

Same review quality in open versus blinded peer review in “Ugeskrift for Læger”

Siri Vinther¹, Ole Haagen Nielsen^{1,2}, Jacob Rosenberg^{1,2}, Niels Keiding³ & Torben V. Schroeder^{2,4}

ABSTRACT

INTRODUCTION: Research into the peer review process has previously been conducted in English-language journals. This study deals with a Danish general medical journal with a relatively small pool of both reviewers and readers. The aim of the study was to compare the quality of reviews produced by identifiable and anonymous reviewers, and further to characterize authors' and reviewers' attitudes towards different peer review systems.

MATERIAL AND METHODS: The study was conducted as a blinded, randomised controlled trial. Each manuscript was reviewed by an identifiable and an anonymous reviewer. Review quality was subsequently assessed by two blinded editors, using the validated Review Quality Instrument. Reviewers' and authors' attitudes towards different peer review systems were characterized using questionnaires.

RESULTS: The study included 364 reviews. There was no statistically significant difference in quality between anonymous and identifiable reviewers' evaluations. 55% of the authors preferred the evaluation produced by the identifiable reviewer ($p < 0.05$). 26% of the identifiable reviewers found it unpleasant that authors knew their identities; 43% of the anonymous reviewers found it reassuring that authors did not know their identities. Regarding reviewers' preferences for different peer review systems, 38% preferred a double-blinded, 34% preferred a single-blinded and 28% preferred an open system. For authors, the corresponding proportions were 43%, 19% and 37%.

CONCLUSION: Implementing open peer review will not affect review quality, but lack of anonymity may cause reviewers, already limited in number, to decline when asked to review. Even though this would be a serious implication for a national journal like the *Ugeskrift for Læger*, the implementation of an open system should be discussed.

FUNDING: not relevant.

TRIAL REGISTRATION: not relevant.

at the UfL is double-blinded, meaning that neither authors, nor reviewers know of each other's identities. Studies have shown that blinded peer reviews (whether double-blinded or single-blinded) reduce bias against certain authors, e.g. female authors, junior authors, controversial authors, authors from less prestigious institutions and authors doing research in “new” areas [1]. However, blinded reviewers are often able to identify the authors from the text and the references. Moreover, most studies have failed to demonstrate that blinded peer reviews improve review quality [2-5].

Previous studies of open versus blinded peer reviews have examined English-language journals with relatively large (international) pools of both readers and reviewers. The findings of these studies are not necessarily applicable to a journal such as the UfL, mainly because of the rather limited size of the Danish biomedical

ORIGINAL ARTICLE

- 1) Department of Gastroenterology, Herlev Hospital
- 2) Journal of the Danish Medical Association (*Ugeskrift for Læger*)
- 3) Department of Biostatistics, Institute of Public Health, Copenhagen University
- 4) Department of Vascular Surgery, Rigshospitalet

Dan Med J
2012;59(8):A4479

The Journal of the Danish Medical Association (*Ugeskrift for Læger*, UfL) is a Danish general medical journal publishing original articles, reviews, case reports and editorials. Issued weekly since 1839, it is one of the oldest peer-reviewed medical journals in the world. The journal is distributed to all members of the Danish Medical Association in addition to a limited number of subscribers, in total 23,000 copies. The current peer-review process



Implementing open peer review will not affect review quality, but lack of anonymity may cause reviewers, already limited in number, to decline when asked to review.


FIGURE 1

Review Quality Instrument (Version 4).

1. Did the reviewer discuss the importance of the research question?				
1	2	3	4	5
Not at all				Extensively discussed
2. Did the reviewer discuss the originality of the paper?				
1	2	3	4	5
Not at all				Extensively discussed with references
3. Did the reviewer clearly identify the strengths and weaknesses of the method (study design, data collection and data analysis)?				
1	2	3	4	5
Not at all				Comprehensively
4. Did the reviewer make specific useful comments on the writing, organisation, tables and figures of the manuscript?				
1	2	3	4	5
Not at all				Extensively
5. Were the reviewer's comments constructive?				
1	2	3	4	5
Not at all				Very constructive
6. Did the reviewer supply appropriate evidence using examples from the paper to substantiate their comments?				
1	2	3	4	5
None substantiated		Some substantiated		All substantiated
7. Did the reviewer comment on the author's interpretation of the results?				
1	2	3	4	5
Not at all				Discussed extensively
8. How would you rate the tone of the review?				
1	2	3	4	5
Abusive				Courteous

Source: [6, 7]

community. Having to sign a review containing critical comments about the work of a colleague or acquaintance – especially a senior colleague on whom your career may depend – may induce potential reviewers to decline the job. Alternatively, the reviews produced may be either too vague or too positive.

