

## Impact factors: use and abuse

MAYUR AMIN<sup>1</sup>, MICHAEL A. MABE

*Elsevier Science, The Boulevard, Langford Lane, Kidlington, Oxford, United Kingdom*

### What is an Impact Factor?

The **impact factor** is only one of three standardized measures created by the Institute of Scientific Information (ISI), which can be used to measure the way a journal receives citations to its articles over time. The build-up of citations tends to follow a curve like that of Figure 1. Citations to articles published in a given year rise sharply to a peak between two and six years after publication. From this peak citations decline exponentially. The citation curve of any journal can be described by the relative size of the curve (in terms of area under the line), the extent to which the peak of the curve is close to the origin and the rate of decline of the curve. These characteristics form the basis of the ISI indicators **impact factor**, **immediacy index** and **cited half-life**.

The **impact factor** is a measure of the relative size of the citation curve in years 2 and 3. It is calculated by dividing the number of current citations a journal receives to articles published in the two previous years by the number of articles published in those same years. So, for example, the 1999 impact factor is the citations in 1999 to articles published in 1997 and 1998 divided by the number of articles published in 1997 and 1998. The number that results can be thought of as the average number of citations the average article receives per annum in the two years after the publication year.

The **immediacy index** gives a measure of the skewness of the curve, that is, the extent to which the peak of the curve lies near the origin of the graph. It is calculated by dividing the citations a journal receives in the current year by the number of articles it publishes in that year, i.e., the 1999 immediacy index is the average number of citations in 1999 to articles published in 1999. The number that results can be thought of as the initial gradient of the citation curve, a measure of how quickly items in that journal get cited upon publication.

The **cited half-life** is a measure of the rate of decline of the citation curve. It is the number of years that the number of current citations takes to decline to 50% of its initial value; the cited half-life is 6 years in the example given in (Figure 1). It is a measure of how long articles in a journal continue to be cited after publication.

---

<sup>1</sup> m.amin@elsevier.co.uk

**Nota de la Redacción:** Mayur Amin y Michael A. Mabe son Director y Director Asociado de Elsevier Science (*publishing strategy, research interests and expertise*). Actualmente dirigen las relaciones académicas del *Elsevier Science Group* con responsabilidad en el estudio del comportamiento de los investigadores y el sistema de publicaciones científicas. Este artículo fue publicado en *Perspectives in Publishing*, N° 1, octubre 2000 y se encuentra en <http://www.elsevier.com/homepage/about/ita/editors/perspectives1.pdf>.

Con autorización de los autores, *Medicina (Buenos Aires)* lo reproduce por considerarlo de interés para sus lectores. Las ilustraciones han sido redibujadas para adaptarlas al formato de la revista.

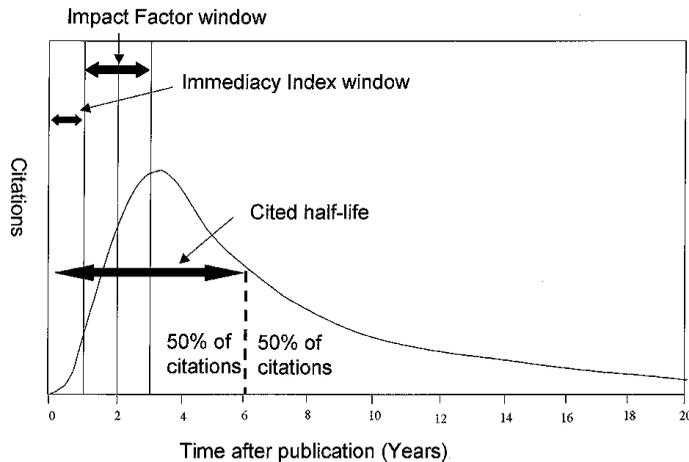


Fig. 1.— Generalized Citation Curve.

### How variable is the impact factor?

Of the three measures described above, the impact factor is the most commonly used and also most misunderstood. This pamphlet addresses some of the factors that affect the impact factor. The value of the impact factor is affected by sociological and statistical factors. Sociological factors include the subject area of the journal, the type of journal (letters, full papers, reviews), and the average number of authors per paper (which is related to subject area). Statistical factors include the size of the journal and the size of the citation measurement window.

