

Peer Review in a Small and a Big Medical Journal: Case Study of the *Croatian Medical Journal* and *The Lancet*

Ana Marušić, Ivan Krešimir Lukić, Matko Marušić, David McNamee¹, David Sharp¹, Richard Horton¹

Croatian Medical Journal, Zagreb, Croatia; and ¹The Lancet, London, UK

Aim. To compare reviewers' recommendations and editorial decisions in *The Lancet*, a high-impact journal, and *Croatian Medical Journal (CMJ)*, a small general medical journal.

Method. Case study of research manuscripts submitted to the *CMJ* (n = 140 manuscripts; 308 review forms) and a sample of similar manuscripts submitted to *The Lancet* (n = 141; 348 review forms) in 1999 and the first half of 2000. Reviewers' recommendations and their influence on editorial decisions on manuscripts were analyzed by logistic regression. Agreement between reviewers was assessed by the kappa statistic.

Results. Although reviewers' scores were identical in the two journals (median = 3 for both journals, range 0 to 5), *Lancet* reviewers more often recommended rejection than *CMJ* reviewers (44% vs 17%; chi-square = 52.1, p = 0.029), and agreed best on rejection (kappa = 0.29 [95% CI = 0.04 to 0.53] vs kappa = -0.04 [-0.45 to 0.36] for *CMJ*). *Lancet* editors were even stricter than their reviewers and accepted for publication only 53% [95% CI = 37 to 68] of manuscripts graded acceptable by the reviewers, compared with 85% [73 to 91] for *CMJ* editors (chi-square = 10.0, p = 0.001). For nine questions about manuscript quality in the review form, multiple regression analysis showed significant association between editorial decision and reviewers' scores for the suitability of research design (B = 0.70, OR = 2.01, 95% CI = 1.40 to 2.89, p < 0.001) and discussion of systematic/random error (B = 0.32, OR = 1.38 [1.03 to 1.85], p = 0.031) for *The Lancet*, and scores for novelty of information (B = 0.56, OR = 1.75 [1.35 to 2.27], p < 0.001) for the *CMJ*.

Conclusion. Reviewers of big journals, such as *The Lancet*, are stricter in their recommendations reviewers of a smaller journal, such as the *CMJ*. *The Lancet* editors rely on reviewers to identify methodologically superior studies, whereas *CMJ* editors look for the novelty of information in a manuscript.

Key words: Croatia; cross-cultural comparison; cultural diversity; journals, scientific; manuscripts, medical; peer review, publishing; peer review, research; publications, medical; United Kingdom

Regardless of their size and importance, most journals depend on peer review in making decisions on publishing research data. The size and importance of the journal determines the number and quality of submissions (1), and possibly influences its peer review process. We compared reviewers' recommendations and their influence on editorial decisions in *The Lancet*, a high-impact journal, and the *Croatian Medical Journal (CMJ)* a small general medical journal outside mainstream science. Identity of the review forms in the two journals enabled us to make quantitative comparison between the two journals.

Methods

In 1999 and the first half of 2000 (January-June), *The Lancet* received 8,852 manuscripts, and published 730 (8%). *CMJ* published 118 (44%) of 269 submitted manuscripts during the same period. Both journals use almost identical review forms for research articles, based on a validated multi-item rating scale (2). We excluded submissions that were either not the type published by one of the journals or had a different review process: manuscript with statistical reviews, review articles, research letters, fast

track manuscripts, and appeals on rejected manuscripts. This left 140 manuscripts out of the total 269 (52%) for *CMJ*. To make a comparable sample of *Lancet* articles, we systematically chose 141 articles to span the same period (every 18th article out of 1,551 eligible articles from 1999 and every 11th article out of 596 from 2000).

To explore the influence of reviewer's final suggestion and reviewer's grade of the individual aspect of the manuscript (independent variables) on the editorial decision (dependent variable), we used the logistic regression analysis or multiple logistic regression analysis, respectively (3). Although the independent variables were categorical, the underlying continuity may be assumed (3). The assumptions of the analyses were met (4). The forward stepwise logistic model was used in calculations. For the analysis of influence of reviewer's final suggestion on editorial decision, we used the indicator-variable coding scheme, with the "accept" grade as the reference category. The regression coefficient "B" is the average amount the dependent variable increases when the independent variable increases one unit. In other words, B coefficient is the slope of the regression line: the larger the B, the steeper the slope, the more the dependent changes for each unit change in the independent variable. To compare the relative importance of independent variables in terms of their effect on the dependent variable, we used the odds ratio.

