
Impact factor: a valid measure of journal quality?

By Somnath Saha, M.D., M.P.H.
sahas@ohsu.edu
Assistant Professor

Division of General Internal Medicine & Geriatrics and
Division of Medical Informatics & Outcomes Research
Portland VA Medical Center and Oregon Health & Science University
Portland VA Medical Center (P3MED)
3710 Southwest U.S. Veterans Hospital Road
Portland, Oregon 97207

Sanjay Saint, M.D., M.P.H.
saint@umich.edu
Associate Professor

Division of General Medicine
University of Michigan
3116 Taubman Center
Ann Arbor, Michigan 48109-0376

Dimitri A. Christakis, M.D., M.P.H.
dachris@u.washington.edu
Assistant Professor

Department of Pediatrics and the Child Health Institute
University of Washington
146 North Canal Street, Suite 300
Seattle, Washington 98103

Objectives: Impact factor, an index based on the frequency with which a journal's articles are cited in scientific publications, is a putative marker of journal quality. However, empiric studies on impact factor's validity as an indicator of quality are lacking. The authors assessed the validity of impact factor as a measure of quality for general medical journals by testing its association with journal quality as rated by clinical practitioners and researchers.

Methods: We surveyed physicians specializing in internal medicine in the United States, randomly sampled from the American Medical Association's Physician Masterfile (practitioner group, $n = 113$) and from a list of graduates from a national postdoctoral training program in clinical and health services research (research group, $n = 151$). Respondents rated the quality of nine general medical journals, and we assessed the correlation between these ratings and the journals' impact factors.

Results: The correlation between impact factor and physicians' ratings of journal quality was strong ($r^2 = 0.82$, $P = 0.001$). The correlation was higher for the research group ($r^2 = 0.83$, $P = 0.001$) than for the practitioner group ($r^2 = 0.62$, $P = 0.01$).

Conclusions: Impact factor may be a reasonable indicator of quality for general medical journals.

INTRODUCTION

The impact factor of a journal reflects the frequency with which the journal's articles are cited in the scientific literature. It is derived by dividing the number

of citations in year 3 to any items published in the journal in years 1 and 2 by the number of substantive articles published in that journal in years 1 and 2 [1]. For instance, the year 2002 impact factor for *Journal X* is calculated by dividing the total number of citations

Figure 1
Calculating impact factor

A = Total citations in 2002 to items published in *Journal X*
B = 2002 citations to items published in *Journal X* in 2000–01 (subset of **A**)
C = Number of substantive articles published in *Journal X* in 2000–01

$$\text{Impact factor} = B/C$$

Example:

Assume that in 2002, there were 3,200 citations to items published in *Journal X*. Of these, 550 were citations to items published in *Journal X* in 2000 and 2001. During those two years, *Journal X* published 72 articles.

$$2002 \text{ Journal X impact factor} = 550/72 = 7.64$$

Adapted from: GARFIELD E. The impact factor. [Internet]. *Curr Contents* 1994 Jun 20;(25):3–7. [cited 16 Aug 2002]. <http://sunweb.isinet.com/isi/hot_essays/journalcitationreports/7.html>.

during the year 2002 to items appearing in *Journal X* during 2000 and 2001 by the number of articles published in *Journal X* in 2000 and 2001 (Figure 1). Conceptually developed in the 1960s, impact factor has gained acceptance as a quantitative measure of journal quality [2]. Impact factor is used by librarians in selecting journals for library collections, and, in some countries, it is used to evaluate individual scientists and institutions for the purposes of academic promotion and funding allocation [3, 4]. Not surprisingly, many have criticized the methods used to calculate impact factor [5, 6]. However, empiric evaluations of whether or not impact factor accurately measures journal quality have been scarce [7].

The use of impact factor as an index of journal quality relies on the theory that citation frequency accurately measures a journal's importance to its end users. This theory is plausible for journals whose audiences are primarily researchers, most of whom write manuscripts for publication. By citing articles from a given journal in their own manuscripts, researchers are in essence casting votes for that journal. Impact factor serves as a tally of those votes.

A journal's impact within clinical medicine, however, depends largely on its importance to practitioners, most of whom never write manuscripts for publication and thus never have a chance to "vote." Citation frequency may therefore better reflect the importance of clinical journals to researchers than practitioners. Because the opinions of both practitioners and researchers are relevant in judging the importance of clinical journals, the validity of impact factor as a measure of journal quality in clinical medicine is uncertain. The authors therefore sought to examine whether impact factor is a valid measure of journal quality as rated by clinical practitioners and researchers.

