There is little evidence from this study to support changing current practice by blinding or unmasking to improve the quality of reviews. Blinding or unmasking might, however, have other advantages in the peer review process, such as ensuring that the review process is seen to be fair. In view of the difference between the results of this study and previous research, it is not possible to generalize from this study to other settings, particularly the many biomedical journals that are more specialized. Further research should encompass a wide variety of different types and sizes of journals.

### References

1. Lock S. *A Difficult Balance: Editorial Peer Review in Medicine.* London, England: Nuffield Provincial Hospitals Trust; 1985.
2. Kassirer JP, Campion EW. Peer review: crude and understudied, but indispensable. *JAMA.* 1994; 272:96-97.
3. Fisher M, Friedman SB, Strauss B. The effects of blinding on acceptance of research papers by peer review. *JAMA.* 1994;272:143-146.
4. Strasburger VC. Righting medical writing. *JAMA.* 1985;254:1789-1790.
5. Yanksuer A. Peer review again. *Am J Public Health.* 1982;72:239-240.
6. Shapiro S. The decision to publish: ethical dilemmas. *J Chronic Dis.* 1985;38:365-372.
7. Ingelfinger FJ. Peer review in biomedical publication. *Am J Med.* 1974;56:686-692.
8. Robin ED, Burke CM. Peer review in medical journals. *Chest.* 1987;91:252-255.
9. Feinstein AR. Some ethical issues among editors, reviewers and readers. *J Chronic Dis.* 1986; 39:491-493.
10. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer review. *JAMA.* 1990;263:1371-1376.
11. Laband DN, Piette MJ. A citation analysis of the impact of blinded peer review. *JAMA.* 1994;272: 147-149.
12. Black N, van Rooyen S, Godlee F, Smith R, Evans S. What makes a good reviewer and a good review for a general medical journal? *JAMA.* 1998;280:231-233.

# 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. Group 5 differed from group 4 only in that its members were unaware that they were taking part in a study.

**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.

*JAMA.* 1998;280:237-240

PEER REVIEW, as usually practiced by biomedical journals, protects the identity of reviewers but not of authors. Concerns about bias, lack of accountability, and poor quality of peer review have brought this practice into question. Two interventions, which together would reverse the balance of anonymity, have been suggested as possible solutions: removing authors' identities from the manuscript (blinding) and asking reviewers to sign their reports (signing).[1-4] We performed a randomized controlled trial to examine the effect on peer review of blinding reviewers and asking them to sign their reports.
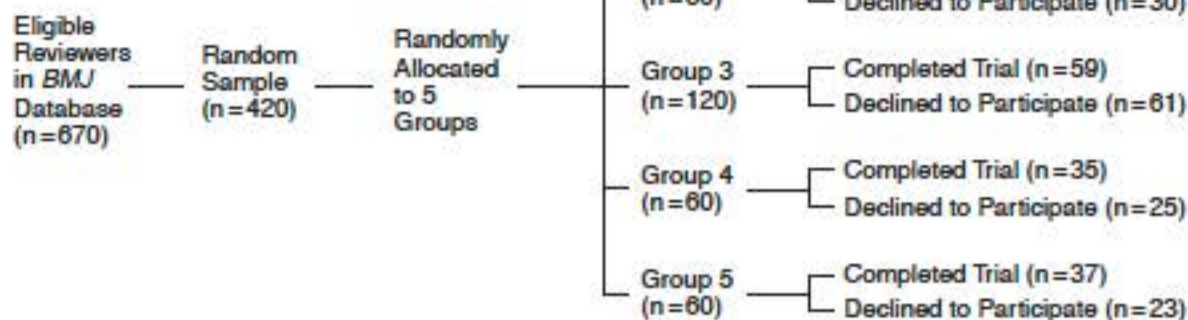
## METHODS

With the authors' consent, a paper already peer reviewed and accepted for publication by *BMJ* was altered to introduce 8 weaknesses in design, analysis, or interpretation.[4] All reviewers whose spe-

Reviewer selection process.

