Effect on the Quality of Peer Review of Blinding Reviewers and Asking Them to Sign Their Reports

A Randomized Controlled Trial

Fiona Godlee, BSc, MRCP; Catharine R. Gale, BSc; Christopher N. Martyn, DPhil, FRCP

Context.—Anxiety about bias, lack of accountability, and poor quality of peer review has led to questions about the imbalance in anonymity between reviewers and authors.

Objective.—To evaluate the effect on the quality of peer review of blinding reviewers to the authors’ identities and requiring reviewers to sign their reports.

Design.—Randomized controlled trial.

Setting.—A general medical journal.

Participants.—A total of 420 reviewers from the journal’s database.

Intervention.—We modified a paper accepted for publication introducing 8 areas of weakness. Reviewers were randomly allocated to 5 groups. Groups 1 and 2 received manuscripts from which the authors’ names and affiliations had been removed, while groups 3 and 4 were aware of the authors’ identities. Groups 1 and 3 were asked to sign their reports, while groups 2 and 4 were asked to return their reports unsigned. The fifth group was sent the paper in the usual manner of the journal, with authors’ identities revealed and a request to comment anonymously.

Main Outcome Measure.—The number of weaknesses in the paper that were commented on by the reviewers.

Results.—Reports were received from 221 reviewers (53%). The mean number of weaknesses commented on was 2 (1.7, 2.1, 1.8, and 1.9 for groups 1, 2, 3, and 4 and 5 combined, respectively). There were no statistically significant differences between groups in their performance. Reviewers who were blinded to authors’ identities were less likely to recommend rejection than those who were aware of the authors’ identities (odds ratio, 0.5; 95% confidence interval, 0.3-1.0).

Conclusions.—Neither blinding reviewers to the authors and origin of the paper nor requiring them to sign their reports had any effect on rate of detection of errors. Such measures are unlikely to improve the quality of peer review reports.

References

©1998 American Medical Association. All rights reserved.
Table 1.—Characteristics of Respondents and Nonrespondents

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Respondents</th>
<th>Nonrespondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD), y</td>
<td>50.8 (10.2)</td>
<td>51.4 (9.7)</td>
</tr>
<tr>
<td>Female, No. (%)</td>
<td>28 (13.9)</td>
<td>25 (16.3)</td>
</tr>
<tr>
<td>Postgraduate training in epidemiology/statistics, No. (%)</td>
<td>128 (63.1)</td>
<td>64 (57.3)</td>
</tr>
<tr>
<td>Currently involved in biomedical research, No. (%)</td>
<td>172 (84.3)</td>
<td>98 (82.3)</td>
</tr>
<tr>
<td>Journals reviewed for in 1995, No. (IQR)</td>
<td>5 (3-6)</td>
<td>4 (2-6)</td>
</tr>
<tr>
<td>Papers refereed in 1995, No. (IQR)*</td>
<td>12 (5-20)</td>
<td>10 (5-20)</td>
</tr>
<tr>
<td>Published papers in past 5 y, No. (IQR)*</td>
<td>17 (8-30)</td>
<td>15 (6-29)</td>
</tr>
<tr>
<td>On editorial board of scientific/medical journal, No. (%)</td>
<td>117 (57.6)</td>
<td>52 (43.3)</td>
</tr>
<tr>
<td>Previously given highest grades by BMJ for reviewer performance, No. (%)</td>
<td>89 (40.3)</td>
<td>46 (24)</td>
</tr>
</tbody>
</table>

*Median number; IQR indicates interquartile range.

RESULTS

Of the 420 reviewers invited to comment on the manuscript, 221 (53%) returned a report. Among the 74 people (18%) who gave a reason for declining to review the manuscript, 46 said they felt they were not competent to comment and 16 said they were too busy. There were no statistically significant differences in response rate among the 5 groups of reviewers. The 5 groups did not differ significantly by demographic or academic characteristic. Those who reported on the manuscript tended to have reviewed for more journals in the previous year than the nonresponders. They were also more likely to be on the editorial board of a biomedical journal and to have received a higher grading for their past performance in writing reports by BMJ. Characteristics of respondents and nonrespondents are shown in Table 1.

In total, 8 areas of weakness in design, analysis, or interpretation had been introduced into the manuscript. The mean number of weaknesses commented on was 2. Only 10% of reviewers identified 4 or more areas of weakness, and 16% failed to identify any.

