Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 80, No. 2 (2009) 343–349
DOI: 10.1007/s11192-008-2064-1

The relationship between reviewers' quality-scores and number of citations for papers published in the journal *Physics in Medicine and Biology* from 2003–2005

MICHAEL S. PATTERSON,[a] SIMON HARRIS[b]

[a] *Juravinski Cancer Centre and McMaster University, 699 Concession Street, Hamilton, Ontario, L8V 5C2, Canada*
[b] *Institute of Physics Publishing, Bristol, UK*

For each of the years 2003, 2004, and 2005 the number of citations for individual papers published in *Physics in Medicine and Biology* was compared to the mean quality-score assigned to the manuscript by two independent experts as part of the normal peer review process. A low but statistically significant correlation was found between citations and quality score (1 best to 5 worst) for every year: 2003: –0.227 ($p < 0.001$); 2004: –0.238 ($p < 0.001$); 2005: –0.154 ($p < 0.01$). Papers in the highest quality category (approximately 10 per cent of those published) were cited about twice as often as the average for all papers. Data were also examined retrospectively by dividing the papers published in each year into five citation quintiles. A paper of the highest quality is about ten times more likely to be found in the most cited quintile than in the least cited quintile. By making the assumption that the mean number of citations per paper is a reasonable surrogate for the impact factor, it was also shown that the impact factor for Physics in Medicine and Biology could be increased substantially by rejecting more papers based on the reviewers' scores. To accomplish this, however, would require a reduction in the acceptance rate of manuscripts from about 50 per cent to near 10 per cent.

Introduction

Although citation frequency is often used as an indicator of the quality of a scientific publication, surprisingly few studies have examined its relation to other measures of quality. Nieminen et al. found no significant relationship between the quality of the statistical analysis in papers published in four psychiatry journals and their subsequent citation frequency [1]. West and McIlwaine [2] reported no correlation between citations to 79 papers published in the journal *Addiction* in 1997 and post-publication evaluation of those papers by two independent experts. On the positive side, Bergh et al. [3] observed that articles in *Strategic Management Journal* were more likely to be cited if they employed more rigorous research methodology. Positive correlations have also been reported between citations and other factors that are, at best, tenuously related to quality including number of authors [4, 5], length of article [5, 6], number of

Received February 2, 2008; Published online April 16, 2009

Address for correspondence:
MICHAEL S. PATTERSON
E-mail: mike.patterson@jcc.hhsc.ca

0138–9130/US $ 20.00

references [5], claims of statistical significance [7], sample size [5, 8], impact factor of the journal [8], country of origin [9], and study design [10] (e.g. randomized trial versus case report).

In this paper, we examine the relationship between numerical scores assigned to papers published in the journal *Physics in Medicine and Biology* by two independent experts as part of the peer review process, and the number of citations to those papers in the two to four years following publication. The original motivation for this study was to explore the relationship between the impact factor of the journal and the threshold for acceptance of manuscripts. The impact factor of a journal for a given year is calculated as the number of citations received in that year to papers published in the preceding two years divided by the number of papers published in those two years [11]. As the impact factor is based on citations over a relatively short time window (two years), we focused our analysis on papers published in the recent years 2003, 2004, and 2005. One of the authors (MSP) is a member of the Editorial Board of the journal and the other is the publisher. *Physics in Medicine and Biology* is published by IOP Publishing on behalf of the Institute of Physics and Engineering in Medicine and its scope is the application of theoretical and practical physics to medicine, physiology and biology.

Methods

Papers submitted to *Physics in Medicine and Biology* are rated by two expert reviewers. The reviewers are requested to assign papers a score of 1 to 10 in each of three categories: originality, soundness, and significance. The total scores (out of 30) are averaged and the paper is classified as Q1 (mean score 26-30), Q2 (24-25), Q3 (21-23), Q4 (18-20), or Q5 (less than 18). Manuscripts rated Q1 – Q3 are normally accepted for publication. Those rated Q4 are referred to a member of the Editorial Board for final adjudication and may be rejected. Q5 papers are usually rejected, but some are published when the two reviewer scores are highly disparate and adjudication is performed by a member of the Editorial Board who decides that the paper is of sufficiently high quality. Reviewers are informed of these criteria when they receive a manuscript. For this study, all papers published in 2003, 2004, and 2005 were considered, but numerical scores were not available for about 10 per cent of papers published each year. This is because one of the assigned reviewers provided only a verbal description of the quality of the manuscript that was, nonetheless, adequate for an editorial decision. These papers were excluded from further analysis. For the remainder, the number of citations from publication to December 2007 was obtained from the Science Citation Index of the Institute of Scientific Information (ISI). The correlation between quality rating and number of citations was analyzed for each year of publication independently. This determines whether the reviewers' quality-score prospectively predicts the frequency of citation. Data were also analyzed

"retrospectively" for each publication year to see whether highly cited papers were more likely to have been rated as high quality. The published papers were grouped into quintiles based on the number of citations and the composition of each quintile was examined to assess the proportion of Q1–Q5 papers.

