Review article

Citation rates and journal impact factors are not suitable for evaluation of research

Per O Seglen

Institute for Studies in Research and Higher Education (NIFU), Hegdehaugsveien 31, NO-0352 Oslo, Norway, and Institute for Cancer Research, Norwegian Radium Hospital, Montebello, NO-0310 Oslo, Norway

Recognition of scientific quality is essential not only as a motivating force for the individual scientist, but also as a key to the funding needed to keep the scientific machine running. Unfair research evaluation is, therefore, a major source of frustration in scientific communities worldwide, and a potential threat to the whole scientific enterprise. The traditional peer review is too often performed on the basis of superficial criteria (personal or institutional reputation, project relevance, journal prestige, crude publication counts, etc.) and is generally regarded as a lottery, not entirely without justification.[1-3] Alternative evaluation methods have, therefore, received increasing attention, mainly those based on seemingly objective, quantitative indicators, like citation rates and journal impact factors.

Measurement of citation rates

Citation data are assembled by the Institute for Scientific Information (ISI) in Philadelphia, on the basis of the reference lists of articles in scientific journals. All references (citations) to a given article in a given year are listed in the annual *Science Citation Index* (SCI), where both cited and citing articles are arranged alphabetically by first-author names. From this list, the annual citation rate of any article or first-author can be obtained, most easily by computerized access to the ISI database or to a CD-ROM version of it.

However, there are many technical problems associated with the recording and retrieval of citation data (Table 1).

The SCI covers about 3,200 journals,[4] which is a small fraction of the world total of 126,000.[5] Different research fields are covered unequally: the coverage for chemistry has been estimated at 90%, as compared to 30% for biology.[6] Although the database attempts to include the most important journals, the journal selection appears to be somewhat random.[7] The journal set in the database may vary in composition from year to year.[8,9] Books are not included as source items, despite their prominent role in many research fields.[10]

The ISI database has a preference for English language journals:[6] for example, it was found to include only two German social science journals, as compared to 542 in a German database.[11] This language bias will favor English-speaking authors, given the strong tendency of authors to cite selectively articles in their own national language.[5,12,13] More than one-half of all citations concern US scientists,[14] who are particularly prone to cite one another,[5,13,15] thereby helping to raise the citation rate of US science 30% above the world average.

The retrieval of citation data raises several problems. Only first-authors can be searched for in CD-ROM or paper versions of the SCI, necessitating the use of reference lists and article-by-article search for compilation of individual citation profiles (although

Table 1. Technical problems associated with citation data

1. Incomplete journal coverage in the database
2. The database coverage differs between research fields
3. The journal set included in the database may vary
4. Books are not included as source items in the database
5. The database is biased towards the English language
6. The database is biased towards US science
7. First-author retrieval only (without direct access)
8. On-line access to the database is expensive
9. Delayed registration of citations
10. Many misprints (up to 25%)
11. Inconsistent foreign language spelling (e.g., æ, ø, å)
12. Synonymy (several variants of the same article)
13. Homonymy (several authors with the same name, e.g., in Japan)

Table 2. Motives and problems in reference selection

1. The primary criterion is not quality, but utility in research
2. Incomplete referencing due to journal space limitations
3. Citation of secondary sources (e.g., reviews) rather than of primary publications
4. Reference copying
5. Established knowledge is not cited ("obliteration by incorporation")
6. Argumentative citation (mainly self-supportive)
7. Flattery (citation of editors, potential referees, etc.)
8. Show-off (citation of "hot" papers)
9. Conventions (methods are cited; reagents are not)
10. Self-citation
11. In-house citation (friends and close colleagues)

on-line searches in the central ISI database offer wider possibilities, they are very expensive). Entry into the database is often delayed, so that articles and citations published at the end of the year may be missing from the annual index. Misprints are frequent (about 25%), often reflecting errors in the original reference lists.[16,17] In one case, misprints caused the same reference to be listed in the SCI under 70 different synonyms.[18] Homonymy (same name) is another problem, especially among Japanese authors: K. Suzuki figures impressively in the database, with several hundred publications each year.