We found no statistically significant difference in performance among the 37 reviewers in group 5 who had been sent the manuscript in the usual manner of BMJ and were unaware that they were taking part in a study and the 85 reviewers in group 4 who had been sent the manuscript under the same conditions but informed that they were participating in a study. These groups were therefore combined when we examined whether blinding reviewers and/or asking them to sign their reports affected the proportion of weaknesses they identified. Table 2 shows the mean number of errors detected by each group and rate ratios for identifying weaknesses. There were no statistically significant differences. Among the 90 respondents who had been blinded to the authors’ identities, 23 (26%) named the authors correctly in their report. Rate ratios were little changed when these people were excluded.

We also examined whether other characteristics of the reviewers were linked with better performance at identifying weaknesses in the manuscript (Table 3). Reviewers who had postgraduate training in epidemiology or statistics tended to comment on more points of weakness, but this relationship was not statistically significant. Neither sex nor experience as a peer reviewer or as a member of an editorial board was associated with the quality of the report. Younger reviewers, those currently involved in biomedical research, those with more publications in the previous 5 years, and those who had received a higher grad-
Table 2.—Poisson Regression Rate Ratios for Identifying Errors in the Manuscript: The Effect of Blinding Reviewers to the Authors’ Identities and Asking Them to Sign Their Reports

<table>
<thead>
<tr>
<th>No.</th>
<th>Mean No. of Errors Identified</th>
<th>Unadjusted Rate Ratios (95% CI)</th>
<th>Adjusted Rate Ratios (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not asked to sign and not blinded</td>
<td>721</td>
<td>1.9</td>
<td>1.0 (0.9-1.1)</td>
</tr>
<tr>
<td>Asked to sign and not blinded</td>
<td>30</td>
<td>1.8</td>
<td>0.9 (0.7-1.1)</td>
</tr>
<tr>
<td>Not asked to sign and blinded</td>
<td>59</td>
<td>2.1</td>
<td>1.1 (0.9-1.4)</td>
</tr>
<tr>
<td>Asked to sign and blinded</td>
<td>60</td>
<td>1.7</td>
<td>0.9 (0.7-1.1)</td>
</tr>
</tbody>
</table>

*Calculated after excluding cases where blinding was unsuccessful.
†Groups 4 and 5 were combined for this analysis.

Table 3.—Poisson Regression Rate Ratios for Identifying Errors in the Manuscript: Factors Associated With Better Performance

<table>
<thead>
<tr>
<th>No.*</th>
<th>Unadjusted Rate Ratios (95% Confidence Interval)</th>
<th>Adjusted Rate Ratios (95% Confidence Interval)†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤44</td>
<td>132</td>
<td>1.0</td>
</tr>
<tr>
<td>45-49</td>
<td>55</td>
<td>1.2 (0.9-1.6)</td>
</tr>
<tr>
<td>50-57</td>
<td>54</td>
<td>1.2 (0.9-1.6)</td>
</tr>
<tr>
<td>&gt;57</td>
<td>44</td>
<td>1.0</td>
</tr>
</tbody>
</table>

*Because of missing values for some variables, numbers do not always add up to 221.
†Adjusted for all variables in the table.

The questionnaire allowed us to investigate which characteristics of reviewers were associated with the most complete detection of errors in a manuscript. As others have shown, younger reviewers performed better than older reviewers. We also found, perhaps unsurprisingly, that reviewers who had themselves published a large number of articles in the past few years tended to identify more errors than those who had published little and that reviewers whose past performance in reviewing for BMJ had been highly rated by the editorial staff also tended to perform better.

One intriguing finding concerns reviewers’ recommendations to the editor about publication. The groups who were blind to authors’ identities were the least likely to recommend rejection. Blinding reviewers to the identities and affiliations of authors and requiring them to sign their reports might prove to be a successful strategy in reducing bias and increasing accountability in peer review, though it is hard to ensure that blinding is complete. It may also make them less likely to recommend rejection. But the results of this study do not suggest that such measures would be very effective in improving the accuracy with which they detect weaknesses in a manuscript.

We are grateful to members of Locknet (the international peer review research network), Richard Smith, BSc, MB, CHB, and Margaret Bendell of BMJ, and Paul Winter, MSc, of the MRC’s Environmental Epidemiology Unit for advice and assistance.

References
Does Masking Author Identity Improve Peer Review Quality?

A Randomized Controlled Trial

Amy C. Justice, MD, PhD; Mildred K. Cho, PhD; Margaret A. Winker, MD; Jesse A. Berlin, ScD; Drummond Rennie, MD; and the PEER Investigators

Context.—All authors may not be equal in the eyes of reviewers. Specifically, well-known authors may receive less objective (poorer quality) reviews. One study at a single journal found a small improvement in review quality when reviewers were masked to author identity.

