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 variables), 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 suggested 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

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

<table>
<thead>
<tr>
<th>Reviewer’s recommendation</th>
<th>CMJ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lancet</strong></td>
<td><strong>CMJ</strong></td>
</tr>
<tr>
<td>Reject</td>
<td>11 (6)</td>
</tr>
<tr>
<td>Major revision</td>
<td>60 (34)</td>
</tr>
<tr>
<td>Minor revision</td>
<td>21 (12)</td>
</tr>
<tr>
<td>Accept</td>
<td>36 (21)</td>
</tr>
<tr>
<td>Total</td>
<td>176 (100)</td>
</tr>
</tbody>
</table>

**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)

<table>
<thead>
<tr>
<th>Quality assessment item</th>
<th>Lancet</th>
<th>CMJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>The subject addressed in this article is worthy of investigation</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>The hypothesis is clearly stated</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>The most important previous studies have been cited as far as I know</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Information presented was new (or a replication deserved to be done)</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>The research design was suitable</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>The methods were described specifically enough to be evaluated</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>The discussion addresses sources of systematic and random error</td>
<td>37</td>
<td>29</td>
</tr>
<tr>
<td>The conclusions were supported by data</td>
<td>25</td>
<td>32</td>
</tr>
<tr>
<td>The summary accurately reflects the content of the paper</td>
<td>7</td>
<td>11</td>
</tr>
</tbody>
</table>

The analysis included review forms with a score for at least one assessment item. Each item was addressed with the question: “To what extent does the article meet this criterion?”

The analysis included review forms with a score for at least one assessment item. Each item was addressed with the question: “To what extent does the article meet this criterion?”

The analysis included review forms with a score for at least one assessment item. Each item was addressed with the question: “To what extent does the article meet this criterion?”

The analysis included review forms with a score for at least one assessment item. Each item was addressed with the question: “To what extent does the article meet this criterion?”

Figure 1. Distribution of reviewers’ recommendation for all manuscripts in The Lancet (open bars; n = 300 reviewers) and the Croatian Medical Journal (closed bars; n = 280).
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. More research is needed into the reliability of manuscript reviews for 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 reviewers 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

2. Strayhorn J Jr, McDermott JF Jr, Tanguay P. An intervention to improve the reliability of manuscript reviews for


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