Journal of Clinical Epidemiology 63 (2010) 660−664

Journal of Clinical Epidemiology

Citation analysis of identical consensus statements revealed journal-related bias

Thomas V. Perneger*

Division of Clinical Epidemiology, University Hospitals of Geneva, Geneva, Switzerland

Accepted 30 September 2009

Abstract

Objective: To examine whether the prestige of a journal, measured by its impact factor, influences the numbers of citations obtained by published articles, independently of their scientific merit.

Study Design and Setting: In this cohort study, citation counts were retrieved for articles describing consensus statements that were published in multiple journals and were correlated with the impact factors of the source journals.

Results: Four consensus statements were published in multiple copies: QUOROM (QUality Of Reporting Of Meta-analyses) was published in three journals, CONSORT (CONsolidated Standards Of Reporting Trials) in eight journals, STARD (STAndards for Reporting of Diagnostic accuracy) in 14 journals, and STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) in eight journals. For each consensus statement, the impact factor of the source journal and the number of citations were highly correlated (Spearman correlation coefficients: QUOROM, 1.00; CONSORT, 0.88; STARD, 0.65; and STROBE, 0.81—all $P < 0.02$). When adjusted for time since publication, each logarithm unit of impact factor predicted an increase of 1.0 logarithm unit of citations (95% confidence interval: 0.7−1.3, $P < 0.001$), and the variance explained was 66% (adjusted $r^2 = 0.66$).

Conclusions: The prominence of the journal where an article is published, measured by its impact factor, influences the number of citations that the article will gather over time. Citation counts are not purely a reflection of scientific merit.

Keywords: Research assessment; Research evaluation; Impact factor; Citations, Consensus statements; Bias

1. Introduction

The assessment of the research output of individuals and institutions often relies on citation counts [1−3]. The assumption is that better research is cited more than less-worthy work, but variables other than research quality also influence the number of citations [2]. In particular, the prestige of the journal in which the article is published may drive citations, independent of the scientific worth of the article. If this were the case, the research assessment would be distorted, and researchers might be tempted to devote a disproportionate amount of energy to finding the most prestigious channel of communication in addition to (or perhaps instead of) doing the best possible research.

Several studies have shown that articles published in journals with high impact factors are cited more often than those published in less prominent journals [4−8], but this may simply reflect the ability of the best journals to attract the best articles [4]. By definition, citations to articles drive journals' impact factors; it is whether the opposite relationship is also true which occupies us here. To show whether the prominence of the journal influences citation counts, one would need to randomize articles to high- or low-profile journals or to publish the same article in several journals, neither of which is practical. However, a recent phenomenon in academic publishing provides an opportunity to clarify this issue: the concomitant publication of the same article in several journals, typically a consensus statement about the reporting of a specific type of research. Anyone who wants to cite such a guideline has a choice between several equivalent citations, which differ only in the source journal. If citations to high-profile journals were

Competing interests: The author has no conflict of interest related to this article.

Contributions: T.V.P. conceived the idea of the study, performed the data retrieval and analysis, interpreted the results, and wrote the article.

Funding: The study received no specific funding.

Data access: T.V.P. had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Ethics committee approval: Approval was not sought.

* Corresponding author. Division of Clinical Epidemiology, University Hospitals of Geneva, 2 rue Gabrielle Perret-Gentil, CH-1211 Geneva, Switzerland. Tel.: +41-22-372-9037; fax: +41-22-372-9035.

E-mail address: thomas.perneger@hcuge.ch

0895-4356/$ − see front matter
doi: 10.1016/j.jclinepi.2009.09.012

What is new?

Key findings

- When identical articles published in multiple journals were considered, the impact factor of source journals predicted the number of citations per article.
- The association was strong: if the impact factor was twice as high, the article received twice as many citations.

What this adds to what was known

- By definition, citations to articles drive journal impact factors, but it appears that the converse is also true.

What is the implication and what should change?