The objective of this study was to compare the quality of reviews produced by identifiable and anonymous reviewers, respectively. We also wanted to assess authors' and reviewers' attitudes towards different peer review systems (open, single-blinded and double-blinded).

MATERIAL AND METHODS

All manuscripts submitted to the UfL between September 2004 and November 2004 were eligible for inclusion. Each included manuscript was reviewed by two peer reviewers selected from the administrative database by UfL editors (the standard procedure at the UfL is to assign one reviewer per manuscript). The peer reviewers

were randomised into two groups: an intervention group consisting of peer reviewers who were asked to have their identity revealed to authors ("identifiable reviewers") and a control group consisting of peer reviewers who remained anonymous ("anonymous reviewers").

All peer reviewers were asked to review the manuscript, give recommendations regarding publication (or revision/rejection) and fill out a questionnaire (the wording of which differed slightly depending on whether the reviewer was identifiable or anonymous). Reminders were sent to those reviewers who did not respond initially.

Review quality was assessed by two editors, neither of whom was aware of the group to which a reviewer had been allocated. Review quality was assessed by means of the validated Review Quality Instrument (RQI) which consists of eight items (Figure 1) [6, 7]. Each of the eight items were scored on a five-point scale (1 = poor, 5 = excellent). The total and mean scores of questions 1-7 were calculated for each of the two editors.

Copies of the reviews (two for each manuscript) were sent to the author (after recommendations regarding publication had been given). Authors were asked which review they found the most constructive/helpful, which type of review process he/she would prefer in the future and, finally, if he/she had any considerations regarding the different types of peer review.

Regarding quality assessment, a difference of 0.3 in RQI score was considered significant. To detect such a difference (level of significance: 0.05; power: 0.9; assumed loss: 20%), 190 manuscripts were included in the study. A total of 182 complete datasets (two reviews and two completed reviewer questionnaires for each manuscript) were available for quality assessment and subsequent statistical analysis (Figure 2). In all, 157 completed author questionnaires were available for analysis.

The responses to the RQI were compared item by item by standard pairwise t-tests. In order to assess the combined results of all items, we used Hotelling's T^2 -test, a multivariate extension of Student's t-test. An associated correlation matrix indicated the extent to which the various items measured similar dimensions of the opinions of the editors.

Reviewers' recommendations regarding publication were scored on a five-point scale (1 = reject, 5 = publish without revision) and compared by standard pairwise t-tests.

This editorial study did not need ethics approval from the Danish Ethics Committee. All study participants (authors and reviewers) were informed about the study and agreed to participate.

Trial registration: not relevant.

RESULTS

The two blinded editors each assessed the quality of 364 reviews using the RQI. For eight manuscripts, it was not possible to obtain complete data (an open *and* a blinded review) despite that several reminders were sent to reviewers; these eight manuscripts (corresponding to 16 reviews) were excluded from the analysis.

Quality scores (means) for each of the eight items are presented in **Table 1**. For each item, there was no statistically significant difference between the reviews produced by the identifiable reviewers and the reviews produced by the anonymous reviewers. Furthermore, as for the mean quality score (items 1-7), the difference between the two groups was statistically insignificant (3.34 versus 3.28, standard deviation of difference = 1.22, $p = 0.51$) (Table 1).

Each editor gave very similar scores to all items, suggesting a high degree of intra-rater correlation between the items (confirmed by a correlation matrix; see also Table 1 in which the values of each column are very similar). Hotelling's T^2 -test showed that there was no statistically significant differences between the eight item responses when taken together ($p = 0.24$).

This gave reason to test whether the average of the eight items was significantly higher for the reviews produced by the identifiable reviewers or the reviews produced by the anonymous reviewers. The average of the

eight differences between the two groups was 0.06, and the t-test (based on the null hypothesis that the average difference was zero) yielded a $p = 0.46$. The 95% confidence interval for the mean of the differences was (-0.10 to 0.22).

Regarding reviewers' publication recommendations, the identifiable reviewers rejected 2.5% of the manuscripts directly (score 1); this proportion was 15% for anonymous reviewers. Yet, anonymous reviewers recommended publication without revision (score 5) just as often as did identifiable reviewers. Comparing mean scores, there was no statistically significant difference between identifiable and anonymous reviewers' recommendations (3.66 versus 3.39, $p = 0.15$).