METHODS

As part of a study assessing the effect of journal prestige on physicians' assessments of study quality [8], we mailed questionnaires to 416 physicians specializing in internal medicine in the United States. We recruited half of our sample (practitioner group) from the American Medical Association's (AMA's) master

list of licensed physicians. This database is not limited to AMA members and is acknowledged to be the most complete repository of physicians' names, addresses, specialties, and primary activities (e.g., office practice, hospital practice, research, etc.) in the United States. The 208 practitioners we surveyed for our study were randomly selected from the approximately 110,000 U.S. physicians who listed internal medicine as their primary specialty, by the vendor who dispenses data from the AMA's master list.

We used a random number generator (STATA, version 5.0) to select the other half of our sample (research group) from the alumni directory of the Robert Wood Johnson Clinical Scholars Program, a national post-doctoral training program for physicians. Graduates of this program have typically received training in clinical research, epidemiological research, health services research, or a combination of these. Our intent in including this group was to enrich the sample with physicians likely to be engaged primarily in research to determine whether the correlation between impact factor and journal quality was higher for researchers than for practitioners, as we hypothesized it would be.

We asked respondents to rate the overall quality of nine general medical journals on a scale from 1 to 10, with 10 being the highest rating. We chose journals that we believed spanned a broad range of perceived quality and would be familiar to internists in the United States: *American Journal of Medicine (AJM)*, *Annals of Internal Medicine (Annals)*, *Archives of Internal Medicine (Archives)*, *Hospital Practice (HP)*, *JAMA*, *Journal of General Internal Medicine (JGIM)*, *The Lancet*, *New England Journal of Medicine (NEJM)*, and *Southern Medical Journal (SMJ)*. All of these journals are categorized as journals of "clinical medicine" by ISI.* Some also publish basic science research and are therefore categorized as "life sciences" journals as well. We included these "hybrid" journals (*AJM*, *Annals*, *Archives*, *JAMA*, *Lancet*, *NEJM*) because we believed they represented journals commonly used and recognized by clinicians in the United States. We compared mean quality ratings between journals using *t* tests.

We obtained each journal's impact factor from 1997 [9], because our survey sampled physicians' opinions in the early part of 1998. We then calculated the correlation (squared value of Pearson's correlation coefficient) between the natural logarithm of impact factor and physicians' mean quality ratings for the nine journals, using first the entire sample and then the practitioner and research groups separately. We transformed impact factor logarithmically because the relationship between impact factor and quality ratings was nonlinear. Such transformations are generally appropriate when using linear modeling to test nonlinear associations [10].

We also asked respondents to report whether they subscribed to and regularly read each of the nine journals. We examined the correlation of reported journal

* ISI's list of clinical medicine journals may be viewed at <http://www.isinet.com/isi/journals/>.

Table 1
Characteristics of survey respondents

Characteristic	Total (n = 264)	Practitioner group (n = 113)	Research group (n = 151)
Average age (standard deviation)	47.1 (9.7)	49.0 (12.0)	45.7 (7.2)
Average medical school graduation year (standard deviation)	1978 (9.6)	1976 (11.5)	1979 (7.6)
Male	84%	85%	83%
Certified in a medical subspecialty	23%	18%	26%

quality with subscription rates and readership rates to evaluate how impact factor compared as a marker of journal quality with these rates, which represented alternate ways in which physicians "vote" for high-quality journals.

To account for the possibility that some respondents were not familiar with *JGIM* (a subspecialty journal), *The Lancet* (a foreign journal), and *SMJ* (a regional journal), we repeated our analyses after excluding ratings for each of these journals.

RESULTS

Thirteen questionnaires were returned with no forwarding address, and four were returned by physicians reporting that they no longer practiced medicine or were retired. Of the 399 eligible participants, 264 returned questionnaires (response rate 66%), 113 from the practitioner group (response rate 58%) and 151 from the research group (response rate 74%). There was no significant difference between respondents and nonrespondents with respect to age, graduation year, or subspecialty training. Men responded more frequently than women (69% versus 55%; $P = 0.03$). In the practitioner group, 105 (93%) of the respondents had a listed primary activity of office or hospital practice. Characteristics of respondents are shown in Table 1.