The Editorial Board of any journal is concerned with its impact factor and how it might be increased. One obvious strategy is to use the reviewers' ratings to reject more papers on the assumption that these "marginal" papers would be cited less frequently. To test this, for each year we calculated the average number of citations per paper should the acceptance threshold be raised from its actual value (Q1 to Q5), to successively higher levels where all Q5 papers are rejected, all Q5 and Q4 papers are rejected, and so on.

Results

Table 1 shows the number of papers in each quality category by publication year. Also listed is the total number of citations for those papers, the mean number of citations per paper, and the standard deviation. Note that more than half of published papers fall into Q3 and only about a tenth into the highest Q1 category. The small number of papers in Q4 and Q5 for 2005 was the result of an Editorial Board decision to reject more marginal papers. In Figure 1 the mean number of citations per paper and the standard error in that mean is plotted versus quality-score for each of the three years. Also shown is the regression line for citations versus quality-score for all the papers published in that year. The correlation coefficients for each year are: 2003, –0.227 ($p < 0.001$); 2004, –0.238 ($p < 0.001$); 2005, –0.154 ($p < 0.01$). Note that we have plotted the standard error in the mean to emphasize the differences between groups but the actual distribution of citations within each quality category is quite broad – typically the standard deviation is comparable to the mean as seen in Table 1.

Table 1. Breakdown of published papers by reviewers' quality-score for 2003, 2004, and 2005. For each Q category, the table shows the number of papers (N), the number of citations from publication to December, 2007 (C), the mean number of citations per paper (m) and the standard deviation (σ)

Year	2003				2004				2005			
Category	N	C	m	σ	N	C	m	σ	N	C	m	σ
Q1	19	326	17.2	16.1	43	536	12.5	13.9	39	257	6.6	11.8
Q2	53	623	11.8	19.6	43	334	7.8	11.8	90	430	4.8	5.3
Q3	196	1697	8.7	9.0	205	1121	5.5	5.2	255	931	3.7	3.9
Q4	29	148	5.1	6.1	55	325	5.9	9.7	11	37	3.4	5.3
Q5	10	39	3.9	3.7	44	171	3.9	3.5	3	13	4.3	3.1
All	307	2833	9.2	12.0	390	2487	6.4	8.3	398	1668	4.2	5.6

