

Review article

Citations and journal impact factors: questionable indicators of research quality

Seglen PO. Citations and journal impact factors: questionable indicators of research quality.

Allergy 1997; 52: 1050–1056. © Munksgaard 1997.

P. O. Seglen

Institute for Studies in Research and Higher Education (NIFU), Oslo, and Institute for Cancer Research, Norwegian Radium Hospital, Oslo, Norway

Dr Per O. Seglen
Institute for Studies in Research and Higher Education
Hegdehaugsveien 31
N-0352 Oslo
Norway

Accepted for publication 9 June 1997

Evaluation of research quality is one of the most important, yet one of the most difficult aspects of the scientific process. The traditional method of evaluation is peer review, whereby research plans or scientific results are scrutinized by scientific colleagues, assumed to possess the necessary expertise. However, peer review is all too often performed either by narrow specialists who are unable to compare different projects, or by people with broad scientific qualifications, but without the specific insight required to evaluate the ingenuity and originality of results generated within a scientific specialty. As a consequence, evaluation committees tend to produce reports lacking in internal consistency and fairness as well as in quality. Resort is often had to secondary, superficial criteria, such as personal or institutional reputation, project relevance, journal prestige, crude publication counts, etc., giving the whole process the character of a lottery (1–4).

Given the problems with peer review, it is not surprising that alternative evaluation methods have been sought. Particular attention has been given to *bibliometric* methods, which, on first impression, appear to be based on objective, quantitative indicators, such as citation rates and journal impact factors.

What is a citation?

The Institute for Scientific Information (ISI) in Philadelphia, USA, has created a large database,

which continuously encodes all references found in the reference lists of a large number of journals within the medical and natural sciences. From this database, ISI publishes an annual list, the *Science Citation Index (SCI)*, which shows how many times the various references (arranged alphabetically by first author) have been cited during the last year, and by whom. Although it was originally intended as a literature research tool, the fact that the *SCI* can provide annual (or cumulative) citation rates for any article or first author has made it tempting to use it for research evaluation, on the assumption that citations indicate scientific quality. The database can also provide mean citation rates for journals, known as *journal impact factors*, which are published separately by ISI as the annual *SCI Journal Citation Reports*. The journal impact factors are widely used as indicators of journal quality, and have also – unfortunately – found widespread use in the quality evaluation of individual scientists and research groups. How justified is this usage of citation data?

The use of citations as a quality measure is founded on the assumption that authors select their references on the basis of quality. However, as practicing scientists, we know that we primarily refer to publications which we use in our own work. The primary criterion for reference selection is thus *utility within research* rather than pure scientific quality. However, journal space limitations prevent us from citing all the sources we draw upon: it has

been estimated that only some 30% of the literature base of a scientific paper is acknowledged in the form of citations (5). The selection that has to be made leaves room for a mixed bag of secondary citer motives (6, 7) (Table 1).

Many scientists have poor knowledge of the primary literature in their field, and may fail to cite seminal source material (8). Resort is often had to reviews and other secondary sources, which are, therefore, generally highly cited (9, 10). 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 (10, 11): for example, the originator of an analytic 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 are usually about 30% of the citations, and often more (12, 13).

Database problems

Several technical problems are associated with the recording and retrieval of citation data (Table 2). About 4500 journals are included in the ISI science database, which is but a small fraction of the world total of 126 000 (14). Different research fields are covered unequally: for example, the coverage of chemistry has been estimated to be 90%, as compared to 30% for biology (12). The database purportedly includes the most important journals, but the journals selected from a given field are not necessarily more cited than those which remain outside the database (15). The set of journals chosen for inclusion in the database may vary from one year to another (16, 17); furthermore, different database products may differ in contents (18). The database does not include books as source items, despite their prominent role in many research fields (19). An Australian study reported that fewer

Table 2. Problems with ISI database

1. The journal coverage in the database is incomplete and somewhat random
2. Different research fields are unequally covered by the database
3. The journal set included in the database may vary with time
4. Books are not included as source items in the database
5. The database has a language bias for English
6. The database is biased for US science
7. Retrieval is only by first author from the standard version of the database
8. On-line access to the database is expensive
9. Different database products differ in contents
10. Delayed registration of citations
11. Many misprints (up to 25%)
12. Inconsistent foreign language spelling (e.g., α , $\varnothing$, $\ddot{a}$)
13. Synonymy (several variants of the same article)
14. Homonymy (several authors with the same name, as in Japan)

than 20% of the country's academic publications were published in journals covered by the ISI database (20).