A questionnaire was sent to all reviewers; answers are summarised in **Table 2**. Regarding reviewer background, 336 reviewers (92%) were specialists, 19 (5%) were in training and seven (2%) of the reviewers were not medical doctors. Regarding attitudes towards anonymity, 26% of the identifiable reviewers found it unpleasant that the author knew their identities; 43% of the anonymous reviewers found it reassuring that authors did not know their identities. Among all reviewers, 8% stated that the wording of their review report would depend on whether the review process was open or not. Among the 182 anonymous reviewers, 119 (65%) stated that they had an idea who wrote the manuscript;

FIGURE 2

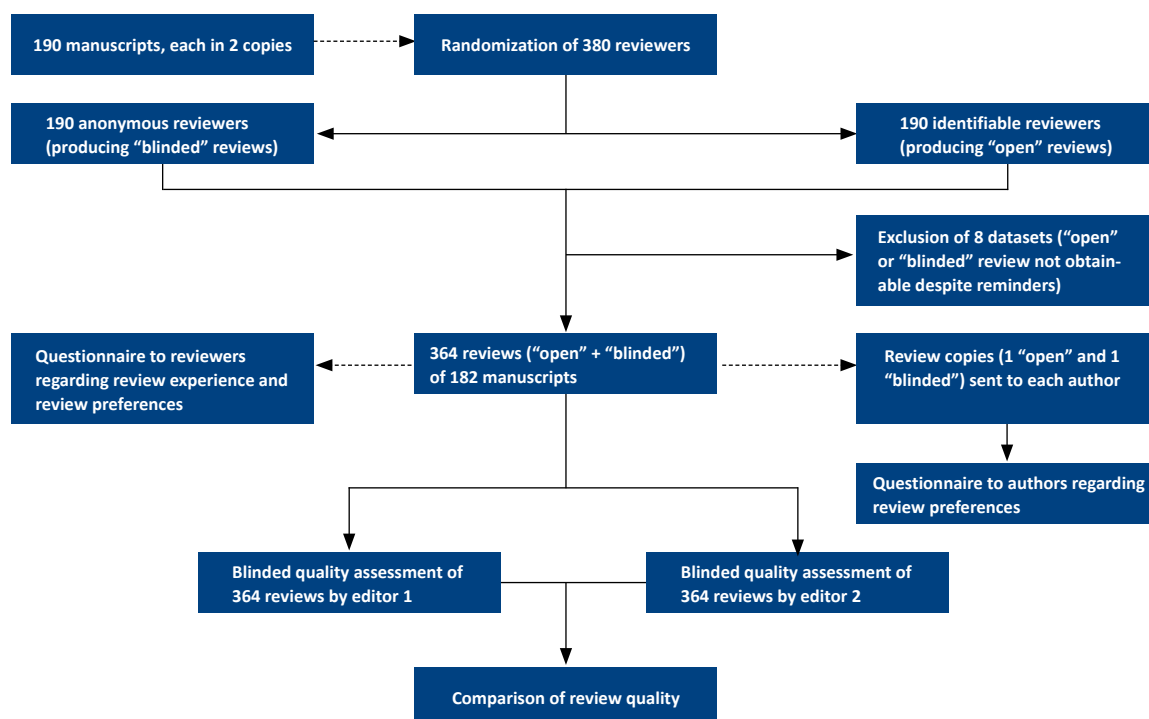


TABLE 1

Review quality scores.^a The values are means of the two editors' assessments (with standard deviations).

Item	Reviews produced by identifiable reviewers (n = 182)	Reviews produced by anonymous reviewers (n = 182)	Difference	p-value
1. Importance	3.32 (0.87)	3.30 (0.79)	0.01 (1.19)	0.91
2. Originality	3.16 (0.83)	3.12 (0.84)	0.04 (1.19)	0.66
3. Method	3.39 (0.91)	3.36 (0.90)	0.03 (1.33)	0.76
4. Presentation	3.48 (0.96)	3.36 (0.91)	0.13 (1.30)	0.18
5. Constructiveness of comments	3.53 (0.90)	3.45 (0.85)	0.08 (1.31)	0.41
6. Substantiation of comments	3.27 (1.00)	3.20 (0.94)	0.07 (1.37)	0.49
7. Interpretation of results	3.13 (0.89)	3.12 (0.84)	0.01 (1.19)	0.91
8. Tone of review	3.44 (0.53)	3.32 (0.60)	0.12 (0.86)	0.06
Mean total score of item 1-7	3.34 (0.82)	3.28 (0.78)	0.06 (1.22)	0.51

a) Each item is scored on a five-point scale (1 = poor, 5 = excellent) [5].