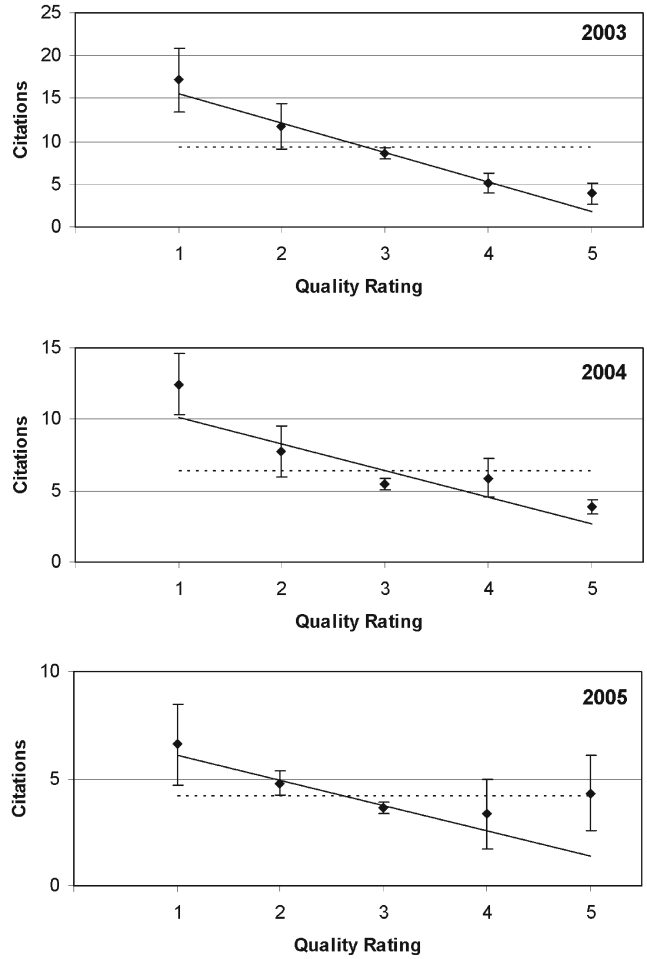

Figure 1. Plot of mean number of citations and standard error in that mean versus quality rating for papers published in 2003, 2004, and 2005. The straight line is a fit to all the data (i.e. 307 points for 2003, not shown) for that year. The horizontal dashed line indicates the average number of citations per paper for all papers published that year

Table 2. Influence of raising the acceptance threshold on the mean number of citations per paper (first entry for each year) and the overall acceptance rate in per cent (second entry for each year)

Categories accepted	2003		2004		2005	
Q1–Q5	9.2	51	6.4	59	4.2	53
Q1–Q4	9.4	49	6.7	52	4.2	53
Q1–Q3	9.9	45	6.8	44	4.2	51
Q1–Q2	13.2	12	10.1	13	5.3	17
Q1	17.2	3	12.5	7	6.6	5

The results of the retrospective analysis are presented in Figure 2. For each year, the composition of the citation quintiles is illustrated, e.g. for 2003, the highest-cited quintile of published papers was composed of 16 per cent Q1 papers, 20 per cent Q2 papers, 61 per cent Q3 papers, 3 per cent Q4 papers, and no Q5 papers.

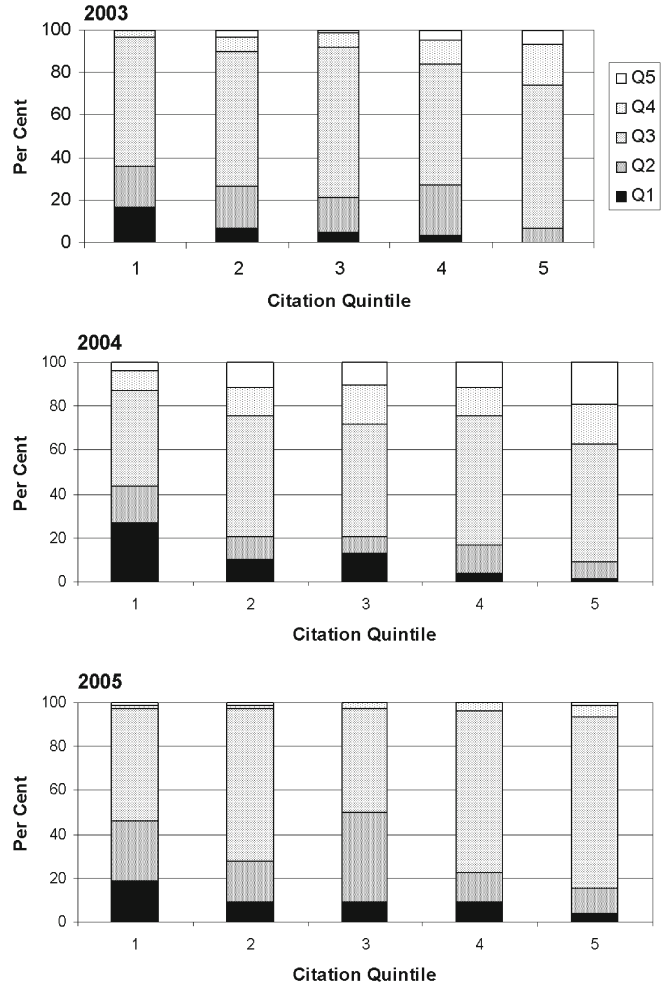

Figure 2. Retrospective analysis of the quality of papers when sorted into citation quintiles. For example, of those papers published in 2003, the fifth of papers that was most cited contained 16 per cent Q1 papers, 20 per cent Q2, etc.

Table 2 shows how the mean number of citations per paper is affected by raising the acceptance threshold from its actual value to successively higher levels. For example, for papers published in 2003, when all 307 papers in Table 1 are included the mean number of citations per paper (till December 2007) is 9.2. The second entry is the acceptance rate – in this case the real rate of 51 per cent. If the 10 Q5 papers are eliminated, the average citation number is 9.4 (not significantly higher) and the acceptance rate is 49 per cent.

