

Education and debate

Why the impact factor of journals should not be used for evaluating research

Per O Seglen, *professor*^a

^a Institute for Studies in Research and Higher Education (NIFU) Hegdehaugsveien 31 N-0352 Oslo Norway

Introduction

Evaluating scientific quality is a notoriously difficult problem which has no standard solution. Ideally, published scientific results should be scrutinised by true experts in the field and given scores for quality and quantity according to established rules. In practice, however, what is called peer review is usually performed by committees with general competence rather than with the specialist's insight that is needed to assess primary research data. Committees tend, therefore, to resort to secondary criteria like crude publication counts, journal prestige, the reputation of authors and institutions, and estimated importance and relevance of the research field,¹ making peer review as much of a lottery as of a rational process.^{2 3}

On this background, it is hardly surprising that alternative methods for evaluating research are being sought, such as citation rates and journal impact factors, which seem to be quantitative and objective indicators directly related to published science. The citation data are obtained from a database produced by the Institute for Scientific Information (ISI) in Philadelphia, which continuously records scientific citations as represented by the reference lists of articles from a large number of the world's scientific journals. The references are rearranged in the database to show how many times each publication has been cited within a certain period, and by whom, and the results are published as the *Science Citation Index (SCI)*. On the basis of the *Science Citation Index* and authors' publication lists, the annual citation rate of papers by a scientific author or research group can thus be calculated. Similarly, the citation rate of a scientific journal—known as the journal impact factor—can be calculated as the mean citation rate of all the articles contained in the journal.⁴ Journal impact factors, which are published annually in *SCI Journal Citation Reports*, are widely regarded as a quality ranking for journals and used extensively by leading journals in their advertising.

Summary points

- Use of journal impact factors conceals the difference in article citation rates (articles in the most cited half of articles in a journal are cited 10 times as often as the least cited half)
- Journals' impact factors are determined by technicalities unrelated to the scientific quality of their articles
- Journal impact factors depend on the research field: high impact factors are likely in journals covering large areas of basic research with a rapidly expanding but short lived literature that use many references per article
- Article citation rates determine the journal impact factor, not vice versa

Since journal impact factors are so readily available, it has been tempting to use them for evaluating individual scientists or research groups. On the assumption that the journal is representative of its articles, the journal impact factors of an author's articles can simply be added up to obtain an apparently objective and quantitative measure of the author's scientific achievement. In Italy, the use of journal impact factors was recently advocated to remedy the purported subjectivity and bias in appointments to higher academic positions.⁵ In the Nordic countries, journal impact factors have, on occasion, been used in the evaluation of individuals as well as of institutions and have been proposed, or actually used, as one of the premises for allocation of university resources and positions.^{1 6 1} Resource allocation based on impact factors has also been reported from Canada⁸ and Hungary⁹ and, colloquially, from several other countries. The increasing awareness of journal impact factors, and the possibility of their use in evaluation, is already changing scientists' publication behaviour towards publishing in journals with maximum impact,^{9 10} often at the expense of specialist journals that might actually be more appropriate vehicles for the research in question.

Given the increasing use of journal impact factors—as well as the (less explicit) use of journal prestige—in research evaluation, a critical examination of this indicator seems necessary (see box).