Mean quality ratings for the nine journals ranged from 4.8 to 8.4; impact factors ranged from 0.17 to 27.8 (Table 2). Pairwise comparisons of quality ratings for individual journals were all significant ($P < 0.01$), except between *AJM* and *Archives* ($P = 0.10$) and between *HP* and *SMJ* ($P = 0.12$). There was a strong correlation between impact factor (logarithmically

transformed) and physicians' ratings of journal quality ($r^2 = 0.82$, $P = 0.001$) (Figure 2). This correlation indicated that impact factor captured 82% of the variation in quality ratings across respondents. The correlation was somewhat higher for the research group ($r^2 = 0.83$, $P = 0.001$) than for the practitioner group ($r^2 = 0.62$, $P = 0.01$). Excluding *JGIM*, *The Lancet*, and *SMJ* did not substantively affect these results.

Subscription rates ranged from 5% to 77% and readership rates from 5% to 75% (Table 2). Physicians' ratings of journal quality correlated more closely with impact factor than with subscription rates ($r^2 = 0.50$, $P = 0.03$) or readership rates ($r^2 = 0.63$, $P = 0.01$). Excluding *The Lancet*, however, substantially increased the correlation of journal quality ratings with both subscription rates ($r^2 = 0.87$, $P = 0.001$) and readership rates ($r^2 = 0.91$, $P < 0.001$). Excluding *JGIM* and *SMJ* did not change these results.

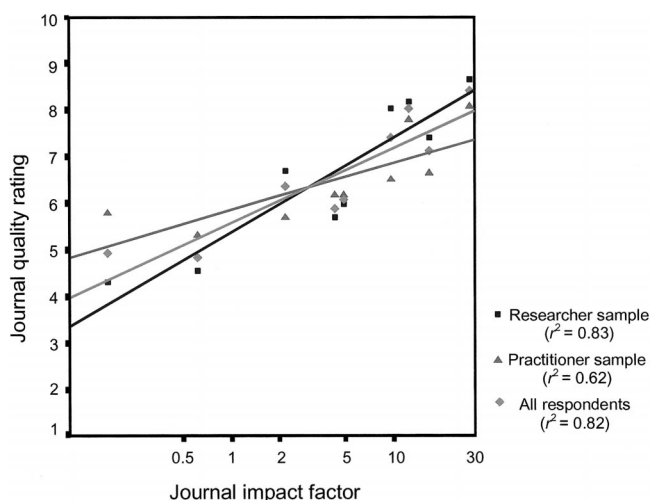
DISCUSSION

Our findings suggest that impact factor may be a reasonable indicator of quality for general medical journals. Subscription rates and readership rates are other potential proxies for journal quality, but these markers are limited by the fact that, for many journals, subscription is not related to physicians' desire to read the journals but rather to society membership, as well as to journal cost and availability. Impact factor performs comparably with subscription and readership rates for U.S. journals. However, when a foreign journal (*The Lancet*) is included, subscription and readership rates lose much of their correlation with journal quality, whereas impact factor retains its correlation. This sug-

Table 2
Journal quality ratings, impact factors, subscribership, and readership (by impact factor, in descending order)

Journal	Quality Ratings (mean \pm standard deviation)			Impact factor (1997)	Subscribe to journal (n, %) (n = 264)	Read journal (n, %) (n = 264)
	Practitioner group (n = 113)	Research group (n = 151)	All respondents (n = 264)			
<i>New England Journal of Medicine</i>	8.1 \pm 1.9	8.7 \pm 0.9	8.4 \pm 1.4	27.8	196 (74%)	199 (75%)
<i>Lancet</i>	6.7 \pm 2.0	7.4 \pm 1.6	7.1 \pm 1.8	16.1	19 (7%)	36 (14%)
<i>Annals of Internal Medicine</i>	7.8 \pm 1.6	8.2 \pm 1.2	8.0 \pm 1.4	12.1	203 (77%)	190 (72%)
<i>JAMA</i>	6.5 \pm 1.8	8.1 \pm 1.2	7.4 \pm 1.7	9.3	187 (71%)	184 (70%)
<i>Archives of Internal Medicine</i>	6.2 \pm 1.7	6.0 \pm 1.5	6.1 \pm 1.6	4.8	106 (40%)	95 (36%)
<i>American Journal of Medicine</i>	6.2 \pm 1.8	5.7 \pm 1.6	5.9 \pm 1.7	4.2	62 (23%)	52 (20%)
<i>JGIM</i>	5.7 \pm 1.8	6.7 \pm 1.3	6.4 \pm 1.6	2.1	90 (34%)	83 (32%)
<i>Southern Medical Journal</i>	5.3 \pm 1.9	4.6 \pm 1.6	4.8 \pm 1.8	0.61	13 (5%)	12 (5%)
<i>Hospital Practice</i>	5.8 \pm 2.0	4.3 \pm 1.6	4.9 \pm 1.9	0.17	81 (31%)	60 (23%)