Such an incomplete database coverage of any significantly self-contained research field, country, or language will cause a corresponding reduction in the recorded, as opposed to the real, citation rate. The ISI database has a clear preference for English-language, in particular North American, journals (12): for example, the database's social sciences section was shown to contain only two German social science journals, as compared to 542 in a German database (21). This language discrimination is probably a major reason for the low recorded citation rates of, e.g., German and Russian journals: in the 1988 *ISI Journal Citation Reports*, the journals with the prefix *Z* (for *Zeitung*, i.e., German) had a mean citedness (impact factor) of 0.88, and those with the prefix *Zh* (i.e., Russian) had an impact factor of 0.37, as compared to 1.93 for journals with the prefix *Am* (i.e., American). The language bias in the database is compounded by the strong tendency of authors to cite selectively articles in their own national language (14, 22, 23). In fact, about one-half of all citations are to US scientists (24, 25), who are particularly prone to cite each other (14, 23, 26), thereby helping to raise the citation rate of US scientists to 30% above the world average (24).

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. Entry of references into the database is often delayed, so that articles and citations published by the end of the year may be missing from the annual index. Misprints occur frequently (about 25%), often reflecting errors in the original reference lists (27, 28). In one case, misprints caused the same reference to be listed in the *SCI* under 70 different synonyms (13). Homonymy (same name) is

Table 1. Motives and problems in reference selection

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

another problem, especially among Japanese authors; for example, “K. Suzuki” is credited in the database with several hundred publications annually.

Research-field effects

One major problem in the use of citations for evaluation purposes is that citations cannot be compared across different research fields (Table 3). Scientific disciplines which use many references per article, such as biochemistry, have higher citation rates than a field such as mathematics, which uses few references. If this is combined with a short-term recording window to favor current, but short-lived articles, the average biochemist will be cited four times as often as the average mathematician (16). In the arts and humanities, article references are given infrequently, meaning that the vast majority of articles in these research fields are never cited (29).

The size of the research field affects citation rates in several ways. In very small, self-contained fields, the citation rates can be expected to be proportional to the number of authors within the field, but only up to a limit set by the length of the reference list. In large research fields, the mean citation rate should be independent of field size (30). However, the range of citedness will be wider in a large field, giving a better opportunity for a few to become highly cited (31). In rapidly expanding fields, citation rates will be high, because the

number of citers is high relative to the amount of citable material (32, 33).

The tendency of a research field to be cited in adjacent fields is probably the most important of the field effects. For example, clinical medicine draws heavily upon basic science, but not vice versa. As a result, papers in basic medical fields are cited more often than those in clinical medicine (Table 4) (31, 34, 35).

There is evidence that field effects extend even to the subdiscipline level (36). Therefore, the choice of research theme will determine, *a priori*, the probability of becoming highly cited. Thus, the citation rates of scientists working on different subjects cannot be compared, even within the same field. Attempts have been made to correct for field effects by dividing article citation rates by journal citation rates (journal impact factors) (12), or by using complex field citation factors (37, 38), but such corrections may introduce new biases, e.g., by punishing authors publishing in highly cited journals. The construction of individualized field factors is not feasible, and probably not very useful: citation rates are determined by so many technical factors that pure scientific quality may be a very minor influence.

