

Citation data: the wrong impact?

Charles Jennings, Ph.D.

Editor, *Nature Neuroscience*, 345 Park Avenue South, New York, NY 10010,
USA. TEL: +1 212 726 9310, FAX: +1 212 696 0978

Every September, a ripple of excitement passes through the scientific community as the Institute of Scientific Information (ISI) publishes its latest set of impact factors, in which some six thousand journals are ranked according to the number of citations they received in the previous year. The release of these results triggers elation or gloom in editorial offices around the world, but for many scientists it is no more than light entertainment, the scientific equivalent of tabloid gossip. For others, however, it represents something more serious, because their career prospects are increasingly affected by the impact factors of the journals in which they publish. Although bibliometric data undoubtedly have the potential to reveal significant insights into the quality of scientific work, they are also susceptible to abuse. It is therefore worth examining in some detail how they are derived and how they are now being applied.

ISI is a commercial company, based in Philadelphia, which publishes Science Citation Index and Current Contents in addition to Journal Citation Reports, where impact factors are reported. The impact factor for a given year—say, 1997—is calculated as follows: ISI counts the number of citations made in 1997 to papers published in the previous two years, 1995 and 1996, and divides by the number of articles published in that two-year period.

The number thus derived is biased in several ways that are not always fully appreciated (1). Most obviously, by the time the impact factors appear, the papers to which they refer are already two to three years old, so any recent changes in a journal's editorial policies will not be reflected in its impact factor. (This is partly avoided by looking at the 'immediacy index', which is the average number of citations in—say—1997 to papers published in 1997, but this number is no more than a snapshot, and papers appearing early in the year will be cited more than those appearing later.)

According to ISI, the great majority of citations are almost invariably to a small fraction of the total articles, and so the impact factor, which is the mean citation rate, is a poor measure of the typical paper in that journal; this is true of high- and low-impact journals alike. In fact, most papers are cited at much lower rates than the journal's impact factor would suggest. Giving a disproportionate weight to the most highly cited papers is not necessarily a disadvantage if the aim is to measure the usefulness of a journal to its field—assuming that the more highly cited papers are likely to be the more significant ones—but it does mean that little can be inferred about the likely citation of an individual paper from simply knowing the impact factor of the journal in which it appeared.

Most importantly, however, different fields have different intrinsic citation rates, and the impact factor for a given journal reflects the topics it covers. Molecular biology, for instance, tends to generate a large number of citations per paper, mainly because there are so many molecular biology papers that can cite each other. There are fewer ecology papers published, so they each receive fewer citations. Neuroscience is somewhere in the middle, but it seems likely that within the field, the most highly cited papers tend to be on molecular and cellular rather than systems or cognitive neuroscience. Although it might be argued that fields become large because they are important, there is a danger (at least when comparing across fields) that impact factors will tend to reward followers rather than leaders, and that papers representing pioneering work in new areas will receive fewer citations than those from fields that are already crowded.

Although these limitations are (or should be) well known, journals routinely use impact factors to evaluate their editorial performance, to attract the best papers and to market themselves to potential subscribers. *Nature Neuroscience* is of course still too young to have an impact factor, but our colleagues on the other *Nature* journals, like publishers elsewhere, do not hesitate to draw attention to numbers that they believe reflect well on their respective titles. There is nothing wrong with a little friendly competition, but it should not be taken too seriously. If readers pay too much attention to the numbers, they may create an incentive for editors to inflate them by artificial means; David Pendlebury, an analyst at ISI, says he has received a number of calls from editors seeking to understand the impact factor calculation so that they can manipulate it to their journal's advantage. Needless to say, ISI does not condone this practice and recommends instead publishing better papers, but for those who may be interested, here are some strategies: publish more reviews, which receive higher citations than original research papers; alter subject coverage in favor of fields with high intrinsic citation rates, such as molecular biology; eliminate topics and sections that generate few citations; and publish controversial editorials. The last method works because when the impact factor is calculated, the numerator is the total number of citations to any item in the journal, whereas the denominator is the number of articles only, and editorials and letters are not normally counted.

Despite these problems, most scientists would agree that journals do vary in quality and that, at least within a given field, there is some correlation between quality and impact factor. Moreover, many studies have shown correlations between citation frequency and significance of individual papers as judged by other means; one, coauthored by Eugene Garfield, the founder of ISI, even reports that publication of highly cited papers is a good predictor of future Nobel prizewinners (2). Why then does it matter that people have become so obsessed with impact factors?

The main problem is that impact factors are being increasingly used for a purpose for which they were never intended, namely to evaluate individual applicants for jobs or funding. The ISI has never advocated this use; they emphasize that there is

no substitute for informed peer review, and that bibliometric data may supplement but should never replace such review. Unfortunately this message is not always heard, and a disturbing trend has emerged over the last few years, in which committees charged with making hiring and funding decisions have come to rely increasingly on impact factors rather than on more direct methods when evaluating the quality of their candidates' research programs.