Problems associated with the use of journal impact factors

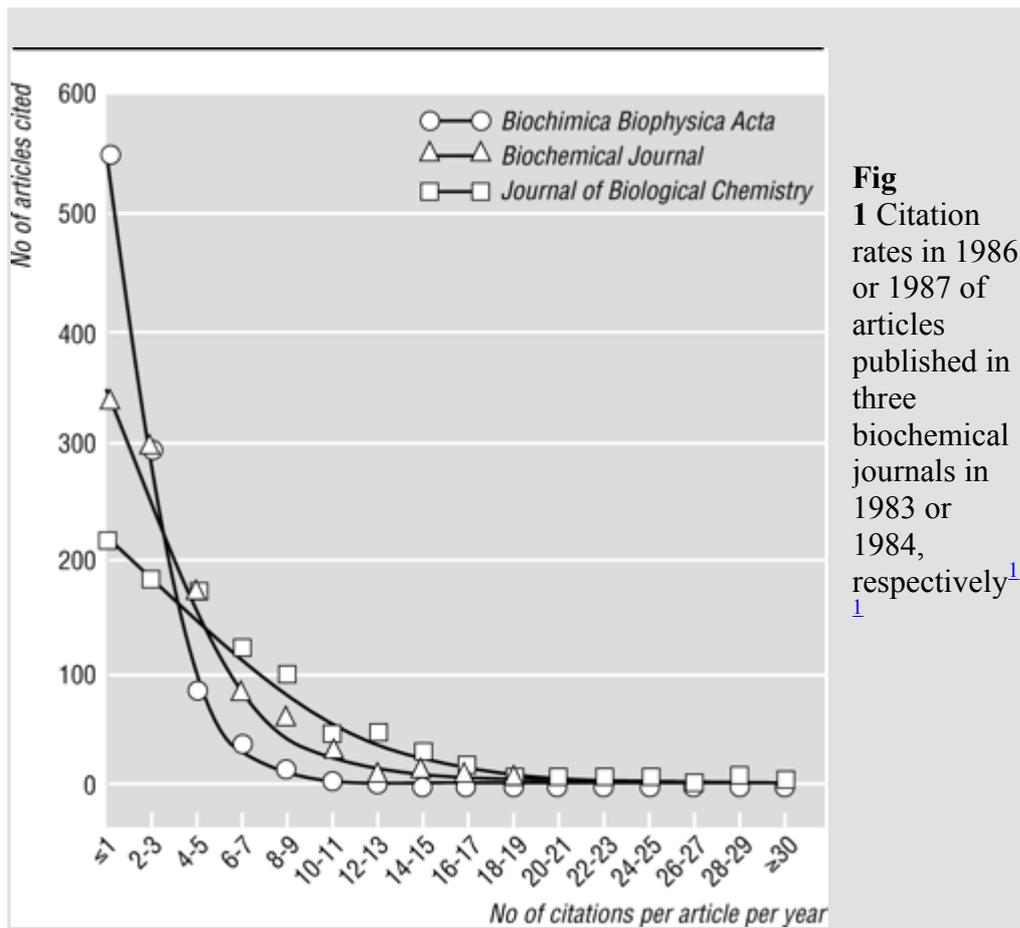
- Journal impact factors are not statistically representative of individual journal articles
- Journal impact factors correlate poorly with actual citations of individual articles
- Authors use many criteria other than impact when submitting to journals
- Citations to "non-citable" items are erroneously included in the database
- Self citations are not corrected for
- Review articles are heavily cited and inflate the impact factor of journals
- Long articles collect many citations and give high journal impact factors
- Short publication lag allows many short term journal self citations and gives a high journal impact factor
- Citations in the national language of the journal are preferred by the journal's authors

- Selective journal self citation: articles tend to preferentially cite other articles in the same journal
- Coverage of the database is not complete
- Books are not included in the database as a source for citations
- Database has an English language bias
- Database is dominated by American publications
- Journal set in database may vary from year to year
- Impact factor is a function of the number of references per article in the research field
- Research fields with literature that rapidly becomes obsolete are favoured
- Impact factor depends on dynamics (expansion or contraction) of the research field
- Small research fields tend to lack journals with high impact
- Relations between fields (clinical v basic research, for example) strongly determine the journal impact factor
- Citation rate of article determines journal impact, but not vice versa

Is the journal impact factor really representative of the individual journal articles?