Journal impact factors

Despite the unsuitability of citations as a measure of research quality, an indirect citation measure, the so-called journal impact factor (39), has found increasing application in research evaluation (4, 40, 41). The journal impact factor is an indicator of journal citedness, calculated as the mean number of citations, in a given year, to journal items published during the preceding 2 years. The impact factors of the journals in the ISI database are published annually in the *SCI Journal Citation Reports*; these factors are, therefore, readily available and

Table 3. Research-field effects that affect citation rates

- 1. Number of references per article within the field
- 2. Obsolescence of references relative to time window for citation recording
- 3. Field size (determines the maximal citation rate obtainable)
- 4. Field dynamics (field expansion or contraction)
- 5. Interfield relations (e.g., basic vs applied)
- 6. Subfield microheterogeneity

Table 4. Citation rates (impact factors) of various biomedical research fields

Research field	No. of ISI journals	No. of articles	No. of citations	Field citedness (citations/article)
Biochemistry	130	53 501	188 597	3.53
Immunology	67	20 882	64 841	3.11
Genetics	48	11 057	24 663	2.23
Oncology	49	18 591	34 205	1.84
Anesthesiology	11	3092	5231	1.69
Urology and nephrology	25	6591	9245	1.40
Pediatrics	41	9610	11 408	1.19
Ophthalmology	27	6641	7670	1.15
Surgery	50	19 313	21 976	1.14
Orthopedics	10	2905	2490	0.86
World total	4398	1 038 485	1 550 778	1.49

Data assembled from *SCI Journal Citation Reports*, 1988, which records citations in 1988 to journal articles in 1986–7.

Table 5. Examples of journal impact factors

Journal	
<i>Cell</i>	30.2
<i>N Engl J Med</i>	23.2
<i>Nature</i>	19.3
<i>EMBO J</i>	12.4
<i>Proc Natl Acad Sci U S A</i>	10.3
<i>J Immunol</i>	7.0
<i>BMJ</i>	3.9
<i>Biochim Biophys Acta</i>	2.5
<i>Allergy</i>	1.1
<i>Acta Chir Scand</i>	0.4
<i>Eur J Surg Oncol</i>	0.0

Data from *SCI Journal Citation Reports*, 1991.

easy to use (Table 5 gives some examples to indicate the range in factor values).

The basic premise for the use of journal impact factors in evaluation is that the journal is representative of its articles, in which case 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 distribute in a Gaussian fashion around the population mean, i.e., the journal impact factor. However, the actual citedness distributions of articles from three different journals were found to be extremely skewed, only a few articles being anywhere near the population

mean (Fig. 1) (42). A cumulative plot revealed that about 15% of the journal articles accounted for about 50% of the citations, and that the most cited half of the articles was cited, on average, about 10 times as often as the least cited half (Fig. 1B). Assignment of the same score (the journal impact factor) to all articles would mask these tremendous differences, thus defeating the very purpose of an evaluation. Since the impact factors are clearly not representative of the individual articles, it is hardly surprising that the correlation between an article's actual citedness and the impact factor of its journal is often very poor (Fig. 2).

Biased calculation of journal impact factors

The limitations associated with the use of citations in evaluation largely apply also to journal impact factors, but there are a few more in addition (Table 6). Although the impact factors are computed by including citations of all types of journal documents in the numerator, items such as editorials, letters, and meeting abstracts are excluded from the denominator (43, 44). As a result, journals with interesting editorials and a lively correspondence section may have their impact factors inflated by up to 75% (45). The ISI database does not correct for self-citations; editors can, therefore, raise the impact of their journals by referring frequently to their previous editorials. Other impact-maximizing