Objectives.—To determine whether masking reviewers to author identity is generally associated with higher quality of review at biomedical journals, and to determine the success of routine masking techniques.

Design and Setting.—A randomized controlled trial performed on external reviews of manuscripts submitted to *Annals of Emergency Medicine, Annals of Internal Medicine, JAMA, Obstetrics & Gynecology, and Ophthalmology.*

Interventions.—Two peers reviewed each manuscript. In one study arm, both peer reviewers received the manuscript according to usual masking practice. In the other arm, one reviewer was randomized to receive a manuscript with author identity masked, and the other reviewer received an unmasked manuscript.

Main Outcome Measure.—Review quality on a 5-point Likert scale as judged by manuscript author and editor. A difference of 0.5 or greater was considered important.

Results.—A total of 118 manuscripts were randomized, 26 to usual practice and 92 to intervention. In the intervention arm, editor quality assessment was complete for 77 (84%) of 92 manuscripts. Author quality assessment was complete on 40 (54%) of 74 manuscripts. Authors and editors perceived no significant difference in quality between masked (mean difference, 0.1; 95% confidence interval [CI], −0.2 to 0.4) and unmasked (mean difference, −0.1; 95% CI, −0.5 to 0.4) reviews. We also found no difference in the degree to which the review influenced the editorial decision (mean difference, −0.1; 95% CI, −0.3 to 0.3). Masking was often unsuccessful (overall, 68% successfully masked; 95% CI, 58%-77%), although 1 journal had significantly better masking success than others (90% successfully masked; 95% CI, 73%-98%). Manuscripts by generally known authors were less likely to be successfully masked (odds ratio, 0.3; 95% CI, 0.1-0.8). When analysis was restricted to manuscripts that were successfully masked, review quality as assessed by editors and authors still did not differ.

Conclusions.—Masking reviewers to author identity as commonly practiced does not improve quality of reviews. Since manuscripts of well-known authors are more difficult to mask, and those manuscripts may be more likely to benefit from masking, the inability to mask reviewers to the identity of well-known authors may have contributed to the lack of effect.

From the Division of General Internal Medicine and Program for Health Care Research, Department of Veterans Affairs Medical Center and University Hospitals of Cleveland and Case Western Reserve University, Cleveland, Ohio (Dr Justice); the Center for Bioethics (Dr Cho) and the Division of Biostatistics, Department of Biostatistics and Epidemiology and Center for Clinical Epidemiology and Biostatistics, School of Medicine, University of Pennsylvania, Philadelphia (Dr Berlin); and the Institute for Health Policy Studies, University of California, San Francisco (Dr Rennie). Dr Winker is Senior Editor and Dr Rennie is Deputy Editor (West) for *JAMA.*

Reprints: Amy C. Justice, MD, PhD, Case Western Reserve University, School of Medicine, WG-29, 10900 Euclid Ave, Cleveland, OH 44106 (e-mail: acj@po.cwru.edu).

IT HAS BEEN suggested that masking reviewers to author identity would improve the fairness and the quality of peer review because well-known authors’ work may be reviewed less critically. Yet only a small fraction of journals routinely mask reviewers.2,3 When editors are asked why they do not mask, they cite an “overwhelming burden” associated with masking.2,3 Some question whether it is possible to mask successfully.2,4

One study, conducted at a single journal,5 demonstrated that the quality of masked reviews was statistically higher than that of unmasked reviews, although that difference was small. We tested the hypothesis that masking peer reviewers to author identity improves the quality of peer review at 5 biomedical journals. To increase the generalizability of our study, we used a masking procedure that is commonly practiced.

METHODS

Journals

Five journals participated in the study: *Annals of Emergency Medicine, Annals of Internal Medicine, JAMA, Obstetrics & Gynecology, and Ophthalmology.* Only 1 of these journals, *Annals of Emergency Medicine,* routinely masks reviewers to author identity.

Manuscript Enrollment

Eligible manuscripts were submitted between November 1995 and March 1996 and met the following inclusion criteria: (1) the manuscript reported original research, including meta-analyses but excluding case reports or letters, (2) the manuscript was sent for external peer review, and (3) the authors did not object to having their manuscripts enrolled. Authors were notified that their manuscripts would be included in a study of the peer review process unless they declined, and that declining to have their...