- Citation counts are not purely a reflection of scientific merit.
- Journal impact factors are in part self-perpetuated.
- Other indicators of scientific values should be considered.

preferred, this would support the existence of a journal-related citation bias.

In this cohort study, I identified consensus statements that were published in multiple journals and explored the associations between the journals' prominence, measured by their impact factors [9], and the numbers of citations to the published articles.

2. Methods

Consensus statements about research reporting were identified from the Equator Network Web site (http://www.equator-network.org). A consensus statement was eligible if it was published in at least three copies by the same core group of authors. Editorials or commentaries that raised awareness of the statements or provided an endorsement or a critique were not included and neither were translations or secondary articles that included additional explanations of the consensus statement. Finally, the published statements had to have accumulated published citations, which excluded the most recent statements.

Published articles that contained each statement were retrieved both on PubMed and on the Web site of the ISI Web of Knowledge (Thomson Reuters), using the acronym of the statement in the title and the names of the authors. Cited references in the identified articles, the Equator Web site, and the dedicated Web sites maintained for some of the consensus statements were checked to make sure that the list of eligible articles was complete.

The main outcome variable was the number of citations obtained by each article, retrieved from the Web of Knowledge, as of April 15, 2009 (date of last verification). The main independent variable was the impact factor of the journal in which each article was published, for the corresponding year, obtained from the ISI Journal Citation Reports. In the few instances where the official impact factor was unavailable, an unofficial impact factor was either obtained from that journal's Web site or computed from the number of articles and reviews published in 2005−2006 and citations to these articles in 2007. The time elapsed since publication was recorded in months as of April 2009.

Because the distributions of both citation counts and impact factors are skewed toward high values, these variables were transformed using the logarithm in base 10. Natural units were nevertheless used for descriptive statistics. The analysis was stratified by consensus statement. For each statement, I obtained a scatterplot of the logarithm-transformed number of citations by impact factor as well as Pearson and Spearman correlation coefficients (Fig. 1). Finally, the logarithm of citations was predicted from the logarithm of the impact factor and the type of statement or time since publication in a general linear model.

3. Results

The following consensus statements were published three times or more: QUOROM (QUality Of Reporting Of Meta-analyses) was published three times, the revised CONSORT (CONsolidated Standards Of Reporting Trials) eight times, STARD (STAndards for Reporting of Diagnostic accuracy) 14 times, and STROBE (STrengthening the Reporting of OBservational studies in Epidemiology) eight times (Table 1). In addition, SQUIRE (Standards for QUality Improvement Reporting Excellence) was published five times since October 2008, and STREGA (STrengthening the REporting of Genetic Association studies) was published seven times since January 2009, but none of these articles had been cited as of April 15, 2009; hence, these statements were excluded from further analysis.

Citation statistics were available for all 33 articles, and the official impact factor was retrieved for all journals but three, in which case approximate impact factors were obtained (Table 1, footnotes). The median number of citations was greater for older consensus statements, and the median impact factors varied from 1.6 to about 7.5 (Table 2). For each consensus statement, the impact factor of the source journal and the number of citations were highly correlated, for example, the Spearman correlation coefficients ranged from 0.65 for STARD to 1.00 for QUOROM, and all four were statistically significant (Table 2). Univariate linear regression models replicated these findings (Table 2). Each logarithm unit of the impact factor of the journal in which

Table 1
Consensus statements and related publications, with citations and impact factors

Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. Quality of Reporting of Meta-analyses.
Lancet 1999;354:1896–1900 (1,264 citations, impact factor: 10.2)
Br J Surg 2000;87:1448–1454 (52 citations, impact factor: 2.9)
Onkologie 2000;23:597–602 (23 citations, impact factor: 0.6)

Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials.
Lancet 2001;357:1191–1194 (971 citations, impact factor: 13.3)
JAMA 2001;285:1987–1991 (715 citations, impact factor: 17.6)
Ann Intern Med 2001;134:657–662 (278 citations, impact factor: 11.1)
BMC Med Res Methodol 2001;1:2 (76 citations, impact factor: 2.0[a])
J Am Podiatr Med Assoc 2001;91:437–442 (28 citations, impact factor: 0.4)
Altern Ther Health Med 2002;8:96–100 (1 citation, impact factor: 0.9)
Clin Oral Investig 2003;7:2–7 (46 citations, impact factor: 1.7[b])
Explore (NY) 2005;1:40–45 (17 citations, impact factor: 0.5[b])

Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis PP, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, De Vet HC. Standards for Reporting of Diagnostic Accuracy. Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy.
Clin Chem 2003;49:1–6 (166 citations, impact factor: 5.5)
Ann Intern Med 2003:138:40–44 (160 citations, impact factor: 12.4)
Radiology 2003;226:24–28 (71 citations, impact factor: 4.8)
BMJ 2003;326:41–44 (252 citations, impact factor: 7.2)
Am J Clin Pathol 2003;119:18–22 (6 citations, impact factor: 2.9)
Clin Biochem 2003;36:2–7 (5 citations, impact factor: 1.2)
Clin Chem Lab Med 2003;41:68–73 (9 citations, impact factor: 1.5)
Acad Radiol 2003;10:664–669 (5 citations, impact factor: 1.4)
AJR Am J Roentgenol 2003;181:51–55 (18 citations, impact factor: 2.5)
Ann Clin Biochem 2003;40:357–363 (8 citations, impact factor: 1.2)
Clin Radiol 2003;58:575–580 (22 citations, impact factor: 1.3)
Croat Med J 2003;44:635–638 (10 citations, impact factor: 0.9)
Fam Pract 2004;21:4–10 (17 citations, impact factor: 1.3)
Vet Clin Path 2007;36:8–12 (3 citations, impact factor: 0.3)

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.
PLoS Med 2007;4:e296 (14 citations, impact factor: 12.7)
Ann Intern Med 2007;147:573–577 (43 citations, impact factor: 15.5)
Lancet 2007;370:1453–1457 (24 citations, impact factor: 28.6)
BMJ 2007;335:806–808 (28 citations, impact factor: 9.7)
Prev Med 2007;45:247–251 (9 citations, impact factor: 2.3)
Epidemiology 2007;18:800–804 (12 citations, impact factor: 5.3)
Bull World Health Organ 2007;85:867–872 (7 citations, impact factor: 4.0)
J Clin Epidemiol 2008;61:344–349 (5 citations, impact factor: 2.6)

[a] Unofficial impact factor from the journal's Web site.
[b] Approximate impact factor computed from number of articles published in 2005–2006 and citations obtained in 2007.

the consensus statement was published was associated with an increment of 0.7–1.4 logarithm units in citation counts.

In multivariate analysis, impact factors, consensus statements, and time since publication were considered as predictors of citations. Because statement and time since publication were closely correlated, only one was entered at a time. When adjusted for consensus statement, each logarithm unit of impact factor predicted an increase of 1.2 logarithm units of citations (95% confidence interval: 0.9–1.5, $P < 0.001$). This model explained 71% of the variance in citations (adjusted $r^2 = 0.71$). The interaction between the effects of the impact factor and consensus statement was not statistically significant ($P = 0.58$). When adjusted for time since publication, each logarithm unit of impact factor predicted an increase of 1.0 logarithm unit of citations (95% confidence interval: 0.7–1.3, $P < 0.001$), and the variance explained was 66% (adjusted $r^2 = 0.66$).

4. Discussion

This analysis shows that the prominence of the journal where an article is published, measured by its impact factor, is positively correlated to the number of citations that the article will gather over time. Because identical articles published in different journals were compared, the characteristics of the articles themselves (be it quality of writing, scientific originality, or repute of the authors) could not have explained the observed differences. Hence, these results reflect pure journal-related bias in citation counts.

The strength of the association was high, with most correlation coefficients around 0.8, and reasonably stable across the consensus statements considered. The regression slope was close to unity on a logarithm scale, meaning that if the impact factor of the journal was twice (or 10 times) as high, there would be twice (or 10 times) as many citations for the published article. Whether this equation has predictive ability can be verified in a few years on the citation statistics of the SQUIRE and STREGA consensus statements, which were too recent to be included in this analysis.