The trend appears to be particularly widespread in Europe. In Italy, for instance, the Italian Association for Cancer Research (AIRC) requires grant applicants to complete worksheets, reminiscent of income tax returns, in which they must calculate the sum of the impact factors for each journal in which they have published for the last five years, then calculate their weighted average impact factor, then repeat the process for special categories such as reviews and first/last authorship publications. According to Antonio Malgaroli, a neuroscientist at the University of Milan, such calculations are widely used in Italy for both hiring and funding decisions, with little attempt to consider the biases inherent in impact factor measurements.

Similar practices are used in other countries of Europe, and also in Japan. Masao Ito, director of the RIKEN Brain Sciences Institute near Tokyo, agrees that there is a serious problem; appointment committees at Japanese universities are often heavily influenced by journal impact factors, and committee members tend to place excessive weight on numbers whose meaning they do not properly understand. The same is true to some extent in the US, according to Zach Hall, vice-chancellor for research at UCSF and former director of the National Institute of Neurological Disorders and Stroke. Hall believes, however, that the practice is less widespread in the US than in some other countries, and in particular that it is relatively rare at the leading universities and research institutes. Nevertheless, Janet Robertson, editor of Journal Citation Reports, says she receives calls almost every week from scientists both in the US and elsewhere, complaining that they have been victims of misinterpreted ISI data.

The motive in all these cases seems to be a desire to make the selection process both efficient and objective, but unfortunately neither outcome is likely. In principle, committees might use citations to individual papers rather than to the journals in which they appeared, but because the relevant papers are often recent, these numbers may not exist, leaving the impact factor as the most readily available surrogate. Numerical methods are particularly tempting for large departments and interdepartmental groups, where hiring committees may have neither the time nor the expertise to evaluate candidates in all the fields for which they are responsible. Faced with an incessant flow of applications, a simple algorithm for ranking candidates has an obvious appeal. Yet, as Richard Frackowiak, dean of the Institute of Neurology at University College London, puts it, although increased objectivity is a reasonable goal, the available tools are still "extremely crude", and relying on them in hiring or funding decisions is "iniquitous and frankly counter-productive". Hall agrees, and considers most numerical methods of evaluation as little more than "excuses for not thinking".

The result of all this numerology has been an increasing obsession among researchers, particularly younger scientists who have not yet established their reputations, to boost their numbers by whatever means possible. Ito, for instance, recounts the case of a young colleague who chose to submit to one journal rather than another based on a difference of 0.2 between their respective impact factors. *Nature Neuroscience* has received at least one inquiry from a prospective author, wondering whether to submit his paper to us and wanting to know what our impact factor would be. These may be extreme examples, but they reflect a more general trend toward placing an increased weight on impact factors relative to more appropriate criteria such as editorial policies or target readership. The situation has reached the point where many scientists (and

most editors) can quote the impact factors of their favorite journals to three significant figures, and the word 'impact' has become a virtual synonym for scientific quality.

There are signs that the situation may be changing, at least in some quarters. Impact factors have been widely used in Germany in the past, but earlier this year, the Deutsche Forschungsgemeinschaft (DFG, Germany's main government research agency) issued new guidelines to universities, requiring that they abandon the practice of evaluating candidates based on impact factors, and instead examine the candidates' top five publications directly. According to Wolf Singer, a neuroscientist at the Max-Planck Institute (MPI) in Frankfurt and a member of the committee that prepared the guidelines, this reflects a broader cultural change in German science. Several high-profile fraud cases led to the conclusion that one motive for scientific misconduct is the pressure to boost bibliometric scores by publishing as many papers as possible in high-impact-factor journals. As a result, both the DFG and the MPI are now looking for ways to reform the research climate in ways that will nurture quality rather than sheer quantity. Similarly, according to Frackowiak, the Wellcome Trust (which funds his work) is exploring ways to use bibliographic methods more intelligently. For instance, applicants for Wellcome fellowships are asked to identify their leading peers in the same discipline, and the citation rates of these people's papers (rather than the journals in which they appeared) form a baseline against which the applicant's publication record can be compared.

On the other hand, governments around the world are increasingly demanding objective indicators of research performance, in the name of increased efficiency. In Britain, for instance, every four years the government conducts a Research Assessment Exercise (RAE), in which research units are evaluated and given a numerical score that determines their future funding. As part of the assessment, individuals must submit four recent publications, and although the RAE does not officially use impact factors in its evaluations, there is a widespread perception that they weigh heavily in many panels' recommendations. In the US, the Government Performance and Results Act requires all federally funded agencies to use performance measures to evaluate themselves, beginning this year. How this should be applied to agencies that fund basic research is not clear, but one obvious possibility is to use bibliometric data; indeed, ISI staff have already given presentations to the National Research Council committee charged with solving this problem.

It may be appropriate to end with a conflict of interest statement. Although *Nature Neuroscience* is now indexed by Current Contents and hopes to be listed on Medline by early 1999, it has no impact factor at present and does not expect to have one until 2001. Whether this constitutes a conflict is for readers to decide; we hope, however, that by then, the uncritical obsession with impact factors that has become so pervasive over the last few years will have been replaced by a more sophisticated approach to the analysis of what is undoubtedly an enormously valuable resource for understanding how science is practiced.

REFERENCES

1. See for instance Seglen, P. O. *Br. Med. J.* 314, 498-502 (1997).
2. Garfield, E. & Welljams-Dorof, A. *Theor. Med.* 13, 117-135 (1992). MEDLINE