

# Scientific Peer Review: an Evaluation

Znanstvena recenzija: ocjena vrijednosti

Allan G. Farman,  
Zoltan T. Bérky,  
Taeko T. Farman

School of Dentistry, The  
University of Louisville,  
Louisville, Kentucky 40292,  
USA

---

## Summary

**Objectives:** Editors of journals depend on the system of peer review to screen out poorly conceived, poorly executed and unoriginal manuscripts. This study was designed to assess the reliability and consistency of reviewers' responses with regard to manuscripts submitted for publication to a leading international dental journal.

**Methods:** Three elements in the peer review process were investigated; namely: (1) reviewers' reports on manuscripts to the oral and maxillofacial radiology section of Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics were studied for consistency in the three reports for each of 78 consecutively submitted manuscripts rated as "full papers"; (2) ten previously accepted and ten previously rejected manuscripts were sent out for re-review by different panels of reviewers who were uninformed of the prior review; and (3) questionnaires were sent to the editors of ten leading dental journals to compare their peer review criteria and outcomes.

**Results:** Reviewers were consistently reliable in their responses to questions regarding originality and scientific merit, and in delineating manuscript acceptability. Of manuscripts that were previously accepted eight of ten were again accepted following a further double blind review. Of manuscripts that were rejected originally, six of ten were again rejected upon a second review by other referees. The use of two reviewers was validated for accepted articles; but validity required at least three reviewers when manuscripts were rejected. The selected journal editors reported acceptance rates of from 30 to 80% with publication waiting times from three to 24 months. Higher acceptance percentages generally were associated with greater delays in publication.

**Conclusions:** Journal peer review is not perfectly reliable; however, it does serve the purpose of reducing the number of poorly conceived and poorly constructed research papers.

Key words: journal editing; literature, dental; literature, scientific; peer review; publication; scientific method

---

Acta Stomatol Croat  
1997; 345—352

ORIGINAL SCIENTIFIC  
PAPER  
IZVORNI ZNANSTVENI  
RAD

Received: October 8, 1997  
Priljeno: 8. listopada 1997.

## Introduction

Status-judges are integral to any system of social control through their evaluation of role-performance and their allocation of rewards for that performance (1). In the case of scientific journals the significant status-judges are the editors and the referees. The referees are "presumed experts" who usually review the manuscript blind to the author and institution making the publication submission. Their decisions are made to the editor in confidence and may be shared with the author(s) by the editor. The authors are usually blind to the names and institutions of the reviewers. In contrast to scientific reviewers, other judges in science make their judgements publicly, as in the case for book reviews and review articles which assess the credibility of recent work in a special field of knowledge.

The journal peer review system originated more than 300 years ago, perhaps with the creation of *the Journal des Scavans* in January 1665, and *the Philosophical Transactions* of the Royal Society two months later (1). In both cases primary responsibility was assigned to one person, although not designated as the "editor" at the time. There were also councils which maintained a measure of control over journal content and became the basis for the peer review mechanism. The major functions of the journal peer review system are to screen out fallacious or poorly constructed investigations, to document priority in scientific discoveries, to assure that results are presented clearly and in appropriate detail, and to direct reports to the most appropriate audiences (2). Nevertheless, the role of peer review in weeding out suspect papers, inclusive of those which are fraudulent or plagiarized, does not always work. It is unreasonable for the review process to catch all errors or even outright fraud in scientific papers (3-5). The "experts" selected by the editor might not be as competent as the editor hopes. The reviewers are expected to possess not only scientific expertise but also literary and stylistic skills plus discriminatory judgments. Additionally, the review work is generally voluntary and unpaid. Inter-referee disagreement can be caused by reviewer bias in scientific beliefs leading to subjective appraisal, and editors might be swayed by the eminence of the author (1,2,6-10). Reviewer bias against authors and institutions can be reduced by blinding them to the

manuscript's origin; however, this is not always achievable. The ideal reviewer should be totally objective: however, since the reviewers are usually selected because they are engaged in similar scientific endeavors to those of the author(s), they are almost always competitors or teammates. It is difficult for even the most conscientious reviewer to enthusiastically assess work that might undermine his own efforts or to degrade a report supporting his own ideas.