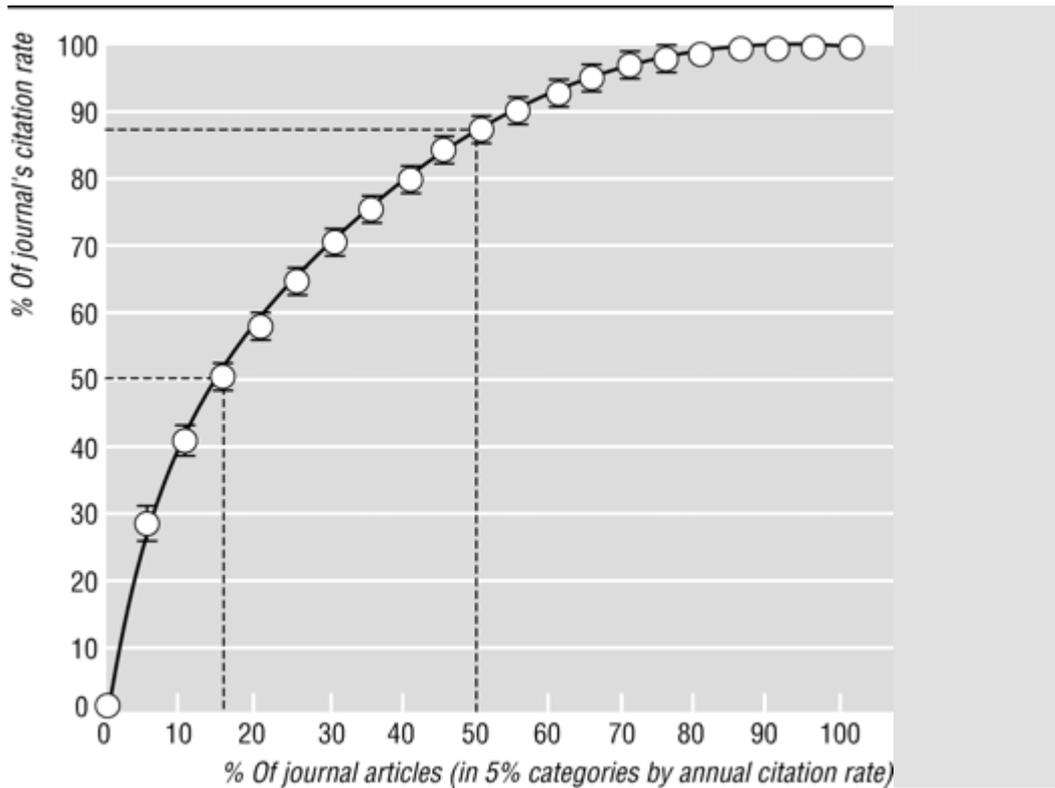
Relation of journal impact factor and citation rate of article

For the journal's impact factor to be reasonably representative of its articles, the citation rate of individual articles in the journal should show a narrow distribution, preferably a Gaussian distribution, around the mean value (the journal's impact factor). Figure 1) shows that this is far from being the case: three different biochemical journals all showed skewed distributions of articles' citation rates, with only a few articles anywhere near the population mean.¹¹



The uneven contribution of the various articles to the journal impact is further illustrated in figure 2): the cumulative curve shows that the most cited 15% of the articles account for 50% of the citations, and the most cited 50% of the articles account for 90% of the citations. In other words, the most cited half of the articles are cited, on average, 10 times as often as the least cited half. Assigning the same score (the journal impact factor) to all articles masks this tremendous difference—which is the exact opposite of what an evaluation is meant to achieve. Even the uncited articles are then given full credit for the impact of the few highly cited articles that predominantly determine the value of the journal impact factor.

Fig 2 Cumulative contribution of articles with different citation rates (beginning with most cited 5%) to total journal impact. Values are mean (SE) of journals in fig 1; dotted lines indicate contributions of 15% and 50% most cited articles¹¹



Since any large, random sample of journal articles will correlate well with the corresponding average of journal impact factors,¹² the impact factors may seem reasonably representative after all. However, the correlation between journal impact and actual citation rate of articles from individual scientists or research groups is often poor^{9 12} (fig 3). Clearly, scientific authors do not necessarily publish their most citable work in journals of the highest impact, nor do their articles necessarily match the impact of the journals they appear in. Although some authors may take journals' impact factors into consideration when submitting an article, other factors are (or at least were) equally or more important, such as the journal's subject area and its relevance to the author's specialty, the fairness and rapidity of the editorial process, the probability of acceptance, publication lag, and publication cost (page charges).¹³