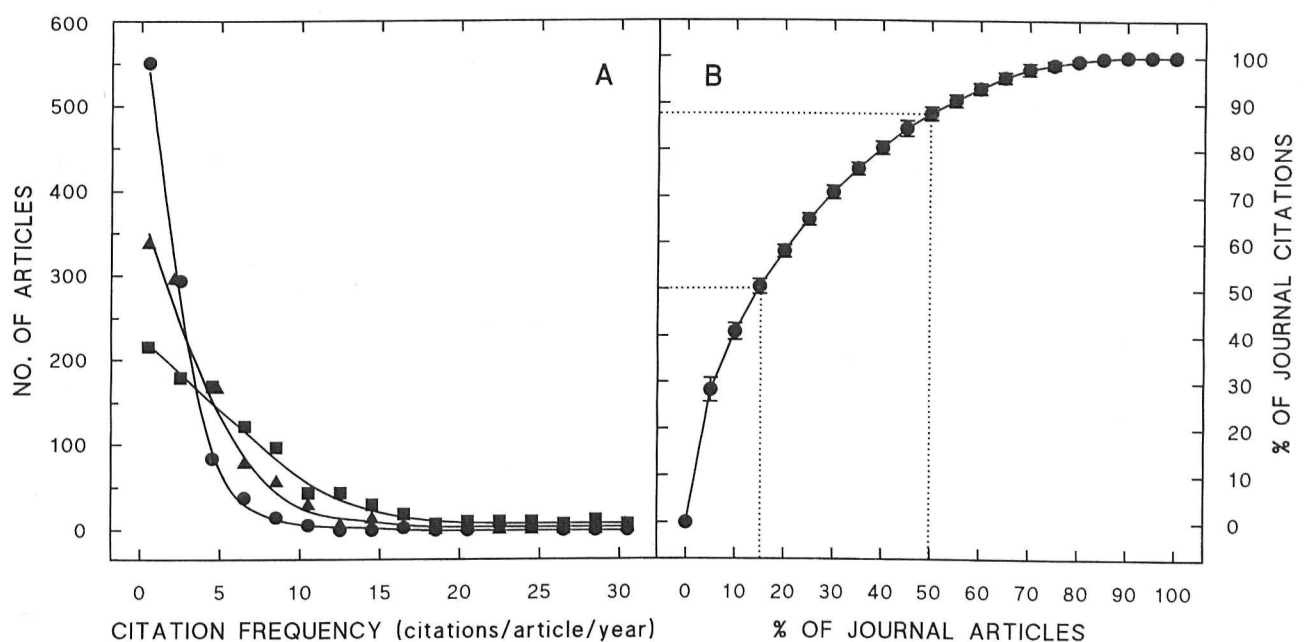


Fig. 1. Skewed distribution of citations to articles within journal. A) Third-year citedness of articles in *Biochim Biophys Acta* (●), *Biochem J* (▲), and *J Biol Chem* (■), arranged by number of articles in each citedness category. B) Cumulative contribution of different citedness categories (beginning with most cited 5%) to total journal impact. Each value is mean $\pm$ SE of three journals in panel A. Dotted lines indicate contributions of 15% and 50% most cited articles. From Seglen (42).

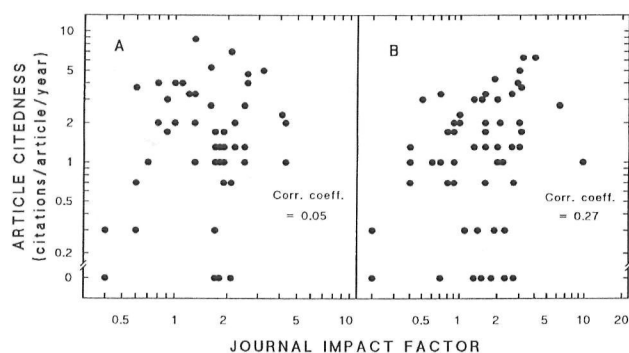


Fig. 2. Correlation between article citedness and journal impact for two individual authors. From Seglen (51).

measures are the inclusion of review articles, which are generally highly cited (44, 46), and of long, rather than short articles, since citedness is roughly proportional to article length (Fig. 3) (36). Relative to the length of articles, biochemical journals of the "short communication" type are actually cited at least as highly as the leading high-impact journals within their field (47, 48).