The review mechanism is not meant to achieve perfection, but rather to improve quality and accuracy (4). Wide distribution of scientific papers perhaps is more important for peer review than is the manuscript review process that journals employ. While most editors faithfully and meticulously depend on systems of peer review, the system as it applies to the health sciences has rarely if ever been subjected to the type of scrutiny that it uses itself (6). Despite the problems inherent in the journal review process, editors depend on these systems. Today, peer review is an integral part of scientific methodology and science itself. Although the beginnings of this system were over 300 years ago, much is still poorly understood about the process and its reliability in screening out poorly conceived, poorly designed, poorly executed, unoriginal and uninterpretable manuscripts.

The study reported here was designed to assess the reliability and consistency of reviewers' responses with regard to manuscripts submitted for publication to a leading international dental journal.

## Material and methods

The study was in three parts: (1) inter rater reliability and agreement using both two and three reviewers; (2) re-review of previously reviewed, but unpublished, manuscripts, and (3) a questionnaire regarding editors' reviewing philosophy and outcomes.

### 1) Inter rater reliability and agreement:

Reviews from a total of 115 consecutively manuscripts submitted to the oral and maxillofacial radiology section of *Oral Surgery, Oral medicine, Oral Pathology, Oral Radiology and Endodontics* we-

re studied to determine the consistency with which different referees responded to the same question on the reviewers' report form. Of these, 78 were rated full manuscripts were included in the complete study. The ten areas addressed were: type of manuscript (e.g. scientific, case report, forum note), subject relevance to the journal section concerned, originality, hypothesis, statistical methods, results, conclusions, scientific merit, accept/reject, and priority. The manuscripts were divided into two groups based upon the number of reviews returned: 44 with three reviews and 34 with only two returned reviews at the time of the editor's decision. For statistical analysis the  $\chi^2$  statistic was used. Cramer's phi prime was used to index the magnitude of relationships with 0.40 set as the minimum level of practical importance (11). The Fisher exact test was used to test contingency tables where the cell frequencies were small and the expected frequencies were less than five. The *a priori*  $\alpha$  was set  $p < 0.05$ .

A table of random numbers was used to select two out of three of the reviews for the 44 having three reviews (12). This was then used to assess whether changes might occur in the publication decision should only two reviewers be utilized.

### 2) Re-review of manuscripts:

Ten manuscripts accepted for publication (but unpublished) and ten rejected manuscripts were sent out to second panels of reviewers who were not told that the manuscripts had already been reviewed. The statistical methods employed were the same as those for the first part of the study.

### 3) Questionnaire:

Journal peer review methods and outcomes (e.g. rejection rate, publication time) were assessed using a questionnaire sent to the editors of the *Journal of the American Dental Association*, the *American Journal of Orthodontics and Dentofacial Orthopedics*, the *British Dental Journal*, *Dentomaxillofacial Radiology*, the *Journal of Oral and Maxillofacial Surgery*, *Pediatric Dentistry*, and to the editors of the oral and maxillofacial surgery, the oral pathology, and the endodontics sections of *Oral Surgery*, *Oral Medicine*, *Oral Pathology*, *Oral Radiology and Endodontics*. The specific questions asked

were: (1) What percentage of submitted manuscripts are accepted and published? (2) What percentage of accepted manuscripts are "scientific"? (3) How many reviewers are used? (4) How long is the review process?. (5) What is the average time from submission to publication? (6) What is the average time from acceptance to publication? (7) What are the three most important criteria used by reviewers to evaluate manuscripts?

## Results

### 1) Inter rater reliability and agreement:

The summary tables are read from left to right with the column(s) representing the individual statistic used. With three reviews available, the questions of originality ( $p < 0.001$ ), scientific merit ( $p < 0.001$ ) and accept/reject ( $p < 0.001$ ) were statistically significant for inter rater reliability with Cramer's phi prime indicating a strong association (Table 1). The question of subject relevance was also

Table 1. *Inter rater reliability with three initial reviews*  
Tablica 1. *Odnos pouzdanosti ocjene triju zasebnih recenzija*