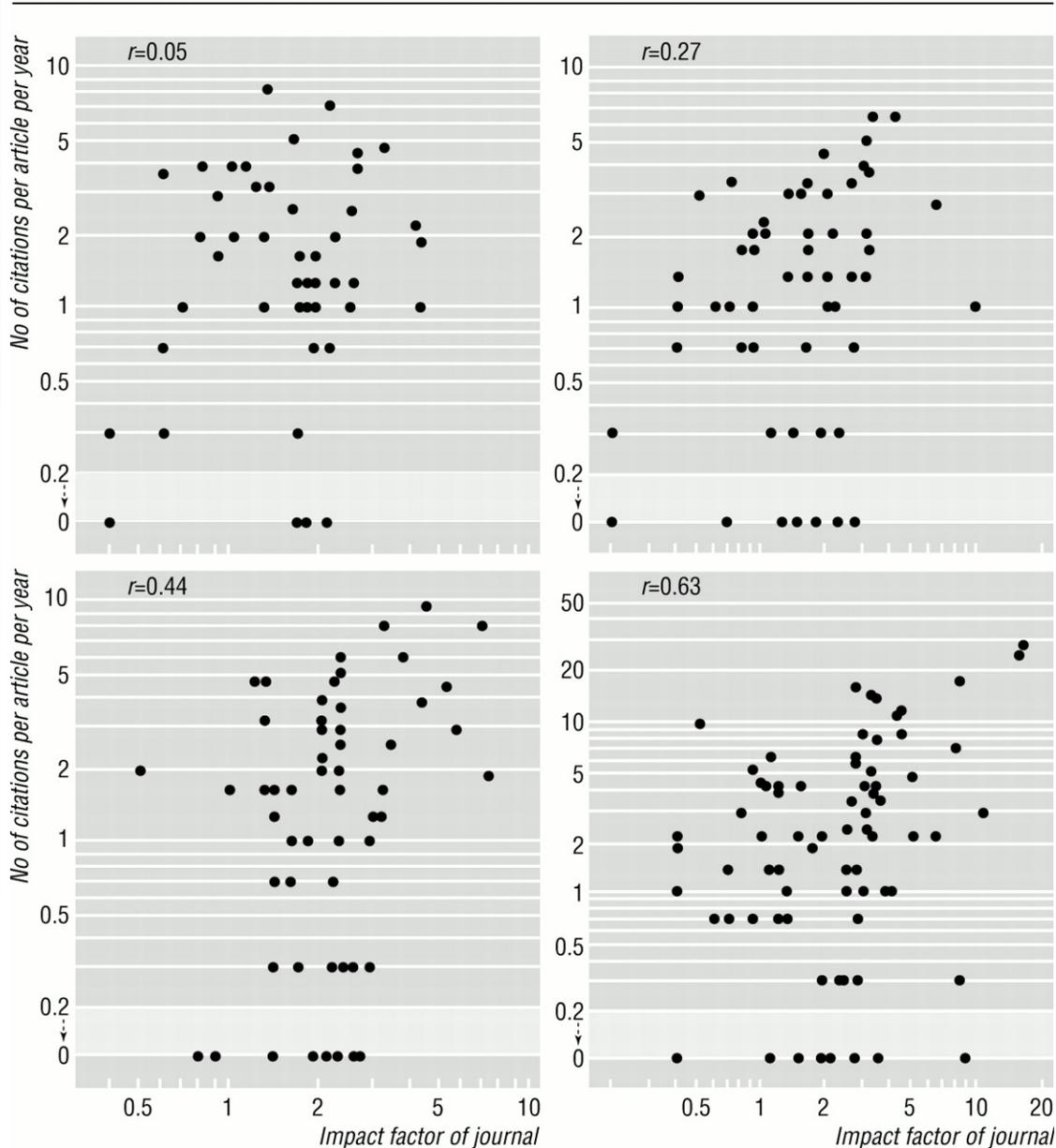


Fig 3 Correlation between article citation rate and journal impact for four authors¹²

Journal impact factors are representative only when the evaluated research is absolutely average (relative to the journals used), a premise which really makes any evaluation superfluous. In actual practice, however, even samples as large as a nation's scientific output are far from being random and representative of the journals they have been published in: for example, during the period 1989-93, articles on general medicine in Turkey would have had an expected citation rate of 1.3 (relative to the world average) on the basis of journal impact, but the actual citation was only 0.3.¹⁴ The use of journal impact factors can therefore be as misleading for countries as for individuals.