The journal impact factor is a short-term index, based on citations recorded during the first 2 years after publication. For journals with short publication lags, a high proportion of the journal self-citations (many of which are author self-citations) will be included, giving the journal a correspondingly high impact factor (49). Russian journals, which are mainly cited by other Russian journals (22), have particularly long publication lags, resulting in low impact factors (50). On the other hand, US journals, which quantitatively dominate the ISI database, have particularly high impact factors as a result of national citation bias.

Journals in the many research fields which are underrepresented in the database (12) receive too

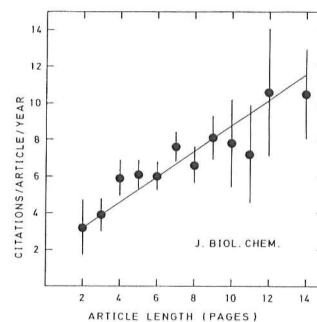


Fig. 3. Dependence of citedness on article length. Third-year citations to *J Biol Chem* articles, given as function of article length (number of pages). Each value is mean $\pm$ SEM of up to 86 articles of equal length. From Seglen (36).

low impact factors. In some fields, books are an important vehicle of publication, but citations in books are not recorded, and therefore do not contribute to journal impact factor values (19). The source journal set in the ISI database is not necessarily selected on the basis of citedness: among leading mathematics journals, those included in the database were actually less cited than those which were not included (15). The impact factors of the latter would even be significantly underestimated, due to the exclusion of their journal self-citations, which are a substantial fraction of the citations (13).

Research-field dependence of journal impact factors

Research-field differences in citation characteristics, such as between the medical fields shown in Table 6, also apply to the fields' journal impact factors, the mean field citedness and the mean journal impact being the same. Basic medical journals tend to have high impact factors, clinical medical journals intermediate ones, and technical fields very low ones. Differences in field size accentuate the heterogeneity, since large fields display a wider range of impact factors, and hence achieve higher values for their top journals (31). For example, there is a 70-fold difference between the top journal in biochemistry (*Annu Rev Biochem*, impact factor 35.6 in 1991) and that in aerospace engineering (*AIAA J*, impact factor 0.5 in 1991). Thus, scientists in large, high-impact fields will have both a higher mean citedness and better access to high-impact journals than scientists in small, low-impact fields.

Does the journal influence the citedness of the article?

A recent study which compared highly cited and less cited scientists who published in similar

Table 6. Problems associated with use of journal impact factors in evaluation

1. Journal impact factors are not representative of the individual journal articles
2. Journal 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 "noncitable" 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 North 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

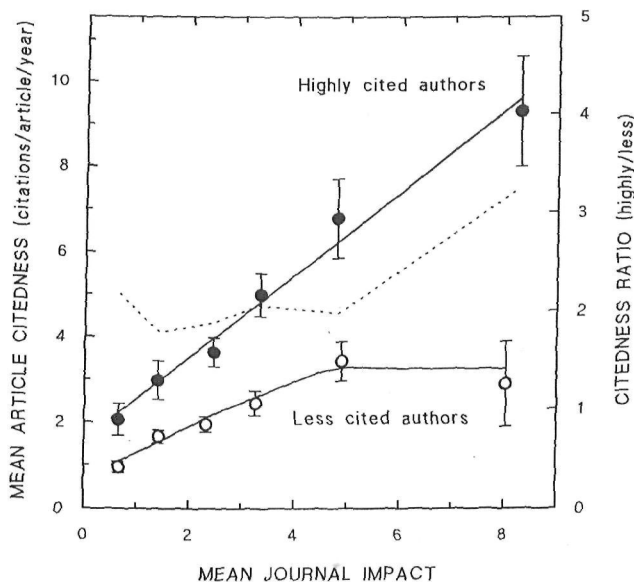


Fig. 4. Article citedness as function of journal impact: comparison between highly cited and less cited authors. Articles written by highly cited (●) or less cited (○) authors were assembled and sorted into six journal-impact categories, each value representing mean $\pm$ SE of from 7 to 145 articles. Dotted line: ratio of citedness between two author groups. From Seglen (51).