CRITERIA	$\chi^2$	Cramer's phi prime	$p \leq$	Fisher Exact Test $p =$
Category	6.26	0.27	0.25	N/A
Agreement	N/A	N/A	N/A	0.07
Relevance	8.19	0.26	0.05	N/A
Agreement	N/A	N/A	N/A	0.05
Originality	21.18	0.41	0.001	N/A
Agreement	1.27	0.17	0.75	N/A
Hypothesis	1.12	0.11	0.75	N/A
Agreement	0.21	0.07	0.90	N/A
Statistics	6.07	0.27	0.25	N/A
Agreement	N/A	N/A	N/A	0.05
Results	2.64	0.16	0.50	N/A
Agreement	1.52	0.19	0.50	N/A
Conclusion	1.13	0.11	0.50	N/A
Agreement	N/A	N/A	N/A	0.40
Science	24.20	0.46	0.001	N/A
Agreement	1.69	0.20	0.50	N/A
Acceptance	30.09	0.50	0.001	N/A
Agreement	8.42	0.44	0.025	N/A
Priority	3.01	0.26	0.25	N/A
Agreement	N/A	N/A	N/A	0.12

N/A = Statistical test not appropriate  
N/A = statistički test nije prikladan

statistically significant ( $p < 0.05$ ); however, the association was weak with a Cramer's phi prime of 0.26. The accept/reject decision showed significance for inter rater agreement ( $p < 0.025$ ). The other questions evaluated were not statistically significant for inter rater agreement.

With only two reviews received (Table 2), there was a strong association for nine of the ten questions studied; however, the chi square results indicated that the reliability of the directional differences are weak for six of these nine questions. Statistical significance combined with strong association was found for the evaluations of scientific merit, accept/reject and priority.

Table 2. *Inter rater reliability with two initial reviews*  
Tablica 2. *Odnos pouzdanosti ocjene dviju zasebnih recenzija*

CRITERIA	$\chi^2$	Cramer's phi prime	$p \leq$	Fisher Exact Test $p =$
Category Agreement	4.62 N/A	0.54 N/A	0.50 N/A	N/A 0.79
Relevance Agreement	3.11 N/A	0.47 N/A	0.10 N/A	N/A N/A
Originality Agreement	4.67 N/A	0.54 N/A	0.10 N/A	N/A 0.21
Hypothesis Agreement	2.50 N/A	0.46 N/A	0.50 N/A	N/A 0.57
Statistics Agreement	2.36 N/A	0.46 N/A	0.75 N/A	N/A 0.29
Results Agreement	1.38 N/A	0.33 N/A	0.75 N/A	N/A 0.21
Conclusion Agreement	7.64 N/A	0.83 N/A	0.10 N/A	N/A 0.57
Science Agreement	6.07 N/A	0.64 0.20	0.05 0.50	N/A N/A
Acceptance Agreement	8.12 8.42	0.74 0.44	0.05 0.025	N/A N/A
Priority Agreement	6.74 N/A	0.67 N/A	0.05 N/A	N/A 0.12

N/A = Statistical test not appropriate  
N/A = statistički test nije prikladan

Randomly selecting two out of the three reviews for the 44 manuscripts having three reviews, resulted in 20 out of 22 accepted papers being again accepted. However, for the 22 manuscripts that were rejected initially three was only complete concurrence for two if one used two randomly selected reviews. It would appear that two reviewers are sufficient for

Table 3. *Journal comparison*  
Tablica 3. *Usporedba časopisa*

Journal	Accept rate (%)	Scientific (%)	Reviewers (No.)	Review time	Submit-to-publication	Accept-to-publication
JADA	34	100	2-3	1-3 mth	2-9 mth	2-9 mth
AJO	80	75	2	1 mth	18-24 mth	12-16 mth
BDJ	50	90	2	3 mth	10 mth	4-5 mth
DMFR	75	80	2	4-6 mth	6-8 mth	6 mth
JOMS	65	33	2	2-3 mth	8-12 mth	6 mth
PediatDent	68	88	2	2 mth	9 mth	6 mth
OOO: OS	50	>50	3	1 wk	N/A	3 mth
OOO: OP	50	100	2	1-2 mth	14 mth	12 mth
OOO: End	30	50	2	1 mth	9 mth	6 mth
OOO: OR	60	59	3	1 mth	13 mth	12 mth

papers that are either of a higher quality or less controversial; but that three reviewers are to be preferred *in toto*.