Citer motives

The use of citations as a quality measure carries the implicit assumption that authors select their references on the basis of quality. However, a little reflection will tell us that we primarily refer to publications that we make use of in our own work. *Utility in research,* rather than pure scientific quality, is therefore the primary criterion for reference selection. Unfortunately, journal space limitations prevent us from citing all the sources we draw on: it has been estimated that only some 30% of the literature base of a scientific paper is rewarded with citations.[19] This strong selection leaves room for a mixed bag of secondary citer motives (Table 2).

An easy way out of the reference choice dilemma is the use of reviews and other secondary sources, which are, therefore, generally much cited. Even reference copying is widespread, as sometimes revealed by particular misprint variants which occur more frequently than the original reference. Certain citing conventions are generally followed: for example, the originator of an analytical method is usually cited, whereas the discoverer of a useful chemical (e.g., an inhibitor) is not. The ISI database does not correct for self-citations, which amount to usually about 30% of the citations, and often more.[6,18]

Table 3. Research field effects that influence citation rates

1. Mean number of references per article in field
2. Reference obsolescence relative to time-window for citation recording
3. Field size (affects mainly citation span, i.e., maximal attainable citation rate)
4. Field dynamics (field expansion or contraction)
5. Interfield relations (e.g., basal vs. applied)
6. Subfield microheterogeneity

Research field effects

A major problem with the use of citations in evaluation is that they are not comparable between different research fields (Table 3). For example, scientific disciplines which use many references per article, such as biochemistry, will have higher citation rates than mathematics, which uses few references. When this is combined with a short-term recording window that favors current, but short-lived articles, the average biochemist will be cited four times as often as the average mathematician.[8] Within the arts and humanities, article references are hardly used at all, leaving these research fields (and others) virtually uncited.[20] In relation to the choice of time-window, it should be noted that the recording of cumulative citations over a long period will be as much a measure of productivity as of "citedness".

The effects of field size are complex. In very small, closed fields, the citation rates can be expected to be proportional to the number of authors in the field, but only up to a limit set by the length of the reference list. Beyond this limit, the field citation rate should be independent of field size.[21] However, the range will be wider in a large field, giving a better opportunity for a few to become often cited.[22] If a field expands rapidly, citation rates will increase, because the number of citers is high relative to the amount of citable material.[23,24]

The main field effect is probably the ability of a research field to be cited by adjacent fields. For example, clinical medicine draws heavily on basic science, but not *vice versa*. The result is that papers in basic medicine are cited 3–5 times more often than those in clinical medicine.[22,25,26] There is evidence that field effects extend even to the subdiscipline level,[27] meaning that the citation rates of scientists working on different subjects cannot be compared, even in the same field. The choice of research theme will determine, *a priori,* the probability of becoming highly cited. Although attempts have been made to correct for field effects—e.g., by dividing article citation rates by journal citation rates (journal impact factors)[6]—or by more sophisticated field citation factors,[28,29] such

Table 4. Journal impact factors, 1991

Cell	30.2
Science	19.6
EMBO J	12.4
J Biol Chem	6.7
Biochem J	3.7
Biochim Biophys Acta	2.5
Acta Anaesth Scand	1.0
Acta Chir Scand	0.4
Eur J Surg Oncol	0.0