#### Subjectivity

Figure 2a shows how the absolute value of the mean impact factor exhibits significant variation according to subject field. In general, fundamental and pure subject areas have higher average impact factors than specialized or applied ones. The variation is so significant that the top journal in one field may have an impact factor lower than the bottom journal in another area.

Closely connected to subject area variation is the phenomenon of multiple authorship. The average number of collaborators in a paper varies according to subject area, from social sciences (with about two authors per paper) to fundamental life sciences (where there are over four). Not unsurprisingly, given the tendency of authors to refer to their own work, there is a strong and significant correlation between the average number of authors per paper and the average impact factor for a subject area (Figure 2b). So comparisons of impact factors should only be made for journals in the same subject area.

#### Article and Journal Type

Even within the same subject area there will be significant **variation according to the journal type or article type**. This is illustrated in Figure 3.

A short or rapid publication journal (often called a “Letters” journal, publishing short papers, not to be confused with letters to the editor) will have greater *immediacy* but a *lower cited half-life* (that is, the peak of the citation curve will be closer to the origin and the curve will decline rapidly after the peak). As a

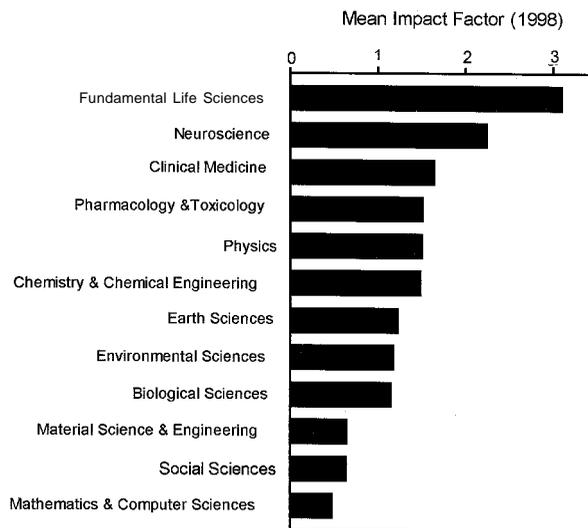


Fig. 2a.- Subject Variation in Impact Factors.

consequence, a large proportion of the citations it receives will tend to fall within the two-year window of the impact factor. By contrast, the full paper journal will have a citation peak around three years after publication, and therefore a lower *immediacy* than the rapid or short paper journal. It will also have a gentle decline after its peak, and consequently a larger *cited half-life*. The proportion of citations that fall within the two-year window will be smaller as a result of the different curve shape, and the *impact factor*

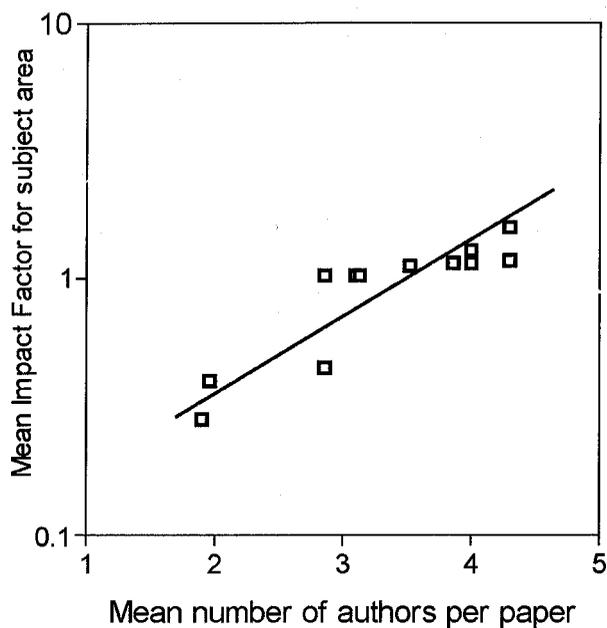


Fig. 2b.- Impact Factors and number of Authors per paper.

