

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

Per O. Seglen
Institute for Studies in Research and Higher Education (NIFU)
Hegdehaugsveien 31
N-0352 Oslo, Norway

9 January 1997

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, making peer review as much of a lottery as of a rational process.

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. 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.

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. 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. Resource allocation based on impact factors has also been reported from Canada and Hungary 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, 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.

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

1 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). 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.

The uneven contribution of the various articles to the journal impact can easily be demonstrated. 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.

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. 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, is also recommended. Furthermore, because citation rate is roughly proportional to the length of the article, 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*.

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). Within the arts and humanities, references to articles are hardly used at all, leaving these research fields (and others) virtually uncited, a matter

of considerable consternation among science administrators unfamiliar with citation kinetics.

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.

2 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. 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. 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 criterion 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. 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."¹

3 Acknowledgements

Funding: None. Conflict of interest: None.

¹Brenner S. Cited [incorrectly] in: Strata P. Citation analysis. Nature 1995;375:624