TABLE 2

Selected reviewer answers. The values are n (%).

	Identifiable reviewers (n = 182)	Anonymous reviewers (n = 182)
<i>Do you find it unpleasant that the author knows your identity?</i>		
Yes	47 (26)	–
No	135 (74)	–
<i>Do you find it reassuring that the author does not know your identity?</i>		
Yes	–	78 (43)
No	–	103 (57)
<i>Is your review any different now that the author knows who you are?</i>		
Yes	12 (7)	–
No	170 (93)	–
<i>Is your review any different now that the author does not know who you are?</i>		
Yes	–	18 (10)
No	–	163 (90)
<i>Even if you had not been told who the author was, would you have any idea who wrote the manuscript and where it was from?</i>		
Yes	111 (61)	–
No	71 (39)	–
<i>Do you have any idea who wrote the manuscript and where it is from?</i>		
Yes	–	119 (65)
No	–	63 (35)
<i>Which peer review system do you prefer in the future?</i>		
Open	54 (30)	50 (27)
Single-blinded	66 (36)	56 (31)
Double-blinded	62 (34)	76 (42)

of these, 84% guessed correctly. Of all reviewers, 38% preferred a double-blinded review system (i.e. to maintain status quo), 34% preferred a single-blinded review system and 28% preferred an open review system (Table 2). These preferences did not depend on whether reviewers were identifiable or anonymous ($\chi^2 = 2.39$, degrees of freedom = 2, $p = 0.30$).

Authors were asked which review they thought was the most constructive/helpful; 8% had no preferences,

55% found the review produced by the identifiable reviewer the most constructive/helpful and 36% found the review produced by the anonymous reviewer the most constructive/helpful ($p < 0.05$). Concerning the preferred review system, 43% of all authors preferred a double-blinded peer review system, 37% preferred an open review system and 19% preferred a single-blinded review system, while 1% had no preferences.

DISCUSSION

This study showed that reviews produced by identifiable reviewers were of the same quality as those produced by anonymous reviewers. Furthermore, the fact that a majority of the authors (55%) found the review produced by the identifiable reviewer more constructive/helpful suggests that an open peer review process is superior to a blinded process. These findings are in line with those reported from previous studies examining peer review in international, English-language journals [2, 5, 8-10].

When asked about their preferred review system, 38% of the reviewers and 43% of the authors stated that they would prefer a double-blinded review system (i.e. to maintain status quo). In theory, anonymity enables reviewers to produce honest evaluations without having to fear personal conflicts and rivalries. Especially younger and less experienced reviewers (i.e. doctors in training) may be reluctant to criticise the work of senior colleagues if this could affect their future career options. Yet, only 5% of the reviewers were doctors in training, which makes this argument rather irrelevant in this study.

Of the authors, 43% preferred a double-blinded review system and 37% preferred an open review system. Additionally, 55% of the authors preferred the review produced by the identifiable reviewer over the review produced by the anonymous reviewer. This may be due to anonymous reviewers being harsher than identifiable reviewers (cf. that anonymous reviewers were more likely to reject manuscripts directly and that authors were aware of the fate of their manuscripts when filling out the questionnaire). Yet, anonymous reviewers recommended publication (without revision) just as often as did identifiable reviewers. This contrasts with previous studies in which anonymous reviewers were less likely to recommend publication than identifiable reviewers [2, 5, 9, 10].

The strengths of this study include its randomised controlled design, the sample size based on initial power calculations, the high response rate, the availability of a validated instrument for review quality assessment and the fact that the quality assessments were performed blindly. Moreover, by examining attitudes towards different peer review processes, this study considered the perspectives of both authors and reviewers.

The study also had a number of limitations. Firstly, the RQI does not allow for investigators to determine whether or not a review is accurate, only review content and tone [5]. Assessing manuscript quality (in addition to review quality) would perhaps lead to other conclusions. Secondly, although blinded, only two editors assessed the quality of the reviews. Thirdly, reviewers could not be blinded (cf. that they had to answer a questionnaire). Fourthly, the questionnaires administered to authors and reviewers were not validated. Fifthly, some degree of underreporting is likely; the peer review process (open or blinded) may, after all, affect how reviewers express themselves. Lastly, authors and reviewers were asked to give their opinion on different peer review systems, but possible reasons for preferences were not examined.