Figure 2
Correlation of impact factor with journal quality ratings



gests that impact factor may be less prone to biases than other available indices and may thus be a more resilient measure of journal quality.

The higher correlation between impact factor and journal quality ratings in the research group as compared to the practitioner group is not surprising. Because impact factor is determined by citation frequency, it makes sense that it would correlate strongly with quality as judged by researchers, who themselves generate citations. The correlation among practitioners, however, is also high, with impact factor capturing more than 60% of the variation in quality ratings.

Impact factor is commonly used as a tool for managing scientific library collections. Librarians faced with finite budgets must make rational choices when selecting journals for their departments and institutions. Impact factor helps guide those choices by determining which journals are most frequently cited. Journals that are cited frequently generally contain articles describing the most notable scientific advances (i.e., those with the greatest "impact") in a given field and are therefore of greatest interest to researchers, teachers, and students in most scientific disciplines. In medical libraries, however, the interests of clinicians must also be considered. Journals publishing "cutting-edge" medical discoveries may be cited frequently and highly valued by researchers but may be of less value to clinicians than journals providing, for instance, concise overviews of common clinical problems. Impact factor may therefore be less valid as a guide to selecting high-quality journals in clinical medicine than in other scientific disciplines.

Few studies have addressed impact factor's validity as a quality measure for clinical journals. Foster reported poor correlation between impact factor and journal prestige as ranked by scientists from the National Institutes of Health (NIH) [11]. While the opinions of NIH scientists may be an appropriate "gold standard" for judging basic science journals, we be-

lieve that clinical researchers and practitioners are the best judges for clinical journals. Our study suggests that impact factor may be an accurate gauge of relative quality as judged by both researchers and practitioners.

Tsay examined the correlation between impact factor and frequency of journal use, as measured by re-shelving rates, in a medical library in Taiwan [12]. Journals were divided into four subject categories—"clinical medicine," "life science," "hybrid" (clinical medicine and life science), or "other" (neither clinical medicine nor life science)—according to their classification in ISI's *Current Contents*[®] journal lists. Most of the general medical journals in our study are "hybrid" journals. In that category, Tsay found a strong correlation between impact factor and frequency of journal use (Pearson's $r^2 = 0.53$, $P = 0.0001$) and concluded that impact factor is a "significant measure of importance that could be used for journal selection" [13]. However, frequency of journal use is influenced not only by the importance of a journal to its end users but also by other factors, such as the frequency of publication [14]. To accurately gauge the importance of a journal, the frequency of its use must be considered alongside its value, or quality, as judged by its readers. Our study therefore complements Tsay's findings by demonstrating that impact factor correlates not only with frequency of journal use but also with journal quality.

Our findings should be considered preliminary in light of several limitations. First, we collected only global ratings of journal quality from our respondents. We were therefore not able to assess correlations between impact factor and specific aspects of journal quality (e.g., immediate usefulness, importance to clinical practice, prestige, methodological rigor, etc.). Further study should evaluate the aspects of quality captured by impact factor.

Second, we asked respondents to assign quality ratings to each journal in our survey, even if they did not regularly read all of the journals. We selected journals with which we believed most U.S. internists would have some familiarity, and we conducted secondary analyses excluding those with which we felt a substantial proportion of our respondents might not have had direct experience. Nevertheless, some of the reported journal quality ratings were likely based on perception rather than experience with the journal, raising concern about the credibility of the ratings. Impact factor might reflect a journal's reputation rather than its actual quality. Future study should examine correlations between impact factor and quality ratings based on critical review of randomly selected samples from each journal to help make this distinction.