Discussion and conclusions

In the prospective analysis, a low but statistically significant correlation was found between reviewers' quality-scores and eventual citation frequency. The correlation for 2005 was somewhat lower than the other two years, possibly because fewer papers of lower quality were published, and because less time was available for consistent citation trends to be established. The mean citation number for papers in the highest quality category is about twice that for all papers. Our results are contrary to those of West and McIlwaine [2] who reported no correlation between citations of articles published in the journal *Addiction* and two expert reviews of quality performed several years after publication. Interestingly, those authors found the correlation between reviewers was only 0.39. Unfortunately, we were not able to analyze the inter-reviewer variation as only the average score is archived. We cannot rule out the possibility that reviewers' scores and citation frequency are both influenced by a third variable such as the reputation of the author(s). It would be interesting to see if the correlation we found exists for journals that use blind review. Note that while the correlation is significant, its value reflects the large spread of the data and low confidence in the ability to predict the citation frequency of an individual paper from its quality-score. For example, a small number of Q1 papers (1 in 2004 and 3 in 2005) have yet to be cited.

In the retrospective analysis a consistent trend was found in the data for all years showing that papers that were cited more often were more likely to be of higher quality according to the reviewers' scores. A paper in the highest-cited quintile is about ten times more likely to be in the highest quality (Q1) category than a paper from the lowest-cited quintile.

Making the reasonable assumption that average citation rate over the first few years after publication is a proxy for impact factor, we showed that the impact factor for *Physics in Medicine and Biology* could likely be increased by using the reviewers' scores to reject papers of lower quality. However, a substantial increase would require a reduction in the acceptance rate from its current 50 per cent to close to 10 per cent. Such a decrease would likely be unacceptable to the research community that supports the journal and to the publisher.

Given the increasing emphasis on citations as a measure of quality, it is somewhat reassuring to find that there is a significant correlation, albeit low, between citations and independent, expert, prospective review. Similar studies for other journals in different disciplines should be relatively easy to perform as long as reviewers are required to rate the quality of accepted manuscripts.

*

The authors thank Christian Box of IOP Publishing for extracting the quality-score data used in this study.

References

1. P. NIEMINEN, J. CARPENTER, G. RUCKER, M. SCHUMACHER, The relationship between quality of research and citation frequency, *BMC Med. Res. Methodol,*. 6 (2006) 42.
2. R. WEST, A. MCILWAINE, What do citation counts count for in the field of addiction? An empirical evaluation of citation counts and their link with peer ratings of quality, *Addiction,* 97 (2002) 501–504.
3. D. D. BERGH, J. PERRY, R. HANKE, Some predictors of SMJ article impact, *Strategic Management Journal,* 27 (2006) 81–100.
4. D. W. AKSNES, Characteristics of highly cited papers, *Research Evaluation,* 12 (2003) 159–170.
5. R. N. KOSTOFF, The difference between highly and poorly cited medical articles in the journal Lancet, *Scientometrics,* 72 (2007) 513–520.
6. J. HUDSON, Be known by the company you keep: Citations – quality or chance?, *Scientometrics,* 71 (2007) 231–238.
7. P. NIEMINEN, G. RUCKER, J. MIETTUNEN, J. CARPENTER, M. SCHUMACHER, Statistically significant papers in psychiatry were cited more often than others, *Journal of Clinical Epidemiology,* 60 (2007) 939–946.
8. M. CALLAHAM, R. L. WEARS, E. WEBER, Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals, *JAMA-Journal of the American Medical Association,* 287 (2002) 2847–2850,
9. G. PASTERKAMP, J. I. ROTMANS, D. V. P. DE KLEIJN, C. BORST, Citation frequency: A biased measure of research impact significantly influenced by the geographical origin of research articles, *Scientometrics,* 70 (2007) 153–165.
10. M. BHANDARI, J. BUSSE, P. J. DEVEREAUX, V. M. MONTORI, M. SWIONTKOWSKI, P. TORNETTA, T. A. EINHORN, V. KHERA, E. H. SCHEMITSCH, Factors associated with citation rates in the orthopedic literature, *Canadian Journal of Surgery,* 50 (2007) 119–123.
11. E. GARFIELD, Journal impact factor: a brief review, *CMAJ,*. 161, 979–980 (1999).