The agreement between the reviewers was tested using kappa statistics (5). The ordinal nature of the data in this study restricted the use of methods for multiple rater comparisons, such as intraclass correlation coefficient (5). As the kappa measures concordance between two raters using ordinal data, we analyzed the manuscripts with two and only two correctly filled-in review forms (55 manuscripts for *The Lancet* and 52 for the *CMJ*). The kappa (κ) value of 1 indicates perfect agreement, values between 0.81 and 1.00 very good agreement, 0.61-0.80 good, 0.41-0.60 moderate, 0.2-0.4 fair, and 0-0.2 poor agreement. Negative κ values indicate agreement worse than chance. Weighted kappa takes into account the seriousness of the disagreement.

The software for statistical analysis was SPSS for Windows v. 7.5 (SPSS Inc., Chicago, IL, USA).

Results

The 141 *Lancet* manuscripts were sent out to 351 reviewers (median = 3 per manuscript, range 1 to 4). The 140 *CMJ* manuscripts were sent out to 392 reviewers (median = 3, range = 1 to 6). Reviewers' response rate was 99% for *The Lancet* and 79% for *CMJ* (chi-square = 72.5, d.f. = 1, $p < 0.001$). Recommendation for publication (accept, minor revision, major revision, or reject) was available on 300 forms out of 348 (86%) returned to *The Lancet*, and 280 out of 308 (91%) for *CMJ*.

Reviewers for the two journals awarded similar grades in the summary score of the manuscripts (median = 3, range 0 to 5, on a 0 to 5 scale for both journals) and the summary score showed good correlation with the reviewers' recommendation for publication in both journals (Spearman's ρ for *The Lancet* was 0.78 and 0.83 for the *CMJ*, $p < 0.001$ for both journals). However, the four recommendations ("reject", "major revision", "minor revision", and "accept") given by the *Lancet* and *CMJ* reviewers were not distributed in the same way (Fig. 1 and Table 1). The *Lancet* reviewers significantly more often sug-

gested rejection (44% vs 17% for *CMJ* reviewers, chi-square = 52.1, d.f. = 3, $p = 0.029$), whereas the recommendations given by the *CMJ* reviewers were clustered in the middle (major and minor revision).

Agreement between reviewers in their recommendation for publication was analyzed for manuscripts with two reviews (55 for *The Lancet* and 52 for the *CMJ*). The overall distribution of the reviewers' recommendation on publication for these manuscripts (data not shown) was very similar to that for the whole manuscript sample. The weighted κ for reliability of reviewers' recommendations was poor: 0.09 [95% CI = -0.11 to 0.28] for *The Lancet* and 0.19 [0.02 to 0.37] for the *CMJ*. *Lancet* reviewers agreed best on the rejection of a manuscript: the reliability of the "reject" grade vs others

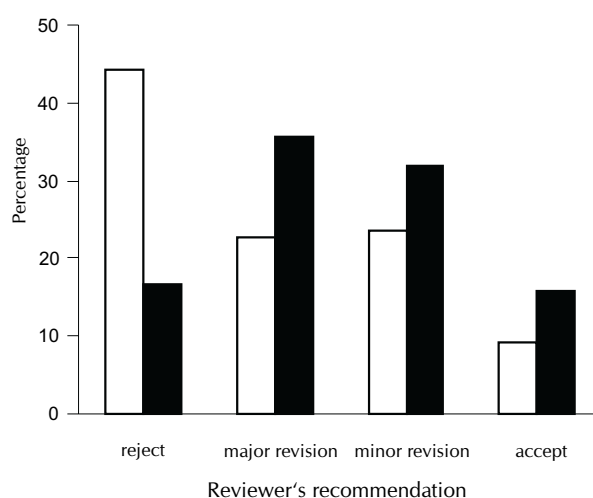


Figure 1. Distribution of reviewers' recommendation for all manuscripts in *The Lancet* (open bars; n = 300 reviews) and the *Croatian Medical Journal* (closed bars; n = 280).

Table 1. Editorial decision on acceptance for publication (n,%) for manuscripts in *The Lancet* and the *Croatian Medical Journal* (*CMJ*) according to reviewers' recommendation

Reviewer's recommendation	Editorial decision				Total	
	accept		reject		<i>Lancet</i>	<i>CMJ</i>
Reject	6 (10)	11 (6)	127 (52)	36 (35)	133 (44)	47 (17)
Major revision	15 (26)	60 (34)	53 (22)	40 (38)	68 (23)	100 (36)
Minor revision	21 (36)	69 (39)	50 (21)	20 (19)	71 (24)	89 (32)
Accept	16 (28)	36 (21)	12 (5)	8 (8)	28 (9)	44 (16)
Total	58 (100)	176 (100)	242 (100)	104 (100)	300 (100)	280 (100)

Table 2. Distribution of reviewers' scores of the quality of manuscripts submitted for publication to *The Lancet* (321 reviews) and the *Croatian Medical Journal* (*CMJ*; 292 reviews)^a