The correlation coefficients between the journal impact factor and the number of citations ranged from 0.82 to 0.92 (Table 2). This is higher than previously reported correlations. At the researcher level, the correlation coefficient between journal impact factor and citation rate per year was 0.41 [5]. Another study of laboratory methods described a positive correlation between impact factor and citations in a figure, but did not report a correlation coefficient [10]. In a cohort of more than 1,200 medical articles [11], the Spearman correlation coefficient between the journal impact factor and the citation count was 0.50 (Cynthia Lokker, PhD, personal communication, 25.8.2009). These heterogenous sets of articles vary naturally in their scientific importance and quality, that is, their propensity to be cited, which is why the journal impact factor explains a smaller proportion of the total variance in citations than among the identical articles analyzed in the present study. More importantly, the correlations in these studies [5,10,11] cannot be interpreted in causal terms (a higher impact factor leads to more citations), because the variable

Table 2
Descriptive statistics, correlations between citations and impact factors, and simple linear regression results

	Consensus statement			
	QUOROM	CONSORT (revised)	STARD	STROBE
Descriptive statistics				
Year of first publication	1999	2001	2003	2007
Number of articles	3	8	14	8
Citations to articles: mean (SD), median	446 (708), 52	266 (373), 61	54 (80), 13.5	18 (13), 13.5
Impact factors of journals: mean (SD), median	4.6 (5.0), 2.9	5.9 (6.9), 1.8	3.2 (3.3), 1.6	10.1 (8.9), 7.5
Correlations between citations and journal impact factor				
Pearson correlation coefficient, logarithms	0.92 ($P = 0.25$)	0.82 ($P = 0.013$)	0.84 ($P < 0.001$)	0.82 (0.012)
Spearman correlation coefficient	1.00 ($P = 0.01$)	0.88 ($P = 0.004$)	0.65 ($P = 0.011$)	0.81 (0.015)
Linear regression coefficients (95% confidence intervals)				
Logarithm of citations per logarithm of impact factor	1.4 (−5.9 to 8.7)	1.2 (0.4–2.1)	1.3 (0.8–1.8)	0.7 (0.2–1.1)

Abbreviations: QUOROM, QUality Of Reporting Of Meta-analyses; CONSORT, CONsolidated Standards Of Reporting Trials; STARD, STAndards for Reporting of Diagnostic accuracy; STROBE, STrengthening the Reporting of OBservational studies in Epidemiology; SD, standard deviation.

quality of the studies will act as a confounder, causing both acceptance in a high- or low-impact journal and subsequent citations.

Other prediction models that take into account the scientific merit of the published articles have not performed much better than ours. The correlation between the author's perception of the importance of an article and the number of citations was 0.55 (i.e., 30% of variance explained) in the field of clinical medicine [12]; a multivariate prediction model based on 20 characteristics of articles and journals (not including the impact factor) explained 60% of variance in citations [11], and models that included several subjective assessments of psychology articles explained between 43% and 66% of variance in citations [13]. Although these numbers cannot be compared directly (the studies differ in their scientific fields, heterogeneity of articles, durations of follow-up, and others), our results suggest that the journal's prominence may play a substantial role in an article's future citation success.