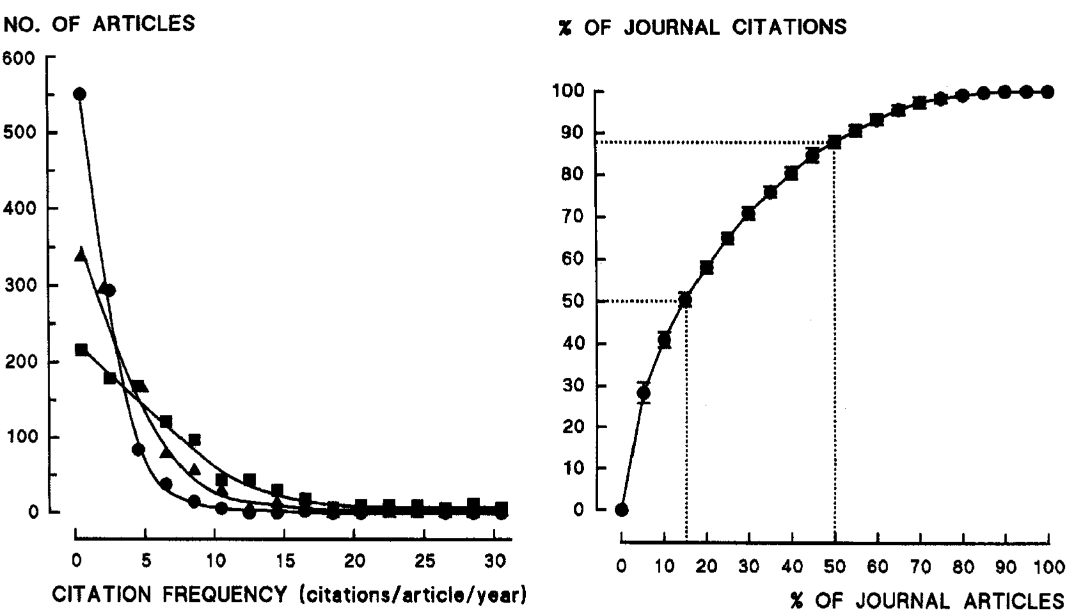
corrections are as likely to introduce new errors (e.g., by punishing authors publishing in much cited journals) as to eliminate old ones. It will never be possible to construct individualized field factors and it is questionable if they would serve any purpose: citation rates are determined by so many technical factors that it is doubtful whether pure scientific quality has any detectable effect at all, in which case all adequately field-corrected citation rates would be reduced to unity.

Use of journal impact factors in research evaluation: are they representative?

Citations are clearly not very useful for the evaluation of research quality. Nevertheless, an indirect citation measure, the so-called *journal impact factor,* supposed to represent the annual mean citation rate of all the articles in a journal,[30] has found increasing application in research evaluation.[3,4,31] Lists of journal impact factors are published annually in the *SCI Journal Citation Reports;* these factors are, therefore, easily available and simple to use (Table 4 gives some examples to indicate the range in factor values).

The use of journal impact factors in evaluation is based on the premise that the journal is representative of its articles, meaning that one can simply add up the journal impact factors of an author's articles to obtain an apparently objective and quantitative measure of the author's scientific achievement. If this premise were valid, the article citation rates would have been distributed in a narrow, Gaussian fashion around the population mean (i.e., the journal impact factor). However, as seen in Figure 1A, the actual distributions recorded for three different journals were very skewed, with only a few articles anywhere near the population mean. A cumulative plot (Figure 1B) shows that 15% of the journal articles account for 50% of the citations and that the most cited 50% of the articles is cited, on average, ten times as often as the least cited half. By giving all articles the same score (the journal impact factor value), this tremendous difference is masked, which contradicts the pur-

Figure 1. Skewed distribution of citations of articles in a journal.



Third-year frequence of citations of articles in Biochim. Biophys. Acta (•), Biochem. J. (▲) and J. Biol. Chem. (■), arranged by number of articles in each citedness category.

Cumulative contribution of the different frequence of citations categories (beginning with the most-cited 5%) to the total journal impact. Each value is the mean ± S.E. of the three journals in (A). Dotted lines indicate the contributions of the 15% and 50% most cited articles. From ref. 44.

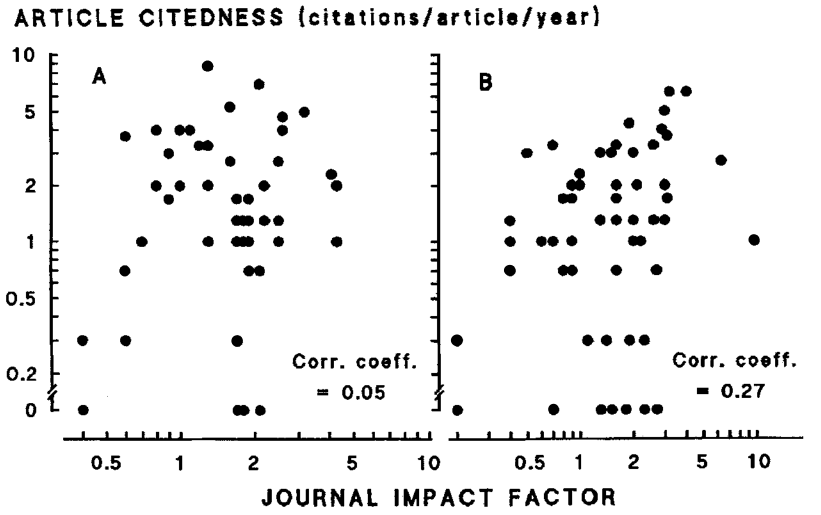


Figure 2. Correlation between article frequence of citations and journal impact for two individual authors (total production). From ref. 41.

