;44:91-98.

9. Strayhorn J Jr, McDermott JF Jr, Tanguay P. An intervention to improve the reliability of manuscript

reviews for the *Journal of the American Academy of Child and Adolescent Psychiatry.* *Am J Psychiatry.* 1993;150:947-952.

10. Cicchetti D. The reliability of peer review for manuscript and grant submissions: a cross-disciplinary investigation. *Behav Brain Sci.* 1991;14:119-186.

11. Nylenna M, Riis P, Karlsson Y. Multiple blinded reviews of the same two manuscripts: effects of referee characteristics and publication language. *JAMA.* 1994;272:149-151.

# 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.

**Design, Setting, and Participants.**—Surveys of reviewers of the 420 manuscripts submitted to *BMJ* between January and June 1997.

**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 reviews they produced (explaining only 8% of the variation), regardless of whether editors or authors defined the quality of the review. In a logistic regression analysis, the only significant factor associated with higher-quality ratings by both editors and authors was reviewers trained in epidemiology or statistics. Younger age also was an independent predictor for editors' quality assessments, while reviews performed by reviewers who were members of an editorial board were rated of poorer quality by authors. 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 as rated by editors and authors.

*JAMA.* 1998;280:231-233

From the London School of Hygiene & Tropical Medicine (Dr Black and Mr Evans) and *BMJ* (Ms Rooyen and Drs Godlee and Smith), London, England.

Presented at the International Congress on Peer Review in Biomedical Publication, Prague, Czech Republic, September 18, 1998.

Reprints: Nick Black, MD, Health Services Research Unit, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, England (e-mail: n.black@lshtm.ac.uk).

ALTHOUGH all editors would like to know how to select good reviewers, there have been only 3 attempts to identify their characteristics.<sup>1-3</sup> Two of these studies found that the best-quality reports were provided by reviewers who were young<sup>1,2</sup> and, therefore, of junior academic status,<sup>1</sup> particularly if they

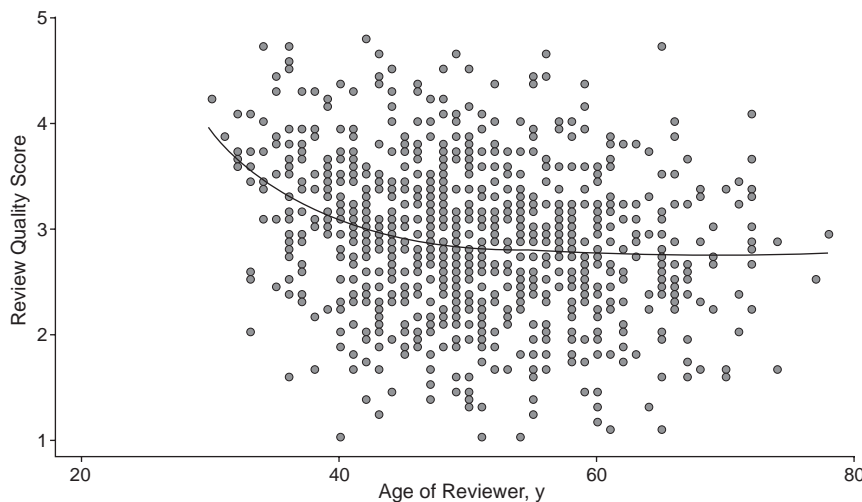
were working at a top academic institution or were known to the editors.<sup>2</sup> The third study demonstrated that younger reviewers with considerable refereeing experience provided stricter assessments of manuscripts than other reviewers.<sup>3</sup> None of the other characteristics examined (such 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,<sup>4-6</sup> 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.<sup>7</sup> Reviewers were supplied with the journal's standard advice and asked to return their



Scatterplot of review quality by age of reviewer. The review quality was scored on a 5-point Likert scale (1 = poor, 5 = excellent).