Journal impact factors are calculated in a way that causes bias

Apart from being non-representative, the journal impact factor is encumbered with several shortcomings of a technical and more fundamental nature. The factor is generally defined as the recorded number of citations within a certain year (for

example, 1996) to the items published in the journal during the two preceding years (1995 and 1994), divided by the number of such items (this would be the equivalent of the average citation rate of an item during the first and second calendar year after the year of publication). However, the Science Citation Index database includes only normal articles, notes, and reviews in the denominator as citable items, but records citations to all types of documents (editorials, letters, meeting abstracts, etc) in the numerator; citations to translated journal versions are even listed twice.^{15 16 17} Because of this flawed computation, a journal that includes meeting reports, interesting editorials, and a lively correspondence section can have its impact factor greatly inflated relative to journals that lack such items. Editors who want to raise the impact of their journals should make frequent reference to their previous editorials, since the database makes no correction for self citations. The inclusion of review articles, which generally receive many more citations than ordinary articles,^{17 18} is also recommended. Furthermore, because citation rate is roughly proportional to the length of the article,¹⁹ journals might wish to publish long, rather than short, articles. If correction were made for article length, "communications" journals like *Biochemical and Biophysical Research Communications* and *FEBS Letters* would get impact factors as high as, or higher than, the high impact journals within the field, like *Journal of Biological Chemistry*.^{20 21}

The use of an extremely short term index (citations to articles published only in the past two years) in calculating the impact factor introduces a strong temporal bias, with several consequences. For example, articles in journals with short publication lags will contain relatively many up to date citations and thus contribute heavily to the impact factors of all cited journals. Since articles in a given journal tend to cite articles from the same journal,²² rapid publication is self serving with respect to journal impact, and significantly correlated with it.²³ Dynamic research fields with high activity and short publication lags, such as biochemistry and molecular biology, have a correspondingly high proportion of citations to recent publications—and hence higher journal impact factors—than, for example, ecology and mathematics.^{23 24} Russian journals, which are cited mainly by other Russian journals,²⁵ are reported to have particularly long publication lags, resulting in generally low impact factors.²⁶ Pure technicalities can therefore account for several-fold differences in journal impact.

Limitations of the database

The Science Citation Index database covers about 3200 journals⁸; the estimated world total is about 126 000.²⁷ The coverage varies considerably between research fields: in one university, 90% of the chemistry faculty's publications, but only 30% of the biology faculty's publications, were in the database.²⁸ Since the impact factor of any journal will be proportional to the database coverage of its research field, such discrepancies mean that journals from an underrepresented field that are included will receive low impact factors. Furthermore, the journal set in the database is not constant but may vary in composition from year to year.^{24 29} In many research fields a substantial fraction of scientific output is published in the form of books, which are not included as source items in the database; they therefore have no impact factor.³⁰ In mathematics, leading publications that were not included in the Science Citation Index database were cited more frequently than the leading publications that were included.³¹ Clearly, such systematic omissions from the database can cause serious bias in evaluations based on impact factor.

The preference of the Science Citation Index database for English language journals²⁸ will contribute to a low impact factor for the few non-English journals that are included,³² since most citations to papers in languages other than English are given by other papers in the same language.^{25 27 33} The Institute for Scientific Information's database for the social sciences contained only two German social science journals, whereas a German database contained 542.³⁴ Specifically, American scientists, who seem particularly prone to citing each other,^{33 35} dominate these databases to such an extent (over half of the citations) as to raise both the citation rate and the mean journal impact of American science 30% above the world average,¹⁴ the rest of the world then falling below average. This bias is aggravated by the use of a short term index: for example, in American publications within clinical medicine, 83% of references in the same year were to other papers by American scientists (many of them undoubtedly self citations), a value 25% higher than the stable level reached after three years (which would, incidentally, also be biased by self citations and citations of other American work).³³ Thus, both the apparent quality lead of American science and the values of the various journal impact factors are, to an important extent, determined by the large volume, the self citations, and the national citation bias of American science,²⁷ in combination with the short term index used by the Science Citation Index for calculating journal impact factors.