Third, we sampled only U.S. internists and queried them about nine general medical journals predominantly published and distributed in the United States. Internists make up the largest pool of primary care physicians and subspecialists in the United States [15]. Furthermore, we drew our sample randomly from national lists, to make it broadly representative. Our re-

sults, nonetheless, might not be generalizable to physicians or journals from other countries or subspecialties.

Finally, the research group in our sample consisted of physicians who had completed research training, not necessarily researchers, per se. Moreover, this group was drawn from a list of graduates from a single national training program, whose opinions might not reflect those of all clinical researchers. These factors might have limited our ability to accurately determine the correlation of impact factor and journal quality for researchers. Our practitioner sample, however, was drawn randomly from a database of all U.S. physicians and contained only eight internists whose primary activity was research or administration. Analyses including and excluding these eight respondents produced nearly identical results (data not shown). Thus, we can confidently assert that impact factor reasonably reflects judgments of overall journal quality among practitioners in internal medicine. We suspect that, if anything, it is an even better indicator among researchers.

It should be emphasized that we logarithmically transformed impact factor in our analyses for statistical purposes and that we did not find a direct, linear correlation between impact factor and journal quality ratings. For instance, there was a forty-six-fold difference in the impact factors for *NEJM* and *SMJ*, but their quality ratings in our survey varied by a factor of only 1.75. Thus, while the quality of these journals may differ, it is unlikely that they differ to the degree that a direct comparison of their impact factors might suggest.

Journal impact factor has its limitations, and we believe that further evaluation of whether and how impact factor measures journal quality is warranted before it is widely adopted as a quantitative marker of journal quality. Nevertheless, our findings provide preliminary evidence that despite its shortcomings, impact factor may be a valid indicator of quality for general medical journals, as judged by both practitioners and researchers in internal medicine. As such, its use as an aid in the selection of journals for medical libraries is probably justified.

ACKNOWLEDGMENTS

■ Funding for this study was provided by the Robert Wood Johnson Clinical Scholars Program.

■ Drs. Saha and Saint are supported by career development awards from the Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs.

■ Dr. Christakis is a Robert Wood Johnson generalist physician faculty scholar.

■ The opinions expressed in this paper are those of the authors and not necessarily those of the Department of Veterans Affairs or the Robert Wood Johnson Foundation.

REFERENCES

1. GARFIELD E. Journal impact factor: a brief review. *Can Med Assoc J* 1999 Oct 19;161(8):979–80.
2. GARFIELD E. The impact factor. [Internet]. *Curr Contents* 1994 Jun 20;(25):3–7. [cited 16 Aug 2002]. <<http://sunweb.isinet.com/isi/hot/essays/journalcitationreports/7.html>>.
3. SEGLEN PO. Why the impact factor of journals should not be used for evaluating research. *BMJ* 1997 Feb 15;314(7079):498–502.
4. LOWY C. Impact factor limits funding. *Lancet* 1997 Oct 4;350(9083):1035.
5. SEGLEN, op. cit.
6. HANSSON S. Impact factor as a misleading tool in evaluation of medical journals. *Lancet* 1995 Sep 30;346(8979):906.
7. FOSTER WR. Impact factor as the best operational measure of medical journals. *Lancet* 1995 Nov 11;346(8985):1301.
8. CHRISTAKIS DA, SAINT S, SAHA S, ELMORE JG, WELSH DE, BAKER P, KOEPEL TD. Do physicians judge a study by its cover? an investigation of journal attribution bias. *J Clin Epidemiol* 2000 Aug;53(8):773–8.
9. INSTITUTE FOR SCIENTIFIC INFORMATION. 1997 SCI Journal Citation Reports. Philadelphia, PA: Institute for Scientific Information, 1998.
10. KLEINBAUM DG, KUPPER LL, MULLER KE. Applied regression and other multivariable methods. Belmont, CA: Duxbury Press, 1988.
11. FOSTER, op. cit.
12. TSAY MY. The relationship between journal use in a medical library and citation use. *Bull Med Libr Assoc* 1998 Jan;86(1):31–9.
13. *IBID.*, 39.
14. *IBID.*, 33–5.
15. AMERICAN MEDICAL ASSOCIATION. Physician characteristics and distribution in the U.S. Chicago, IL: The Association, 1998.

Received May 2002; accepted August 2002