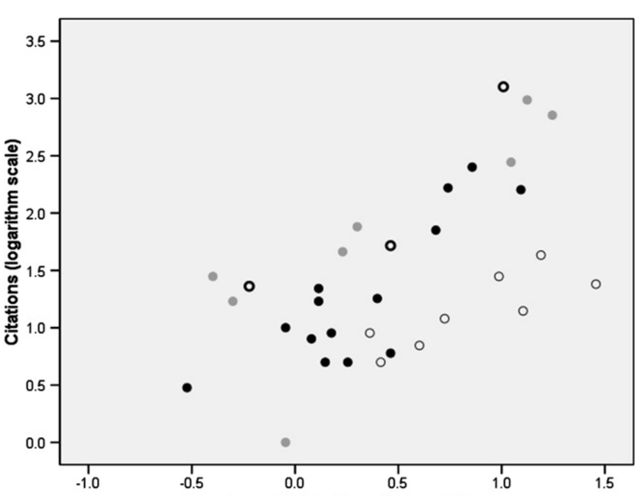

Fig. 1. Scatterplot of citations to publications as a function of journal impact factors for four consensus statements about publication of research (QUOROM: thick circles, CONSORT: grey dots, STARD: black dots, STROBE: thin circles).

By design, this study says nothing about the relationship between scientific merit and citations. The scientific value of the articles that were compared was held constant; hence, this variable could not be analyzed. It is quite possible that scientific arguments trump arguments based on the prestige of the cited source or other social constructs [14]. Our study could not explore other factors that may influence citation counts, such as study design [11,15–17], methodological quality [4,6,11], statistical significance of results [11,18], the cooperation between research groups [19], university prestige [7], online accessibility [20,21], early online access rates [22], or national origin of the citing author and journal [23]. It should be noted that an objective assessment of the quality of a study is not necessarily associated with higher citation counts [6,24,25].

If the impact factor drives citations, researchers who are mindful of their productivity indicators should consider carefully the impact factor of the journal in which they aim to publish their study. Journal prestige was the first criterion that authors considered when submitting their paper even regardless of numerical bibliometric arguments [26]. Furthermore, some institutions use the impact factor of journals in which scientists publish their work as an early indicator of scientific achievement. Although this practice is criticized [4,27], it could be defended if acceptance by a high-impact journal reflected only the high quality of the work. However, the analysis presented here suggests that the association between journal impact factors and future citations is in part self-fulfilling prophecy, as publication in a high-impact journal will push upward the citations of any given article, independent of its value. This creates an additional but perverse incentive to pursue publication in high-impact journals.

These results cast doubt not only on the interpretation of citations as indicators of scientific value but also on the interpretation of impact factors as objective measures of the journals' scientific importance. If indeed an article is cited more often simply for being published in a high-profile journal, then the impact factor becomes auto-correlated—it is high in year Y in good part, because it was high in year Y-1, Y-2, and so on. This phenomenon may help explain the remarkable stability of journal rankings. The indicator has become the thing itself.

The main limitation of this study is that it was performed on consensus statements, not on original research. Whether the same relationship between impact factor and citations would hold for research articles is not established. Such evidence may never be obtained, because multiple publication of the same content is usually considered to be a form of scientific misconduct [28]. However, the same people who cite consensus statements also cite original research and are faced with the recurrent dilemma of selecting 20 or 30 references among the numerous candidates. It is difficult to see why they should be immune to the prestige of the cited journal in doing that.

Selecting citations for inclusion in a manuscript is not a simple deterministic process guided by the scientific merit of available articles; rather, it is a complex phenomenon where individual motivations of the authors and social rules play an important role [2,12,14,29]. A review of citing behavior [2] has identified several classes of factors that influence citation, which relate to time of publication, scientific field, type of article, author characteristics, accessibility, and also the journal where an article was published (as illustrated in this study). Scientific value is one item in this long list. Until we understand better the citation behaviors of medical researchers, the interpretation of citation statistics should be made cautiously, and other performance indicators should be considered as well in assessing institutions, programs, or individuals.

Acknowledgment

Cynthia Lokker, PhD, McMaster University, provided an unpublished analysis cited in the discussion.