**Table 1.—Characteristics of Respondents and Nonrespondents**

| Characteristic | Respondents | Nonrespondents |
|---|---|---|
| Age, mean (SD), y | 50.8 (10.2) | 51.4 (9.7) |
| Female, No. (%) | 28 (13.9) | 25 (16.3) |
| Postgraduate training in epidemiology/statistics, No. (%) | 128 (63.1) | 64 (57.3) |
| Currently involved in biomedical research, No. (%) | 172 (84.3) | 98 (82.3) |
| Journals reviewed for in 1995, No. (IQR)* | 5 (3-6) | 4 (2-6) |
| Papers refereed in 1995, No. (IQR)* | 12 (5-20) | 10 (5-20) |
| Published papers in past 5 y, No. (IQR)* | 17 (8-30) | 15 (6-29) |
| On editorial board of scientific/medical journal, No. (%) | 117 (57.6) | 52 (43.3) |
| Previously given highest grades by *BMJ* for reviewer performance, No. (%) | 89 (40.3) | 48 (24) |

*Median number; IQR indicates interquartile range.

cialities seemed broadly relevant to the subject of the paper were selected from the journal's database (n = 834); 164 were excluded because they were known to be retired, dead, colleagues, or friends of the authors. Where more than 1 potential reviewer came from the same institution, the person whose name was lower in the alphabet was excluded. A statistician gave each reviewer a random number. These 670 reviewers were ordered and the first 420 were selected for allocation to 5 groups in random number sequence (Figure). Four of these groups were constructed, using a factorial design, to investigate the effects of blinding reviewers to the authors' identities and asking them to sign their reports. These reviewers were sent the paper by the editorial staff of the journal with a letter asking them to comment on the paper as part of a study into ways of improving peer review. Groups 1 and 2 were asked to comment on a version of the paper 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. A fifth group was sent the paper in the usual manner of the journal, with authors' identities revealed and on the understanding that the reviewer's name would be removed before the report was sent to the authors. Group 5 differed from group 4 only in that its

members were unaware that they were taking part in a study. A power calculation suggested that we would need 50 reviewers in each group to detect a difference in mean error score of 1 point ($\alpha = .05$, $\beta = .9$)—the smallest difference that journal editors judged worthwhile detecting. Because a pilot study had suggested that about half of the reviewers would decline to sign their reports, randomization was weighted 2 to 1 in favor of groups 1 and 3. All reviewers were asked to complete a questionnaire on their qualifications, academic position, and reviewing experience.

The main outcome measure was the number of weaknesses in the paper that were commented on by the reviewers. We also examined the recommendations the reviewers had made to the editors in regard to publication. Each reviewer's report was assessed independently by an editor and an epidemiologist, neither of whom was aware of the group to which the reviewer had been allocated. Where there was disagreement, the report was re-examined and consensus reached.

The frequency distribution of numbers of weaknesses detected followed a Poisson distribution so we used means to summarize these results. Analysis of variance, Mann-Whitney $U$ tests, $\chi^2$ tests, and logistic regression were used to examine response rate and recommendations regarding publication. Poisson regression was used to examine performance at identifying weaknesses in the manuscript.

## 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 35 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-

Effect on Quality of Peer Review of Blinding Reviewers—Godlee et al

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

| | No. | Mean No. of Errors Identified | Unadjusted Rate Ratios (95% Confidence Interval) | Adjusted Rate Ratios (95% Confidence Interval)* |
|---|---|---|---|---|
| Not asked to sign and not blinded | 72† | 1.9 | 1.0 | 1.0 |
| Asked to sign and not blinded | 30 | 1.8 | 0.9 (0.7-1.5) | 0.8 (0.6-1.2) |
| Not asked to sign and blinded | 59 | 2.1 | 1.1 (0.9-1.4) | 1.1 (0.9-1.4) |
| Asked to sign and blinded | 60 | 1.7 | 0.9 (0.7-1.2) | 0.9 (0.6-1.1) |

*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

| | No.* | Unadjusted Rate Ratios (95% Confidence Interval) | Adjusted Rate Ratios (95% Confidence Interval)† |
|---|---|---|---|
| **Age, y** | | | |
| ≤44 | 52 | 1.4 (1.0-1.8) | 1.3 (0.9-1.8) |
| 45-49 | 55 | 1.2 (0.9-1.6) | 1.2 (0.9-1.6) |
| 50-57 | 54 | 1.2 (0.9-1.6) | 1.2 (0.8-1.6) |
| >57 | 44 | 1.0 | 1.0 |
| | | $P = .05$‡ | $P = .26$ |
| **Currently involved in biomedical research** | | | |
| Yes | 172 | 1.3 (1.0-1.8) | 1.0 (0.7-1.5) |
| No | 32 | 1.0 | 1.0 |
| | | $P = .06$ | $P = .83$ |
| **BMJ grade** | | | |
| 0 | 132 | 1.0 | 1.0 |
| 6 | 53 | 1.3 (1.0-1.6) | 1.2 (0.9-1.5) |
| 7 | 36 | 1.3 (1.0-1.7) | 1.3 (1.0-1.7) |
| | | $P = .006$ | $P = .03$ |
| **Publications in previous 5 y, No.** | | | |
| ≤8 | 46 | 1.0 | 1.0 |
| 9-17 | 50 | 1.0 (0.7-1.4) | 0.9 (0.6-1.3) |
| 18-30 | 43 | 1.3 (0.9-1.8) | 1.1 (0.8-1.6) |
| >30 | 54 | 1.4 (1.1-1.9) | 1.2 (0.9-1.7) |
| | | $P = .001$ | $P = .009$ |