pose of evaluation. Since the impact factors are not representative, it may come as no surprise that the correlation between these factors and actual article "citedness" is often very poor (Figure 2).

Journal impact factors are inappropriately calculated, causing bias

Most of the problems associated with the use of citations in evaluation also extend to journal impact factors, but there are a few more (Table 5). For example, although the impact factors are computed by including citations of all types of journal documents in the numerator, items like editorials, letters and meeting abstracts are excluded from the denominator.[32,33] As a result, journals with interesting editorials and a lively correspondence section can have their impact factors inflated by up to 75%.[34] Since the ISI database does not correct for self-citations, editors can raise the impact of their journals further by referring frequently to their previous editorials. Other impact-maximizing measures would be the inclusion of review articles, which are generally highly cited,[33,35] and of long, rather than short articles, since citedness is roughly proportional to article length.[27,36] On a per word basis, biochemical journals specializing in short communications are actually cited at least as often as the leading high-impact journals in their field.[37,38]

Since the journal impact factor is a short-term index (based on citations during the first two years following publication), journals with short publication lags will have a high proportion of their self-citations recorded, and a correspondingly high journal impact factor.[39] Russian journals, which are mainly cited by other Russian journals,[12] have particularly long publication lags, resulting in generally low impact factors.[40] US journals, on the other hand, are favored by national citation bias, since they quantitatively dominate the ISI database.

Many research fields are underrepresented in the database[6] and all journals in such fields will receive too low impact factors. In some fields, books can be an important vehicle of publication, but citations in books are not recorded and therefore do not contribute to journal impact factor values.[10] The source jour-

Table 5. Problems associated with the use of journal impact factors

1. Journal impact factors are not representative of the individual journal articles
2. Impact factors correlate poorly with actual article citation rates
3. Journal impact factors are research field-dependent
4. Small research fields often lack high-impact journals
5. Citations of "non-citable" documents are included in the impact factor calculation
6. Journals inflate their impact factor through self-citation bias
7. Journals not included as source documents in database are deprived of self-citations
8. Review articles are much cited and give high impact factors to their journals
9. Long articles are more cited than short ones and give higher journal impact factors
10. Short publication times give high impact factors
11. Preference for national language references favors English-language journals
12. National bias in reference selection favors American journals
13. Authors' choice of journals is (was) not primarily based on impact factors
14. Article "citedness" is not affected by the journal impact factor

Table 6. Research field dependence on journal impact factors

Research field	No. of ISI journals	Journal impact factor, mean (SEM)	
		Median 5 journals	Top 5 journals
Immunology	69	1.9 (0.11)	13 (3.5)
Biochemistry	134	1.6 (0.02)	17 (4.2)
Genetics	49	1.5 (0.08)	8.3 (1.5)
Biophysics	33	1.2 (0.18)	6.0 (0.79)
Anesthesiology	10	0.90 (0.18)	2.6 (0.84)
Dermatology	22	0.66 (0.03)	2.3 (0.37)
Ophthalmology	25	0.55 (0.05)	2.1 (0.33)
Mathematics	60	0.46 (0.01)	1.5 (0.16)
Engineering	26	0.41 (0.01)	1.0 (0.14)
Aerospace eng.	8	0.24 (0.03)	0.36 (0.05)

Data calculated from SCI Journal Citation Reports, 1986.

nals in the ISI database are not necessarily selected on the basis of frequency of citations: among leading mathematics journals, those included in the database are, in fact, cited less than those not included.[7] The impact factor values of the latter journals will be substantially underestimated, because their journal self-citations[18] are not recorded in the database.