References

[1] Hendrix D. An analysis of bibliometric indicators, National Institutes of Health funding, and faculty size at Association of American Medical Colleges medical schools, 1997-2007. J Med Libr Assoc 2008;96:324–34.
[2] Bornmann L, Daniel HD. What do citation counts measure? A review of studies on citing behavior. J Doc 2008;64:45–80.
[3] Moed HF. Citation analysis in research evaluation. Dordrecht, The Netherlands: Springer; 2005.
[4] Callaham ML, Wears RL, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. JAMA 2002;287:2847–50.
[5] Seglen PO. Causal relationship between article citedness and journal impact. J Am Soc Inform Sci 1994;45:1–11.
[6] Nieminen P, Carpenter J, Rucker G, Schumacher M. The relationship between quality of research and citation frequency. BMC Med Res Methodol 2006;6:42.
[7] Filion KB, Pless IB. Factors related to the frequency of citation of epidemiologic publications. Epidemiol Perspect Innov 2008;5:3.
[8] Etter JF, Stapleton J. Citations to trials of nicotine replacement therapy were biased toward positive results and high impact-factor journals. J Clin Epidemiol 2009;62:831–7.
[9] Saha S, Saint S, Christakis DA. Impact factor: a valid measure of journal quality? J Med Libr Assoc 2003;91:42–6.
[10] Huang Y, Yang Q, Ao X. Bibliometric analysis of pentachlorophenol remediation methods during the period of 1994 to 2005. Scientometrics 2008;77:177–86.
[11] Lokker C, McKibbon KA, McKinlay RJ, Wilczynski NL, Haynes RB. Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: retrospective cohort study. BMJ 2008;336:655–7.
[12] Aksnes DW. Citation rates and perceptions of scientific contribution. J Am Soc Inform Sci Technol 2006;57:169–85.
[13] Shadish WR, Tolliver D, Gray M, Gupta SKS. Author judgements about works they cite: three studies from psychology journals. Soc Stud Sci 1995;25:477–98.
[14] Baldi S. Normative versus social constructivist processes in the allocation of citations: a network analytic model. Am Soc Rev 1998;63: 829–46.
[15] Lee KP, Schotland M, Bacchetti P, Bero LA. Association of journal quality indicators with methodological quality of clinical research articles. JAMA 2002;287:2805–8.
[16] Patsopoulos NA, Analatos AA, Ioannidis JPA. Relative citation impact of various study designs in the health sciences. JAMA 2005;293:2362–6.
[17] Bhandari M, Busse J, Devereaux PJ, Montori VM, Swiontkowski M, Tornetta P, et al. Factors associated with citation rates in the orthopaedic literature. Can J Surg 2007;50:119–23.
[18] Nieminen P, Rucker G, Miettunen J, Carpenter J, Schumacher M. Statistically significant papers in psychiatry were cited more often than others. J Clin Epidemiol 2007;60:939–46.
[19] Figg WD, Dunn L, Liewehr DJ, Steinberg SM, Thurman PW, Barrett JC, et al. Scientific collaboration results in higher citation rates of published articles. Pharmacotherapy 2006;26:759–67.
[20] Eysenbach G. Citation advantage of open access articles. PLoS Biol 2006;4:e157.
[21] Davis PM, Lewenstein BV, Simon DH, Booth JG, Connolly MJL. Open access publishing, article downloads, and citations: randomised controlled trial. BMJ 2008;337:a568.
[22] Perneger TV. Relation between online "hit counts" and subsequent citations: a prospective study of research papers in the BMJ. BMJ 2004;329:546–7.
[23] Grange RI. National bias in citation in urology journals: parochialism or availability? BJU Int 1999;84:601–3.
[24] Bain CR, Myles PS. Relationship between journal impact factor and levels of evidence in anaesthesia. Anaesth Intensive Care 2005;33: 567–70.
[25] Barbui C, Cipriani A, Malvini L, Tansella M. Validity of the impact factor of journals as a measure of randomized controlled trial quality. J Clin Psychiatry 2006;67:37–40.
[26] Frank E. Authors' criteria for selecting journals. JAMA 1994;272: 163–4.
[27] Seglen PO. Why the impact factor of journals should not be used for evaluating research. BMJ 1997;314:498–502.
[28] Kassirer JP, Angell M. Redundant publication—a reminder. N Engl J Med 1995;333:449–50.
[29] Cronin B. Metatheorizing citation. Scientometrics 1998;43:45–55.