Impact factors—a critique

The so-called impact factor (IF) started as a rather esoteric scientometric tool, itself derived from the science citation index (SCI), which originated with Eugene Garfield in the 1950s. The IF of a journal for a particular year is defined as the quotient of the number of citations received in that year by papers published in that journal in the two preceding years and the number of "citable" papers published in those two years. Had the IF remained, as might have been reasonable to expect, a specialist statistic mainly of interest to information professionals as represented by librarians and others there would be no need to expend energy on criticizing it. Unfortunately its use—or rather misuse—has vastly expanded in recent years, which makes criticism very necessary.

Following the failure to secure public funding for the establishment of the SCI, it was launched as a commercial venture in 1961 as the Institute for Scientific Information (ISI). In 1992 the ISI was sold to the Thomson Corporation, well known as the publisher of newspapers such as the *Sunday Times*. Thomson Corporation merged with Reuters in 2008 to form Thomson Reuters, which now claims to be "the world's leading source of intelligent information *(sic)* for businesses and professionals".

The IF has a number of technical defects. Some of the more obvious ones are: (i) the IF of journals covering a broad area of science with a rapidly growing but ephemeral literature that tends to cite many articles will inevitably be higher than the IF of more specialist journals whose articles may reach peak citation many years after their publication; (ii) review articles tend to be much more heavily cited than primary research articles and hence journals carrying some review articles, and especially journals exclusively devoted to them, will have high IF—it seems absurd to include them both in the same statistic; (iii) as a cursory check will quickly verify, the distribution of citations received by individual articles in a journal is so broad as to make the mean almost meaningless. Technical reports describing a new methodology and "data-rich" papers such as the articles reporting the human genome sequence tend to be very heavily cited. Indeed, it has been admitted by Thomson that the median is very different from the mean [1] without, however, this admission having been followed by revealing the actual distributions. In one field at least, high energy physics, analysis of citations has revealed a power law distribution with at least two exponents [2], hence an ill-defined variance.

Some of the less obvious technical defects are: (i) inconsistency with respect to papers being cited and receiving citations. The citable papers in the denominator contain only papers categorized as primary research articles or review articles; "news and views" articles are excluded [1]. On the other hand, the numerator contains all citations, even when they are made by such news and views articles. Furthermore, categorization of article type is made by Thomson Reuters staff according to a procedure not publicly available and that moreover appears to be subject to commercial negotiation [1]; (ii) only journals included in the Thomson Reuters database are considered. The database is dominated by North American publications [3]. This not only means that many important journals not in the database do not even receive impact factors, but also that citations in journals outside the database, and in other sources such as books, are not counted at all. Furthermore, the database changes from year to year and the criterion or criteria for including a journal are not publicly available; all one knows is that inclusion depends on some subjective factors enshrined within the company's policy, such as a desire to promote a particular field, much as a newspaper proprietor typically imposes a certain party line on editorial policy.

Because of the extreme tediousness of counting citations, it has been practically impossible to check the accuracy of the published citation figures. In the very rare cases where this has been done, alarming discrepancies have been uncovered [1]. Perhaps the recent general availability of powerful data searching services such as Google Scholar, which make it practicable for individual scientometricians and even scientists interested in the matter to develop and calculate their own indices, will enable more reliable citation metrics based on open and transparent procedures to be developed. It should, however, be appreciated that this is a tricky matter indeed. Some of the alternatives that have already been proposed, such as the h- and m-indices (Hirsch), the g-index (Egghe) and the w-index (Woeginger) have also been subject to severe criticism [4].

There are also defects of a semantic nature. Garfield himself wrote "the citation index...may help a historian to measure the influence of the article—that is, its 'impact factor'." [5]. This vagueness regarding the link between citations (i.e., the IF) and actual impact has not been made more precise by subsequent pronouncements (see [4]). No attempt appears to be made to distinguish between positive and negative influence. Another defect of a semantic nature is the astonishing fact that citations to retracted articles are counted [1]! This is all the more regrettable because the scientist but not the bureaucrat knows that retraction is almost always a consequence of the perpetration of fraud having been uncovered.

It would have been more accurate to have called the IF "popularity factor". If one accepts the so-called IF as a

Journal of Biological Physics and Chemistry **9** (2009) 139–140
Received 31 December 2009

measure of quality, one must also accept that, in the UK, the *Sun* is a higher-quality daily newspaper than the *Financial Times*, or (in Switzerland) that *Blick* is of a higher quality than the *Neue Zürcher Zeitung*.

It should also be recalled that different fields of science have very different citation rates. The mean IF of mathematics journals in the Thomson Reuters database is less than one and of physics journals is less than two; chemistry fares somewhat better with a mean IF approaching three and biology almost attains a value of five. This of course makes comparisons between different fields meaningless. Popular magazines like *Nature* and *Science* that combine the functions of a weekly news bulletin and a scientific journal covering many fields have much higher IF (around 30), presumably because they are widely read and the scholarly practice of only citing what one has read still largely holds sway.

However nebulous the meaning of impact in the term "impact factor", undoubtedly the IF has acquired a great impact, especially in connexion with the bureaucratic controls and financial restrictions imposed on scientists. In these contexts the IF is recklessly used by officials not in a position to appreciate those defects that make it quite unsuitable for the purpose of assessing research quality. As the IMU report eloquently states, "Much of modern bibliometrics seems to rely on experience and intuition about the interpretation and validity of citation statistics." [4].

One can indeed admire Garfield's entrepreneurial success in having turned an obscure, specialist and flawed metric into a highly profitable business. Given that Thomson Reuters is a commercial organization with the goal of maximizing profits it should not be expected that it should follow standards of scholarly rigour. What is surprising is that the majority of scientists themselves appear to have acquiesced in the general misuse of the IF [6]. For the reasons discussed above, most of which have already been amply pointed out in the literature, the IF is incompatible with the scientific method. Acceptance of the IF by a scientist must therefore cast doubt on his or her professional integrity. Long ago the then Director of Research of the Institute for Scientific Information pointed out that "the SCI would work perfectly if every author meticulously cited only the earlier work related to his theme; if it covered every scientific journal published anywhere in the world; and if it were free from economic constraints." [7]. It was already realized then that these conditions were not fulfilled—hence the SCI on which the IF is based is flawed even at the most basic level of the scientist choosing which papers to cite.

J.J. Ramsden

References

1. Rossner, M., Van Epps, H. and Hill, E. Show me the data. *J. Cell Biol.* **179** (2007) 1091–1092.
2. Lehmann, S., Lautrup, B. and Jackson, A.D. Citation networks in high energy physics. *Phys. Rev. E* **68** (2003) 026113.
3. Seglen, P.O. Why the impact factor of journals should not be used for evaluating research. *Br. Med. J.* **314** (1997) 497.
4. Adler, R., Ewing, J. and Taylor, P. *Citation Statistics*. Washington: International Mathematical Union (IMU) (Joint Committee on Quantitative Assessment of Research) (2008).
5. Garfield, E. Citation indexes for science: A new dimension in documentation through association of ideas. *Science* **122** (1955) 108–111.
6. Not only merely passively, but sometimes to the extent of going to great lengths to manipulate citations—see Z. Corbyn, *Times Higher Education*, 7 February 2008, p. 6.
7. Cawkell, A.E. Science perceived through the Science Citation Index. *Endeavour (new ser.)* **1** (1977) 57–62.

Impact factors—a critique

Ramsden, Jeremy J.

2009