Journal impact factors are research field-dependent

The differences in citation habits and citation dynamics between research fields (as discussed above) are reflected in the journal impact factors, which are highly field-dependent. Table 6 shows that research fields may differ up to eightfold in their median journal impact factor value. (These unweighted median values can serve only as rough indicators: since a field like biochemistry has a major fraction of its papers published in high-impact journals, its mean field impact factor is actually more than twice as high as the journal median value given in Table 6). Basic biomedicine tends to top the list, clinical medicine has an intermediate position and technical fields show the lowest impact factors. The difference between the fields' top journals is even greater (nearly 50-fold), to some extent depending on field size: large fields have a wider range of impact factors and hence achieve higher values for their top journals. Scientists in such fields will have both a higher mean impact and a better access to high-impact journals than scientists in low-impact fields.

Why publish in high-impact journals?

It is obvious that so long as journal impact factors (or journal reputation in general) continue to be used as a major criterion for evaluation of scientific quality, it will be unwise to publish in low-impact journals, regardless of their scientific suitability. From a purely scientific point of view, however, the journal impact may be relatively unimportant. In a study comparing groups of often cited and less cited scientists who published in similar journals, it was found that the twofold difference in citation rate between the two groups persisted throughout the journal impact range, indicating that the high-impact journals bestowed no «free citations» on their authors.[41] Similarly, for cardiology papers authored by international working groups and published simultaneously in high-impact American journals and low-impact European journals, only a fraction of the impact difference (<20%) rubbed off on the papers[42] (probably reflecting a national bias in journal selection, since a choice was available in this particular case). It would thus seem that scientific papers receive their due citations largely independently of the journals in which they appear, i.e., the journal impact is determined by the articles, not vice versa. Scientists should, therefore, feel confident in making their journal choice on the basis of scientific suitability and the quality of the editorial process, as they did in the past,[43] rather than looking to the journals' impact factors - provided the unfortunate use of the latter in evaluation can be brought to a halt.

References

1. Cole S, Cole JR, Simon GA. Chance and consensus in peer review. Science 1981; 214: 881-886.
2. Ernst E, Saradeth T, Resch KL. Drawbacks of peer review. Nature 1993; 363: 296.
3. Hansen HF, Jørgensen BH. Styring af forskning: Kan forskningsindikatorer anvendes? Frederiksberg, Denmark, Samfundslitteratur; 1995: 1-204.
4. Taubes G. Measure for measure in science. Science 1993; 260: 884-886.
5. Andersen H. ACTA Sociologica på den internationale arena – hvad kan SSCI fortælle? Dansk Sociologi 1996; 2: 72-78.