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

ing for their past performance in writing reports by *BMJ* were all more likely to identify weaknesses in the manuscript. In multivariate analysis, *BMJ* grade and number of publications remained statistically significant predictors of the number of errors detected.

Despite the weaknesses of the manuscript, 73 reviewers (33%) recommended that it be published with minor revision. Twenty-seven reviewers (12%) recommended that the manuscript should be published with major revision and 66 (30%) advised that it be rejected. Fifty-five (25%) made no recommendations regarding publication.

Reviewers in groups 1 to 4 who were aware they were taking part in a study tended to recommend outright rejection of the manuscript more often than those in group 5 who were unaware that they were participating in a study. We calculated odds ratios (ORs) for recommending rejection according to whether the reviewers had been blinded to the authors' identities or asked to sign their reports,

restricting the analysis to groups 1 to 4. Reviewers who had been blinded were less likely to recommend rejection than those who were not blinded (OR, 0.5; 95% confidence interval [CI], 0.3-1.0). This association was strengthened when reviewers who identified the authors' correctly were excluded (OR, 0.3; 95% CI, 0.1-0.6). Those who were asked to sign their reports were slightly less likely to recommend rejection than anonymous reviewers, but this relation was not statistically significant (OR, 0.7; 95% CI, 0.3-1.4).

## COMMENT

This randomized controlled trial was designed to investigate the effects of blinding reviewers to the authors' identities and asking them to sign their reports in a context as close as possible to the usual way in which *BMJ* operates peer review. We found that these interventions had no effect on the quality of reviewers' reports as judged by the number of weaknesses that they identified in the manuscript. Although 4 of the 5 groups of reviewers

knew that they were taking part in a study, their performance in detecting errors did not differ from those of the fifth group who were unaware that they were participating. We do not think that knowledge that they were taking part in the study is likely to have influenced these results. However, the response rate in the study was lower than usual for *BMJ* reviewers. Part of the explanation may be that the expertise of some of the reviewers we approached may have been peripheral to the subject of the manuscript. We know that 46 potential reviewers declined to write a report because they felt unqualified to comment. This point should also be borne in mind when considering the quality of the reports. Similar findings on reviewers' shortcomings at detecting weaknesses in a manuscript have been reported previously.[5] While a reviewer who discovered 4 or 5 serious flaws in a manuscript might reasonably feel that it was unnecessary to mention more, this can hardly explain the low mean score. Had we chosen reviewers with more relevant expertise, the score might have been higher.

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,[6] 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.

### References

1. Fisher M, Friedman SB, Strauss B. The effects of blinding on acceptance of research papers by peer review. *JAMA.* 1994;272:143-146.

2. Blank RM. The effects of double blind versus single blind reviewing: experimental evidence from the *American Economic Review. Am Econ Rev.* 1991;81:1041-1067.
3. McNutt RA, Evans AT, Fletcher RH, Fletcher SW. The effects of blinding on the quality of peer

review. *JAMA.* 1990;263:1371-1376.
4. Gale CR, Martyn CN, Cooper C. Cognitive impairment and mortality in a cohort of elderly people. *BMJ.* 1996;312:608-611.
5. Nylenna M, Riis P, Karlsson Y. Multiple blinded reviews of the same two manuscripts. *JAMA.* 1994;

272:149-151.
6. van Rooyen S, Godlee F, Evans S, Black N. What makes a good reviewer and what makes a good review? Paper presented at: Third International Congress on Peer Review in Biomedical Publication; September 17, 1997; Prague, Czech Republic.

# 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.

*JAMA.* 1998;280:240-242

IT HAS BEEN suggested that masking reviewers to author identity would improve the fairness and the quality of peer review[1] 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