Quality assessment item ^b	<i>Lancet</i> ^c					<i>CMJ</i> ^c						
	0	1	2	3	4	5	0	1	2	3	4	5
The subject addressed in this article is worthy of investigation	3	10	6	30	104	157	13	4	11	37	99	97
The hypothesis is clearly stated	12	10	20	57	91	114	39	19	21	50	56	56
The most important previous studies have been cited as far as I know	6	18	23	48	96	99	10	20	19	45	75	66
Information presented was new (or a replication deserved to be done)	9	15	18	62	106	96	20	21	25	52	80	50
The research design was suitable	25	27	45	55	74	70	25	27	30	46	68	39
The methods were described specifically enough to be evaluated	20	23	46	41	99	77	25	24	27	48	65	46
The discussion addresses sources of systematic and random error	37	29	55	64	71	37	37	27	33	45	51	17
The conclusions were supported by data	25	32	49	70	77	53	27	35	38	34	68	49
The summary accurately reflects the content of the paper	7	11	19	72	109	89	14	16	29	47	83	71
Summary score	7	33	38	71	49	37	19	24	33	61	75	30

^aThe analysis included review forms with a score for at least one assessment item.

^bEach item was addressed with the question: "To what extent does the article meet this criterion?"

^cResponses offered to reviewers: 0 – fails by a large amount; 1 – fails by a moderate amount; 2 – fails by a small amount; 3 – succeeds by a small amount; 4 – succeeds by a moderate amount; 5 – succeeds by a large amount. The answers N – not applicable, and D – do not know are not shown (did not exceed 5% for any question in either of the two journals). The numbers across each row do not add up to the total number of reviews because some reviewers did not give scores to all assessment items.

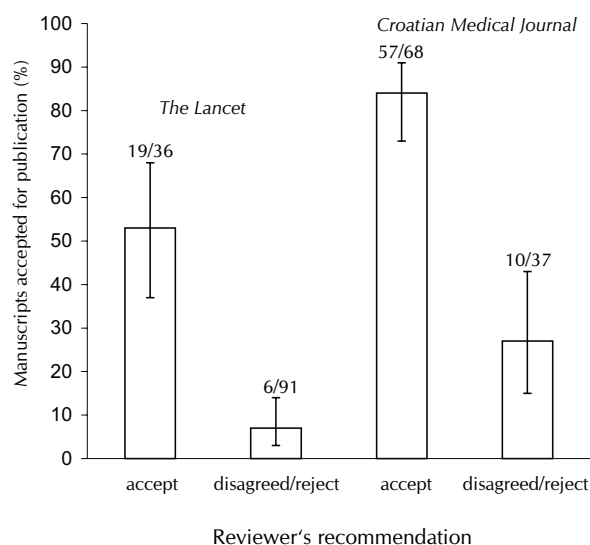


Figure 2. Reviewers' recommendations for manuscripts in *The Lancet* and the *Croatian Medical Journal*, and editorial decision on acceptance for publication. Manuscripts with two or more valid review forms were used in analysis (127 for *The Lancet* and 105 for *Croatian Medical Journal*). Bars represent 95% confidence intervals. Reviewers' agreement on acceptability of a manuscript was considered when they recommended any of the following grades: "accept", "minor revision", or "major revision". Disagreement on rejection was defined as the recommendation to "reject" by one or more but not all reviewers.

for *The Lancet* was $\kappa=0.29$ [0.04 to 0.53], compared with *CMJ* $\kappa=-0.04$ [-0.45 to 0.36].

We next assessed how the agreement of reviewers influenced the editorial decision (Table 1). The "reject" grade had the greatest influence on the editor's decision to reject a manuscript in both journals (*Lancet*, $B=-3.34$, $OR=0.03$ [95% CI 0.01 to 0.11]; *CMJ*, $B=-2.66$, $OR=0.07$ [0.02 to 0.19]; $p<0.001$ for both journals). We next looked at editorial decision for two groups of manuscripts: those for which the reviewers agreed to be suitable for publications (with or without any kind of revision) and those for which they disagreed or all recommended rejection (Fig. 2). *Lancet* editors were stricter than *CMJ* editors in their final decision, rejecting 47% of the manuscripts graded acceptable by the reviewers, compared with 16% for the *CMJ* editors (chi-square=10.0, d.f. = 1, $p=0.002$).

Separate multiple logistic regression analysis was carried out to analyze which of the 9 questions on the quality of a manuscript (Table 2) had the greatest influence on the editorial decision. For *The Lancet*, reviewers' assessment of the research design had a significant influence on the editors decision (question: "suitability of research design" $B=0.70$, $OR=2.01$ [95% CI = 1.40 to 2.89], $p<0.001$; and question: "addressing sources of systematic and random error", $B=0.32$, $OR=1.38$ [1.03 to 1.85], $p=0.031$). For *CMJ* editors, only the score on the novelty of information had a significant influence on their decision ($B=0.56$, $OR=1.75$ [1.35 to 2.27], $p<0.001$).