journals found that the twofold citedness ratio between the two groups was maintained throughout all journal impact categories, indicating that the high-impact journals did not contribute any "free citations" to the articles they contained (Fig. 4) (51). In the case of internationally coauthored cardiology articles published simultaneously in North American high-impact journals and European low-impact journals, less than 20% of the journal impact difference (probably representing the remaining national bias) was reflected in the citedness of the articles (52). Scientific publications thus appear to receive their citations largely independently of the journals in which they appear, i.e., the journal impact is determined by the articles, not vice versa. From a purely scientific point of view, therefore, the choice of a journal for publication can safely be based on criteria such as scientific suitability and fairness in editorial processing, rather than on impact factors. However, as long as journal impact factors or journal reputation continue to be used in science evaluation, scientific authors will shun journals of inferior impact, regardless of their quality. Editors of journals in low-impact fields (clinical; technical) should resist the temptation to raise their journal impact artificially (by the means discussed above), and instead move to halt the detrimental misuse of impact factors. Science deserves to be judged by its contents, not by its wrapping.

References

1. Cole S, Cole JR, Simon GA. Chance and consensus in peer review. *Science* 1981;214:881-6.
2. Peters DP, Ceci SJ. Peer-review practices of psychological journals: the fate of published articles, submitted again. *Behav Brain Sci* 1982;5:187-95.
3. Ernst E, Saradeth T, Resch KL. Drawbacks of peer review. *Nature* 1993;363:296.
4. Hansen HF, Jørgensen BH. Styring af forskning: kan forskningsindikatorer anvendes? Frederiksberg, Denmark: Samfundslitteratur, 1995:1-204.
5. MacRoberts MH, MacRoberts BR. Problems of citation analysis: a critical review. *J Am Soc Inform Sci* 1989;40: 342-9.
6. Brooks TA. Evidence of complex citer motivations. *J Am Soc Inform Sci* 1986;37:34-6.
7. Liu M. Progress in documentation – the complexities of citation practice: a review of citation studies. *J Documentation* 1993;49:370-408.
8. MacRoberts MH, MacRoberts BR. Quantitative measures of communication in science: a study of the formal level. *Soc Stud Sci* 1986;16:151-72.
9. MacRoberts MH, MacRoberts BR. Another test of the normative theory of citing. *J Am Soc Inform Sci* 1987;38: 305-6.
10. MacRoberts MH, MacRoberts BR. Problems of citation analysis. *Scientometrics* 1996;36:435-44.
11. Tainer JA. Science, citation, and funding. *Science* 1991; 251:1408.
12. Moed HF, Burger WJM, Frankfort JG, van Raan AFJ. On the measurement of research performance: the use of bibliometric indicators. Leiden, Science Studies Unit, LIS-BON-Institute, University of Leiden: 1987:1-199.
13. Seglen PO. Kan siteringsanalyse og andre bibliometriske metoder brukes til evaluering av forskningskvalitet? *NOP-Nytt (Helsingfors)* 1989;15:2-20.
14. Andersen H. ACTA Sociologica på den internasjonale arena – hvad kan SSCI fortælle? *Dansk Sociol* 1996;2:72-8.
15. Korevaar JC, Moed HF. Validation of bibliometric indicators in the field of mathematics. *Scientometrics* 1996;37: 117-30.
16. 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.
17. Sivertsen G. Norsk Forskning på den Internasjonale Arena. En Sammenlikning av 18 OECD-lands Artikler og Siteringer i Science Citation Index 1973-86. Oslo: Institute for Studies and Research in Higher Education, 1991:7-90.
18. Moed HF. Differences in the construction of SCI based bibliometric indicators among various producers: a first overview. *Scientometrics* 1996;35:177-91.
19. Andersen H. Tidsskriftpublicering og forskningsevaluering. *Biblioteksarbejde* 1996;17:19-26.
20. Murphy PS. From bibliometrics to research productivity: benchmarks and self-assessment. In: Koenig MED, Bookstein A, editors. *Proceedings of the Fifth Biennial Conference of the International Society for Scientometrics and Informetrics*. Medford, NJ: Learned Information, Inc., 1995:349-55.
21. Artus HM. Science indicators derived from databases. *Scientometrics* 1996;37:297-311.
22. Lange L. Effects of disciplines and countries on citation habits. An analysis of empirical papers in behavioural sciences. *Scientometrics* 1985;8:205-15.
23. Narin F, Hamilton KS. Bibliometric performance measures. *Scientometrics* 1996;36:293-310.