The findings of this study should lead to a discussion about whether or not to implement open peer review at the UfL. The BMJ decided to open up the process after a study similar to the present concluded that review quality would not be affected [5, 11, 12]. Yet, in small biomedical communities, the lack of anonymity may cause reviewers, already limited in number, to decline when asked to review. This would be a much more serious implication for a national journal like the UfL than for English-language journals with international and thus much larger pools of reviewers.

Even though a considerable proportion of reviewers (and authors) preferred anonymity, and thus the current double-blinded system, implementing an open peer review system should be considered anyway. A majority of the authors in this study found the review produced by the identifiable reviewer to be the most constructive/helpful. Moreover, "blinded" reviewers were obviously able to identify the authors anyway, thus undermining this argument for maintaining a double-blinded system. Open peer review would make it impossible for reviewers to hide behind anonymity, thus allowing reviewers and authors to have a professional, collegial and constructive dialogue. Peer reviewers would be able to put a manuscript into the relevant context, ask appropriate questions and raise potential conflicts of interest to the editor if relevant. Furthermore, open peer review would make it easier to establish potential conflicts of interests of reviewers just as it would be easier to credit reviewers. Finally, opening up the peer review process would reduce administrative procedures, allowing already inadequate editorial office resources to be allocated elsewhere.

In conclusion, the findings of this study suggest that substituting the current double-blinded peer review system with open review at UfL will not affect review quality. Yet, a considerable proportion of reviewers and authors preferred anonymity and, thus, a blinded peer

review system. Implementing open peer review may thus reduce the already limited number of reviewers, a serious implication for a national journal like the UfL. In small biomedical communities, such as the Danish, blinded peer review thus has its advantages, but, nevertheless, the implementation of an open system should be discussed.

CORRESPONDENCE: Siri Vinther, Gastroenterologisk Afdeling, Herlev Hospital, Herlev Ringvej 75, 2730 Herlev, Denmark.
E-mail: sirivinther@hotmail.com

ACCEPTED: April 27, 2012

CONFLICTS OF INTEREST: Disclosure forms provided by the authors are available with the full text of this article at www.danmedj.dk.

LITERATURE

1. Triggler CR, Triggler DJ. What is the future of peer review? Why is there fraud in science? Is plagiarism out of control? Why do scientists do bad things? Is it all a case of: "all that is necessary for the triumph of evil is that good men do nothing"? *Vasc Health Risk Manag* 2007;3:39-53.
2. Godlee F, Gale CR, Martyn CN. Effect on the quality of peer review of blinding reviewers and asking them to sign their reports – a randomized controlled trial. *JAMA* 1998;280:237-40.
3. McNutt RA, Evans AT, Fletcher RH et al. The effects of blinding on the quality of peer-review – a randomized trial. *JAMA* 1990;263:1371-6.
4. Regehr G, Bordage G. To blind or not to blind? What authors and reviewers prefer. *Med Educ* 2006;40:832-9.
5. van Rooyen S, Godlee F, Evans S et al. Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *BMJ* 1999;318:23-7.
6. van Rooyen S, Godlee F, Evans S et al. Effect of blinding and unmasking on the quality of peer review. *J Gen Intern Med* 1999;14:622-4.
7. van Rooyen S, Black N, Godlee F. Development of the Review Quality Instrument (RQI) for assessing peer reviews of manuscripts. *J Clin Epidemiol* 1999;52:625-9.
8. Jefferson T, Alderson P, Wager E et al. Effects of editorial peer review – a systematic review. *JAMA* 2002;287:2784-6.
9. Justice AC, Cho MK, Winker MA et al. Does masking author identity improve peer review quality? A randomized controlled trial. *JAMA* 1998;280:240-2.
10. Walsh E, Rooney M, Appleby L et al. Open peer review: a randomised controlled trial. *Br J Psychiatry* 2000;176:47-51.
11. Smith R. Opening up BMJ peer review – a beginning that should lead to complete transparency. *BMJ* 1999;318:4-5.
12. Smith R. Peer review: a flawed process at the heart of science and journals. *J R Soc Med* 2006;99:178-82.