Reviewer Characteristics Associated With Higher-Quality Reviews Based on Editors' (n = 670) and Author's Assessments (n = 507)

Variable	$\beta$	SE $\beta$	P Value
<b>Editors' Assessment</b>			
Age, y			
>60	-.104	.03	$\leq .001$
40-60	.001	.0003	$\leq .001$
Postgraduate training in epidemiology and/or statistics	.201	.05	$\leq .001$
North American resident	.229	.10	.02
<b>Author's Assessment</b>			
Postgraduate training in epidemiology and/or statistics	.180	.07	.01
Member of editorial board	-.166	.07	.02

review within 3 weeks. In addition, two thirds of the reviewers (the third included in the "uninformed" portion of the trial had to be omitted to determine any Hawthorne effect) were asked to report how long they spent carrying out the review (including reading the manuscript, making notes, and writing the review). The time taken to return the review to the journal was recorded.

The quality of each review was assessed independently by 2 editors and by the corresponding author of the manuscript, using a new review-quality instrument. It considers the following 7 aspects of a review: the extent to which the reviewer addressed the importance of the research question, the originality of the question, the strengths and weaknesses of the method, the presentation of the paper (writing, organization, illustrations), the interpretation of the results, and the extent to which the reviewer provided constructive comments and substantiated the comments. Each item is scored on a 5-point Likert scale (1 = poor, 5 = excellent). The total score is the mean of the 7 item scores. In addition, there is a global item seeking an

overall assessment of the quality of the review. The internal consistency of the instrument is high (Cronbach  $\alpha = .84$ ) as is the interrater reliability of the total score (Kendall coefficient,  $\tau = 0.83$ ). Full details of its psychometric properties is available from the authors.

Information on the characteristics of reviewers was obtained from a mailed questionnaire survey carried out in 1996. This covered demographic characteristics (sex, age, country of residence), education and familiarity with research (age when received first degree, postgraduate qualifications, postgraduate training in epidemiology or statistics, current academic appointment, appointment in university or teaching center, current research investigator), publication experience (number of peer-reviewed research publications in past 5 years, number of papers reviewed in past year, number of journals where participated as a reviewer, member of a journal editorial board, member of a research funding body), and willingness to review blinded papers and to have identity revealed to authors.

Review quality was measured in the

following 2 ways: the mean of the 2 editors' total scores and the author's total score. Initially, the relationship between the total quality score and each reviewer characteristic was assessed using linear regression analysis. Characteristics that were statistically significant ( $P < .01$ ) were then entered stepwise using multiple regression (SPSS [Statistical Package for the Social Sciences] for Microsoft Windows, Release 6.1, SPSS, Inc, Chicago, Ill). The significance of comparisons of categorical variables was assessed using  $\chi^2$  tests.

## 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.

### 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 ( $r^2 = 0.03$ ), resident in North America (mean score, 3.22 vs 2.88;  $r^2 = 0.02$ ), training in epidemiology or statistics (3.03 vs 2.74;  $r^2 = 0.04$ ), and current research investigator (2.92 vs 2.74;  $r^2 = 0.006$ ). 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 ( $r^2 = 0.08$ ).

**Author's Assessments.**—Four of the 16 characteristics of reviewers were significantly associated ( $P < .05$ ) with the author's assessment of review quality: age ( $r^2 = 0.008$ ), training in epidemiology or statistics (2.94 vs 2.77;  $r^2 = 0.009$ ), not a member of a journal editorial board (2.96 vs 2.81;  $r^2 = 0.011$ ), and resident in

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<sup>1,2</sup>), 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.<sup>7</sup> The mean total score for the unblinded, masked reviewers was 2.79 and for the uninformed 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 is 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

1. Stossel TP. Reviewer status and review quality: experience of the *Journal of Clinical Investigation*. *N Engl J Med*. 1985;312:658-659.
2. Evans AT, McNutt RA, Fletcher SW, Fletcher RH. The characteristics of peer reviewers who produce good quality reviews. *J Gen Intern Med*. 1993; 8:422-428.
3. Nylenna M, Riis P, Karlsson Y. Multiple blinded reviews of the same two manuscripts: effects of referee characteristics and publication language. *JAMA*. 1994;272:149-151.
4. Yankauer A. Who are the peer reviewers and how much do they review? *JAMA*. 1990;263:1338-1340.
5. Lock S, Smith J. What do peer reviewers do? *JAMA*. 1990;263:1341-1343.
6. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review: a randomized trial. *JAMA*. 1990;263:1371-1376.
7. van Rooyen S, Godlee F, Evans S, Smith R, Black N. Effect of blinding and unmasking on the quality of peer review: a randomized trial. *JAMA*. 1998;280: 234-237.