Journal impact factors depend on the research field

Citation habits and citation dynamics can be so different in different research fields as to make evaluative comparisons on the basis of citation rate or journal impact difficult or impossible. For example, biochemistry and molecular biology articles were cited about five times as often as pharmacy articles.³³ Several factors have been found to contribute to such differences among fields of research.

The citation impact of a research field is directly proportional to the mean number of references per article, which varies considerably from field to field (it is twice as high in biochemistry as in mathematics, for example).²⁴ Within the arts and humanities, references to articles are hardly used at all, leaving these research fields (and others) virtually uncited,³⁶ a matter of considerable consternation among science administrators unfamiliar with citation kinetics.³⁷

In highly dynamic research fields, such as biochemistry and molecular biology, where published reports rapidly become obsolete, a large proportion of citations are captured by the short term index used to calculate journal impact factors, as previously discussed³⁸ –but fields with a more durable literature, such as mathematics, have a smaller fraction of short term citations and hence lower journal impact factors. This field property combines with the low number of references per article to give mathematics a recorded citation impact that is only a quarter that of biochemistry.²⁴

In young and rapidly expanding research fields, the number of publications making citations is large relative to the amount of citable material, leading to high citation rates for articles and high journal impact factors for the field.^{39 40}

In a largely self contained research field, the mean article (or journal) citation rate is independent of the size of the field,⁴¹ but the absolute range will be wider in a large field, meaning higher impact factors for the top journals.⁴² Such differences become obvious when comparing review journals, which tend to top their field (table 1)).

Leading scientists in a small field may thus be at a disadvantage compared with their colleagues in larger fields, since they lack access to journals of equally high citation impact.⁴³

Table 1 Journal impact factors and research field

Journal	1986	1987
<i>Annual Review of Biochemistry</i>	31.6	35.1
<i>Annual Review of Immunology</i>	26.5	25.2
<i>Annual Review of Cell Biology</i>	14.1	22.8
<i>Annual Review of Genetics</i>	14.0	14.3
<i>Annual Review of Neuroscience</i>	15.4	13.7
<i>Annual Review of Pharmacology</i>	10.1	9.9
<i>Annual Review of Physiology</i>	7.8	9.1
<i>Annual Review of Biophysics</i>	7.2	7.7
<i>Annual Review of Microbiology</i>	4.9	6.4

Most research fields are, however, not completely self contained, the most important field factor probably being the ability of a research field to be cited by adjacent fields. The relation between basic and clinical medicine is a case in point: clinical medicine draws heavily on basic science, but not vice versa. The result is that basic medicine is cited three to five times more than clinical medicine, and this is reflected in journal impact factors.^{42 44 42} The outcome of an evaluation based on impact factors in medicine will therefore depend on the position of research groups or institutions along the basic-clinical axis.³³

In measures of citation rates of articles, attempts to take research field into account often consist of expressing citation rate relative to some citation impact specific to the field.⁴⁶ Such field corrections range from simply dividing the article's citation rate by the impact factor of its journal²⁸ (which punishes publication in high impact journals) to the use of complex, author specific, field indicators based on reference lists^{47 48} (which punishes citations to high impact journals). However, field corrections cannot readily be applied to journal impact factors, since many research fields are dominated by one or a few journals, in which case corrections might merely generate relative impact factors of unit value. Even within large fields, the tendency of journals to subspecialise with certain subjects is likely to generate significant differences in journal impact: in a single biochemical journal there was a 10-fold difference in citation rates in subfields.¹⁹

▲ Top
▲ Introduction
▲ Is the journal impact...
▪ Is the impact of...
▼ References

Is the impact of an article increased by publication in a high impact journal?