## 2) Re-review of manuscripts:

As with all manuscript reviewed by three reviewers, the original reviews for the 20 papers sent out for reappraisal showed statistical significance in inter rater reliability for the questions of originality and scientific merit and for the accept/reject decision. The Fisher exact test also identified relevance, hypothesis and priority as significant for inter rater reliability in the selected sample. The re-reviews provided significant inter rater reliability only for originality and priority, with a strong association (Cramer's phi prime = 0.43) for originality. Using a majority decision by the second sets of reviewers, eight out of 10 accept/reject decisions were unaltered by the re-review for previously accepted manuscripts. Six of 10 rejected manuscripts would still have been rejected using this same criterion.

## 3) Questionnaire

Table 3 details the responses of the various editors who were questioned. The acceptance rate for submitted papers ranged from 30 to 80%, and of accepted papers 50% or more were rated "scientific". For the most part only two reviews were used. The review process ranged from one week to six months with accepted manuscripts being published form

two to 24 months following submission or two to 16 months following acceptance. The criteria for acceptability specified most frequently were "relevance to the journal", "scientific validity" and "methodological soundness", followed by "originality" and "style".

## Discussion

The literature on reliability in the journal peer review process is truly inadequate in light of its tremendous impact on the scientific community and on science in general. The few studies that have evaluated agreement found that the concurrence between two reviewers was largely little better than chance (2,6,7,13-19). The present study concurs with this finding.

The acceptance criteria perceived as being most important according to the responses to the editors' questionnaire were "relevance to the journal", "scientific validity" and "methodological soundness". This compares to the present study in which "originality" and "scientific merit", were found to be the significant predictors of outcome.

Most journals use only two reviewers. This practice was validated for acceptable manuscripts but

produced inconsistency of outcomes for rejected papers. Based on the results of resubmission of previously accepted manuscripts, there was a moderately high degree of validity for the review process; however, the validity for rejected manuscripts was only 60%. Papers were less likely to be rejected outright with two reviews than with three. With two reviews there were many split votes and no possibility of a majority decision being achieved.

Rejection by one journal does not necessarily preclude ultimate publication somewhere else. A review of the *Dental Index* showed that 15 of 30 full manuscripts that were rejected for publication in the oral and maxillofacial radiology section of *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics* were actually published elsewhere within a period of two years following rejection. On the 15 published elsewhere, only six demonstrated significant changes in the text when compared to the originally rejected manuscript. On the 15 published elsewhere, seven were published in national or regional journals within the USA, four were published in other national journals and four were published in other sections of *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics*.

---

## ZNANSTVENA RECENZIJA: OCJENA VRIJEDNOSTI

### Sažetak

**Cilj istraživanja:** Urednici časopisa ovise o sustavu recenzija kako bi se zaštitili od loše zamišljenih, loše izvedenih i neoriginalnih rukopisa. Cilj ove studije bio je procijeniti pouzdanost i dosljednost recenzenta prema rukopisima poslanim kako bi se objavili u vodećem međunarodnom stomatološkom časopisu.

**Postupak:** Istraživana su tri elementa u postupku recenzije; poglavito: (1) proučeni su izvještaji recenzenta o rukopisima poslanim odjeku za radiologiju časopisa *Oral Surgery, Oral medicine, Oral Pathology, Oral Radiology and Endodontics*, kako bi se ustanovila dosljednost triju recenzija za svaki od 78 redom pristiglih radova označenih kao "full papers"; (2) deset već prihvaćenih i deset prethodno odbijenih radova poslano je na ponovnu recenziju različitim recenzentima koji nisu bili obaviješteni o prethodnim recenzijama; i (3) poslani su upitnici

---

Address for correspondence:  
Adresa za dopisivanje:

Allan G. Farman, BDS,  
PhD(odont), DSc(odont)  
Professor and Head: Division  
of Radiology and Imaging  
Sciences  
School of Dentistry  
The University of Louisville  
Louisville  
Kentucky 40292, USA  
Tel/Fax: +1(502)852.1241  
e.mail:agfarm01@ulkyvm.louisville.edu

urednicima vodećih stomatoloških časopisa kako bi se usporedili kriteriji i rezultati recenzija.