Discussion

Our case study showed that reviewers and editors of a big and a small journal differed in their recommendations or decisions, respectively. The distribution of reviewers' recommendations suggest that the reviewers for *The Lancet* perceived it as a rigorous journal that sought critical reviews, and more often suggested rejection. *CMJ* reviewers tried to identify valuable elements even in weak manuscripts, choosing to suggest revision rather than rejection. The editors of the two journals also took on different roles. *The Lancet* editors may be even stricter than their reviewers, looking for the methodologically superior studies. They acted as true "gatekeepers" (6). By contrast, both *CMJ* reviewers and editors seem to take great efforts to publish potentially interesting articles, acting as shepherds in their scientific community (7).

Our study also showed that the reliability of peer review was poor in general medical journals, regardless of their size and importance. This agrees with studies of peer review reliability in non-medical fields, such as sociology and psychology (8). Agreement between reviewers for a general medical journal may not be expected, with the argument that the editor often deliberately chooses reviewers from different fields and different points of view (9). However, agreement between reviewers is also poor in different and very specialized fields in the broader biomedical area, such as neuroscience (10), psychiatry (11), and basic biomedical science (12). Poor reliability of peer review regardless of the size and importance of the journal suggests that it may perhaps be inherent to the peer review process, in which reviews serve to supplement and complement the final editorial decision (13).

As a first exploration into the cross-cultural differences between small and big journals, this case study has a number of limitations, including the difference in the number and quality of submitted manuscripts, as well as referees' response rate and possible differences in the quality of reviews. However, our study provides an insight into the different roles of peer review in influencing editorial decisions in general medical journals at opposite ends of international prestige and influence. More research is needed into these sociological and behavioral issues of peer review and editorial decision making in journals from different scientific communities and for different audiences.

Acknowledgment

AM was supported by a Wellcome Trust Travel Grant. We thank Prof Alexander Walker, Ingenix, Newton Lower Falls, MA, USA; and Dr Christopher Palmer, Centre for Applied Medical Statistics, University of Cambridge, UK, for their valuable advice on research design and statistical analysis.

References

- Marušić A, Marušić M. Small scientific journals from small countries: breaking from the vicious circle of inadequacy. *Croat Med J* 1999;40:508-14.
- Strayhorn J Jr, McDermott JF Jr, Tanguay P. An intervention to improve the reliability of manuscript reviews for

- the Journal of the Academy of Child and Adolescent Psychiatry. *Am J Psychiatry* 1993;150:947-52.
- 3 SPSS Inc. Staff. SPSS professional statistics 7.5. Chicago (IL): SPSS Inc. 1997.
 - 4 Garson GD. PA 765 Statnotes: An online textbook. [cited 2001 Aug 24]. Available from: <http://www2.chass.ncsu.edu/garson/pa765/statnote.htm>. Accessed: March 25, 2002.
 - 5 Altman DG. Practical statistics for medical research. London: CRC Press; 1991. p. 403-9.
 - 6 Crane D. The gatekeepers of science: some factors affecting the selection of articles of scientific journals. *Am Sociol* 1967;2:195-201.
 - 7 Marušić M, Marušić A. Good editorial practice: editors as educators. *Croat Med J* 2001;42:113-20.
 - 8 Cicchetti DV. The reliability of peer review for manuscripts and grant submissions: a cross-disciplinary investigation. *Behav Brain Sci* 1991;14:179-86.
 - 9 Fletcher RH, Fletcher SW. Who's responsible? *Ann Intern Med* 1993;118:645-6.
 - 10 Rothwell PM, Martyn CN. Reproducibility of peer review in clinical neuroscience. Is agreement between reviewers any greater than would be expected by chance alone? *Brain* 2000;123:1964-9.
 - 11 Howard L, Wilkinson G. Peer review and editorial decision-making. *Br J Psychiatry* 1998;173:110-3.
 - 12 Scharschmidt BF, DeAmicis A, Bacchetti P, Held MJ. Chance, concurrence, and clustering. Analysis of reviewers' recommendations on 1,000 submissions to The Journal of Clinical Investigation. *J Clin Invest* 1994;93:1877-80.
 - 13 Bornstein RF. The predictive validity of peer review; a neglected issue. *Behav Brain Sci* 1991;14:138-9.

Received: March 14, 2002

Accepted: March 26, 2002

Correspondence to:

Ana Marušić
Croatian Medical Journal
Zagreb University School of Medicine
Šalata 3b
10000 Zagreb, Croatia
marusica@mef.hr