It is widely assumed that publication in a high impact journal will enhance the impact of an article (the "free ride" hypothesis). In a comparison of two groups of scientific authors with similar journal preference who differed twofold in mean citation rate for articles, however, the relative difference was the same (twofold) throughout a range of journals with impact factors of 0.5 to 8.0.¹² If the high impact journals had contributed "free" citations, independently of the article contents, the relative difference would have been expected to diminish as a function of increasing journal impact.⁴⁹ These data suggest that the journals do not offer any free ride. The citation rates of the articles determine the journal impact factor (a truism illustrated by the good correlation between aggregate citation rates of article and aggregate journal impact found in these data), but not vice versa.

If scientific authors are not detectably rewarded with a higher impact by publishing in high impact journals, why are we so adamant on doing it? The answer, of course, is that as long as there are people out there who judge our science by its wrapping rather than by its contents, we cannot afford to take any chances. Although journal impact factors are rarely used explicitly, their implicit counterpart, journal prestige, is widely held to be a valid evaluation criterion⁵⁰ and is probably the most used indicator besides a straightforward count of publications. As we have seen, however, the journal cannot in any way be taken as representative of the article. Even if it could, the journal impact factor would still be far from being a quality indicator: citation impact is primarily a measure of scientific utility rather than of scientific quality, and authors' selection of references is subject to strong biases unrelated to quality.^{51 52} For evaluation of scientific quality, there seems to be no alternative to qualified **experts** reading the publications. Much can be done, however, to improve and standardise the principles, procedures, and criteria used in evaluation, and the scientific community would be well served if efforts could be concentrated on this rather than on developing ever more sophisticated versions of basically useless indicators. In the words of Sidney Brenner, "What matters absolutely is the scientific content of a paper, and nothing will substitute for either knowing or reading it."⁵³

Acknowledgements

Funding: None.

Conflict of interest: None.

References

1. Hansen HF, Jørgensen BH. *Styring af forskning: kan forskningsindikatorer anvendes?* Frederiksberg, Denmark: Samfundslitteratur, 1995.
2. Cole S, Cole JR, Simon GA. Chance and consensus in peer review. *Science* 1981;214:881-6. [[Medline](#)]
3. Ernst E, Saradeth T, Resch KL. Drawbacks of peer review. *Nature* 1993;363:296. [[Medline](#)]
4. Garfield E. Citation analysis as a tool in journal evaluation. *Science* 1972;178:471-9. [[Medline](#)]
5. Calza L, Garbisa S. Italian professorships. *Nature* 1995;374:492.
6. Drettner B, Seglen PO, Sivertsen G. Inverkanstal som fördelningsinstrument: ej accepterat av tidskrifter i Norden. *Läkartidningen* 1994;91:744-5.
7. Gøtzsche PC, Krog JW, Moustgaard R. Bibliometrisk analyse af dansk sundhedsvidenskabelig forskning 1988-1992. *Ugeskr Læger* 1995;157:5075-81.
8. Taubes G. Measure for measure in science. *Science* 1993;260:884-6. [[Medline](#)]
9. Vinkler P. Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics* 1986;10:157-77.
10. Maffulli N. More on citation analysis. *Nature* 1995;378:760.
11. Seglen PO. The skewness of science. *J Am Soc Information Sci* 1992;43:628-38.
12. Seglen PO. Causal relationship between article citedness and journal impact. *J Am Soc Information Sci* 1994;45:1-11.
13. Gordon MD. How authors select journals: a test of the reward maximization model of submission behaviour. *Social Studies of Science* 1984;14:27-43.