24. Braun T, Glänzel W, Grupp H. The scientometric weight of 50 nations in 27 science areas, 1989-1993. II. Life sciences. *Scientometrics* 1996;34:207-37.
25. May RM. The scientific wealth of nations. *Science* 1997; 275:793-6.
26. Møller AP. National citations. *Nature* 1990;348:480.
27. Wade N. Citation analysis: a new tool for science administrators. *Science* 1975;188:429-33.
28. Evans JT, Nadjari HI, Burchell SA. Quotational and reference accuracy in surgical journals. A continuing peer review problem. *JAMA* 1990;263:1353-4.
29. Hamilton DP. Research papers: who's uncited now? *Science* 1991;251:25.
30. Gomperts MC. The law of constant citation for scientific literature. *J Documentation* 1968;24:113-17.
31. Seglen PO. Bruk av siteringsanalyse og andre bibliometriske metoder i evaluering av forskningsaktivitet. *Tidsskr Nor Laegeforen* 1989;31:3229-34.
32. Hargens LL, Felmlee DH. Structural determinants of stratification in science. *Am Sociol Rev* 1984;49:685-97.
33. Vinkler P. Relationships between the rate of scientific development and citations. The chance for a citedness model. *Scientometrics* 1996;35:375-86.
34. Narin F, Pinski G, Gee HH. Structure of the biomedical literature. *J Am Soc Inform Sci* 1976;27:25-45.
35. Folly G, Hajtman B, Nagy JI, Ruff I. Some methodological problems in ranking scientists by citation analysis. *Scientometrics* 1981;3:135-47.
36. Seglen PO. Evaluation of scientists by journal impact. In: Weingart P, Sehringer R, Winterhager M, editors. *Representations of science and technology*. Leiden: DSWO Press, 1992:240-52.
37. Lewison G, Anderson J, Jack J. Assessing track records. *Nature* 1995;377:671.
38. Vinkler P. Model for quantitative selection of relative scientometric impact indicators. *Scientometrics* 1996;36: 223-36.
39. Garfield E. Citation analysis as a tool in journal evaluation. *Science* 1972;178:471-9.
40. Vinkler P. Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics* 1986; 10:157-77.
41. Taubes G. Measure for measure in science. *Science* 1993; 260:884-6.
42. Seglen PO. The skewness of science. *J Am Soc Inform Sci* 1992;43:628-38.
43. 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.
44. 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;37:105-16.
45. Moed HF, van Leeuwen TN. Impact factors can mislead. *Nature* 1996;381:186.
46. Bourke P, Butler L. Standard issues in a national bibliometric database: the Australian case. *Scientometrics* 1996; 35:199-207.
47. Sengupta IN. Three new parameters in bibliometric research and their application to rerank periodicals in the field of biochemistry. *Scientometrics* 1986;10:235-42.
48. Semenza G. Editorial. *FEBS Lett* 1996;389:1-2.
49. Metcalfe NB. Journal impact factors. *Nature* 1995;377:260-1.
50. Sorokin NI. Impact factors of Russian journals. *Nature* 1996;380:578.
51. Seglen PO. Causal relationship between article citedness and journal impact. *J Am Soc Inform Sci* 1994;45:1-11.
52. Opthof T. Sense and nonsense about the impact factor. *Cardiovasc Res* 1997;33:1-7.

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.