**Rezultati:** Recenzenti su bili dosljedno pouzdani u odgovorima o izvornosti, znanstvenom doprinosu i odluci o prihvatljivosti rukopisa. Od deset prethodno prihvaćenih rukopisa osam je bilo ponovno prihvaćeno nakon dvostruke slijepe recenzije. Od deset rukopisa koju su prethodno bili odbijeni šest je po odluci drugih referenata bilo ponovno odbijeno. Korištenje dva recenzenta bilo je punovrijedno za prihvaćene članke, ali da bi se postigla pravovaljanost kod odbijenih rukopisa potrebna su barem tri recenzenta. Urednici izabranih časopisa izvijestili su o prihvaćanju radova u 30 do 80% slučajeva, a vrijeme do tiska radova bilo je od 3 do 24 mjeseca. Viši postotak prihvaćenih radova općenito je bio povezan s dužim razdobljem do objave radova.

**Zaključak:** Recenzija u časopisu nije savršeno pouzdana; međutim, ona služi da bi se i smanjio broj loše zamišljenih i loše izvedenih znanstvenih radova.

Ključne riječi: uređenje časopisa; literatura, stomatološka; literatura, znanstvena; recenzija; tiskovina; znanstveni postupak

## References

- ZUCKERMAN H, METRON RK. Patterns of evaluation in science: institutionalization, structure and function of the referee system. *Minerva* 1972;9:66-100.
- BAILAR JC, MOSTELLAR F. Medical uses in statistics. *New Eng J Med* 1986;349-369.
- CULLITON BJ. NIH ses plagiarism in vision paper. *Science* 1989;244:120-122.
- SUN M. Peer review comes under peer review. *Science* 1989;244:910-912.
- RELMAN AS. Are journals really quality filters? In: Warren KS. *Coping with the biomedical literature explosion: a qualitative approach*. New York: The Rockefeller Foundation, 1978;54-60.
- SCOTT WA. Interreferee agreement on some characteristics of manuscripts submitted to *Journal of Personality and Social Psychology*. *Amer Psychol* 1974;29:698-702.
- PATTERSON K, BAILAR JC. A review of journal peer review. In: Warren KS. *Selectivity in information systems: survival of the fittest*. New York: Praeger, 1985;64-82.
- PETERS DP, CECI SJ. Peer-review practices of psychology journals: the fate of published articles submitted again. *Behaviour Brain Sci* 1982;5:185-186.
- PRESSOR S. Collaboration and quality of research. *Soc Study Sci* 1980;10:95-101.
- GORDON MD. A critical reassessment of the inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication. *Scientometrics* 1980;1:193-210.
- COHEN J. *Statistical power analysis for the behavioral sciences*. 2<sup>nd</sup> ed. Hillsdale, NJ: Lawrence Erlbaum 1988;215-271.
- ZAR JH. *Biostatistical analysis*, 2<sup>nd</sup> ed. Englewood Cliffs, NJ: Prentice Hall 1985.
- JUHASZ S. Acceptance and rejection of manuscripts. *IEEE Trans Prof Comm* 1975;18:175-185.
- CICCHETTI DV, ERON LD. The reliability of manuscript reviewing for the *Journal of Abnormal Psychology*. *Proceedings of the Social Statistics Section Washington D.C., American Statistical Association* 1979;569-600.
- SCARR S, WEBBER BL. The reliability of reviews for the *American Psychologist*. *Amer Psychol* 1978;33:935.
- CICCHETTI CV. Reliability of reviews for the *American Psychologist*: a biostatistical assessment of the data. *Amer Psychol* 1980;35:300-303.
- GOTTFREDSON SD. Evaluating psychological research reports: dimension, reliability and correlates of quality judgements. *Amer Psychol* 1978;33:920-934.
- WILSON JD. Peer review and publication. *J Clin Invest* 1978;61:1697-1701.
- BEYER JM. Editorial policies and practices among leading journals in four scientific fields. *Social Quart* 1978;19:68-88.