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;34:207-37.
15. 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.
16. Moed HF, Van Leeuwen TN. Impact factors can mislead. *Nature* 1996; **381**:186.
17. 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.
18. Bourke P, Butler L. Standard issues in a national bibliometric database: the Australian case. *Scientometrics* 1996;35:199-207.
19. **Seglen** PO. Evaluation of scientists by journal impact. In: Weingart P, Srehinger R, Winterhager M, eds. *Representations of science and technology*. Leiden: DSWO Press, 1992;240-52.
20. Sengupta IN. Three new parameters in bibliometric research and their application to rerank **p**eriodicals in the field of biochemistry. *Scientometrics* 1986;10:235-42.
21. Semenza G. Editorial. *FEBS Lett* 1996;389:1-2.
22. **Seglen** PO. Kan siteringsanalyse og andre bibliometriske metoder brukes til evaluering av forskningskvalitet? *NOP-Nytt (Helsingfors)* 1989;15:2-20.
23. Metcalfe NB. Journal impact factors. *Nature* 1995; **77**:260-1.
24. 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.
25. Lange L. Effects of disciplines and countries on citation habits. An analysis of empirical **p**apers in behavioural sciences. *Scientometrics* 1985;8:205-15.
26. Sorokin NI. Impact factors of Russian journals. *Nature* 1996;380:578.
[\[Medline\]](#)
27. Andersen H. ACTA Sociologica på den internationale arena-hvad kan SSCI fortælle? *Dansk Sociologi* 1996;2:72-8.

28. 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.
29. 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 in Resesearch and Higher Education, 1991.
30. Andersen H. Tidsskriftpublicering og forskningsevaluering. *Biblioteksarbejde* 1996;17:19-26.
31. Korevaar JC, Moed HF. Validation of bibliometric indicators in the field of mathematics. *Scientometrics* 1996;37:117-30.
32. Bauin S, Rothman H. "Impact" of journals as proxies for citation counts. In: Weingart P, Sehringer R, Winterhager M, eds. *Representations of science and technology*. Leiden: DSWO Press, 1992;225-39.
33. Narin F, Hamilton KS. Bibliometric performance measures. *Scientometrics* 1996;6:293-310.
34. Artus HM. Science indicators derived from databases. *Scientometrics* 1996;37:297-311.
35. Møller AP. National citations. *Nature* 1990;348:480.
36. Hamilton DP. Research papers: who's uncited now. *Science* 1991;251:25.
37. Hamilton DP. Publishing by—and for?—the numbers. *Science* 1990;250:1331-2. [[Medline](#)]
38. Marton J. Obsolescence or immediacy? Evidence supporting Price's hypothesis. *Scientometrics* 1985;7:145-153.
39. Hargens LL, Felmlee DH. Structural determinants of stratification in science. *Am Sociol Rev* 1984;49:685-97.
40. Vinkler P. Relationships between the rate of scientific development and citations. The chance for a citedness model. *Scientometrics* 1996;35:375-86.
41. Gomperts MC. The law of constant citation for scientific literature. *J Documentation* 1968;24:113-7.
42. Seglen PO. Bruk av siteringsanalyse og andre bibliometriske metoder i evaluering av forskningsaktivitet. *Tidsskr Nor Laegeforen* 1989;31:3229-34.
43. Seglen PO. Evaluering av forskningskvalitet ved hjelp av siteringsanalyse og andre bibliometriske metoder. *Nordisk Medicin* 1989;104:331-5.

44. Narin F, Pinski G, Gee HH. Structure of the biomedical literature. *J Am Soc Inform Sci* 1976;27:25-45.
45. Folly G, Hajtman B, Nagy JI, Ruff I. Some methodological problems in ranking scientists by citation analysis. *Scientometrics* 1981;3:135-47.
46. Schubert A, Braun T. Cross-field normalization of scientometric indicators. *Scientometrics* 1996;36:311-24.
47. Lewison G, Anderson J, Jack J. Assessing track records. *Nature* 1995;377:671.
48. Vinkler P. Model for quantitative selection of relative scientometric impact indicators. *Scientometrics* 1996;36:223-36.
49. **Seglen** PO. How representative is the journal impact factor? *Research Evaluation* 1992;2:143-9.
50. Martin BR. The use of multiple indicators in the assessment of basic research. *Scientometrics* 1996;36:343-62.
51. MacRoberts MH, MacRoberts BR. Problems of citation analysis: a critical review. *J Am Soc Information Sci* 1989;40:342-9.
52. **Seglen** PO. Siteringer og tidsskrift-impakt som kvalitetsmål for forskning. *Klinisk Kemi i Norden* 1995;7:59-63.
53. Brenner S. Cited [incorrectly] in: Strata P. Citation analysis. *Nature* 1995;375:624. [[Medline](#)]

(Accepted 9 January 1997)