6. Moed HF, Burger WJM, Frankfort JG, Van Raan AFJ. On the Measurement of Research Performance: The Use of Bibliometric Indicators. Leiden, Science Studies Unit, LISBON-Institute, University of Leiden; 1987:1-199.
7. Korevaar JC, Moed HF. Validation of bibliometric indicators in the field of mathematics. Scientometrics 1996; 37: 117-130.
8. Moed HF, Burger WJM, Frankfort JG, Van Raan AFJ. The application of bibliometric indicators: important field- and time-dependent factors to be considered. Scientometrics 1985; 8: 177-203.
9. Sivertsen G. Norsk Forskning på den Internasjonale Arena. En Sammenlikning av 18 OECD-lands Artikler og Siteringer i Science Citation Index 1973-86, Oslo, Inst.Stud.Res.Higher Educat.; 1991: 1-90.
10. Andersen H. Tidsskriftpublicering og forskningsevaluering. Biblioteksarbejde 1996; 17: 19-26.
11. Artus HM. Science indicators derived from databases. Scientometrics 1996; 37: 297-311.
12. Lange L. Effects of disciplines and countries on citation habits. An analysis of empirical papers in behavioural sciences. Scientometrics 1985; 8:2 05-215.
13. Narin F, Hamilton KS. Bibliometric performance measures. Scientometrics 1996; 36: 293-310.
14. Braun T, Glänzel W, Grupp H. The scientometric weight of 50 nations in 27 science areas, 1989-1993. Part II. Life sciences. Scientometrics 1996; 34: 207-237.
15. Møller AP. National citations. Nature 1990; 348: 480.
16. Wade N. Citation analysis: A new tool for science administrators. Science 1975; 188: 429-433.
17. Evans JT, Nadjari HI, Burchell SA. Quotational and reference accuracy in surgical journals. A continuing peer review problem. J Am Med Ass 1990;263:1353-1354.
18. Seglen PO. Kan siteringsanalyse og andre bibliometriske metoder brukes til evaluering av forskningskvalitet? NOP-Nytt (Helsingfors) 1989; 15: 2-20.
19. MacRoberts MH, MacRoberts BR. Problems of citation analysis: a critical review. J Am Soc Information Sci 1989; 40: 342-349.
20. Hamilton DP. Research papers: Who's uncited now. Science 1991; 251: 25.
21. Gomperts MC. The law of constant citation for scientific literature. J Documentation 1968; 24: 113-117.
22. Seglen PO. Bruk av siteringsanalyse og andre bibliometriske metoder i evaluering av forskningsaktivitet. Tidsskr Nor Laegeforen 1989; 31: 3229-3234.
23. Hargens LL, Felmlee DH. Structural determinants of stratification in science. Am Sociol Rev 1984; 49: 685-697.
24. Vinkler P. Relationships between the rate of scientific development and citations. The chance for a citedness model. Scientometrics 1996; 35: 375-386.
25. Narin F, Pinski G, Gee HH. Structure of the biomedical literature. J Am Soc Inform Sci 1976; 27: 25-45.
26. Folly G, Hajtman B, Nagy JI, Ruff I. Some methodological problems in ranking scientists by citation analysis. Scientometrics 1981; 3: 135-147.
27. Seglen PO. Evaluation of scientists by journal impact, in Weingart P, Sehringer R, Winterhager M (eds): Representations of Science and Technology. Leiden, DSWO Press; 1992: 240-252.
28. Lewison G, Anderson J, Jack J. Assessing track records. Nature 1995; 377: 671.
29. Vinkler P. Model for quantitative selection of relative scientometric impact indicators. Scientometrics 1996; 36: 223-236.
30. Garfield E. Citation analysis as a tool in journal evaluation. Science 1972; 178: 471-479.
31. Vinkler P. Evaluation of some methods for the relative assessment of scientific publications. Scientometrics 1986; 10: 157-177.
32. Magri M-H, Solari A. The SCI Journal Citation Reports: A potential tool for studying journals? I. Description of the JCR journal population based on the number of citations received, number of source items, impact factor, immediacy index and cited half-life. Scientometrics 1996; 35: 93-117.
33. Moed HF, Van Leeuwen TN, Reedijk J. A critical analysis of the journal impact factors of Angewandte Chemie and The Journal of the American Chemical Society. Inaccuracies in published impact factors based on overall citations only. Scientometrics 1996; 37: 105-116.
34. Moed HF, Van Leeuwen TN. Impact factors can mislead. Nature 1996; 381: 186.
35. Bourke P, Butler L. Standard issues in a national bibliometric database: the Australian case. Scientometrics 1996; 35: 199-207.
36. Hayes SC. When more is less: Quantity versus quality of publications in the evaluation of academic vitae. Am Psychologist 1983; 38: 1398-1400.
37. Sengupta IN. Three new parameters in bibliometric research and their application to rerank periodicals in the field of biochemistry. Scientometrics 1986;10:235-242.
38. Semenza G. Editorial. FEBS Lett 1996; 389: 1-2.
39. Metcalfe NB. Journal impact factors. Nature 1995; 377: 260-261.
40. Sorokin NI. Impact factors of Russian journals. Nature 1996; 380: 578.
41. Seglen PO. Causal relationship between article citedness and journal impact. J Am Soc Information Sci 1994; 45: 1-11.
42. Opthof T. Sense and nonsense about the impact factor. Cardiovascular Res 1997; 33: 1-7.
43. Gordon MD. How authors select journals: A test of the reward maximization model of submission behaviour. Social Studies of Science 1984; 14: 27-43.
44. Seglen PO. The skewness of science. J Am Soc Information Sci 1992; 43: 628-638.