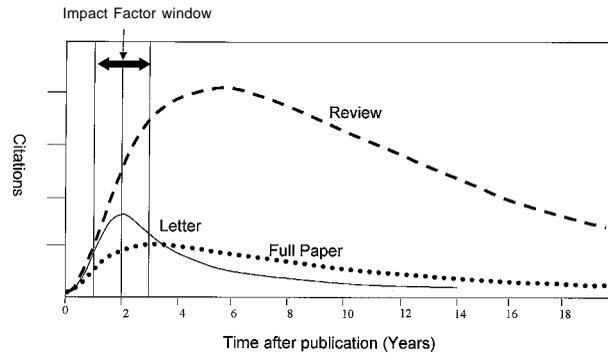


Fig. 3.– Impact Factors and Journal Type.

of such journal will tend to be smaller than its rapid or short paper relative. In the case of a review journal, the immediacy index relative to other measures is very low, citations slowly *rising* to peak many years after publication. The *cited half-life* is also correspondingly long, as the citations decline equally slowly after the peak. The proportion of the curve that sits within the two-year impact factor window is also relatively small, but because the absolute number of citations to reviews is usually very high, even this proportion results in higher average impact factors for review journals over all other journal types. So, given that the impact factor measures differing proportions of citations for different article types, care should be taken when comparing different journal types or journals with different mixes of article types.

#### *Size Matters*

As the impact factor is an average value, it also shows **variation due to statistical effects**. These relate to the number of items being averaged, that is the size of the journal in terms of articles published per annum, or the size of the measurement window (which for the standard or JCR impact factor is two years; in fact, a one year citing window and a two year cited window).

The effects of journal size can be seen quite clearly in Figure 4a. If a large number of journals (4000, arranged in quartiles based on size of journal) are examined and the mean variation in impact factor from one year to the next is plotted against size of the journal, there is a clear correlation between the extent of the impact factor fluctuation and the size of the journal. This means that when impact factors are compared between years it is important to consider the size of the journal under consideration. Small titles (less than 35 papers per annum) on average vary in impact factor by more than  $\pm 40\%$  from one year to the next. Even larger titles are not immune, with a fluctuation of  $\pm 15\%$  for journals publishing more than 150 articles per annum. Does this mean that smaller journals on average are more inconsistent in their standards? The answer is “no”. Any journal in effect takes a small, (biased sample in that subjective selection criteria are involved) of articles from a finite but large pool of articles. The impact factor and any fluctuation in it from one year to the next can be considered a result of that biased sample. However, what fluctuation in impact factor would one see if random (or unbiased) samples of articles were taken? This sampling error is estimated and shown in Figure 4b. Here the observed fluctuation lines represent the actual mean change in impact factor seen in a study of 4000 journals ordered into groups according to size. The shaded area approximates the fluctuation in impact factor that would result from random samples of articles, i.e. the difference between the impact factor values in separate random

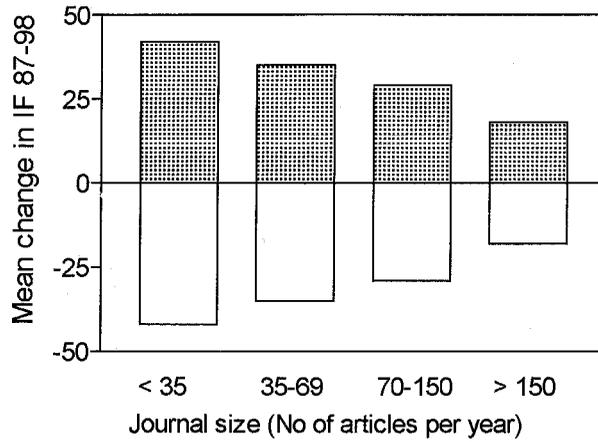


Fig. 4a.- Impact Factor Fluctuations vs Journal Size. Based on sample of 4000 journals.

samples of the same size. Therefore, a change in impact factor for any journal of a given size is no different to the average journal if within the observed fluctuation, and could happen randomly if the fluctuation is within the shaded areas. For example, the impact factor of a journal of 140 articles would need to change by more than  $\pm 22\%$  to be significant. By the same token, differences in impact factor between two journals of the same size and in the same subject area should be viewed with this fluctuation in mind. An impact factor of 1.50 for a journal publishing 140 articles is not significantly different to another journal of the same size with an impact factor of 1.24. While the exact numbers given here are based on an approximate model, care should be exercised to avoid inferring too much from small changes or differences in impact factors.

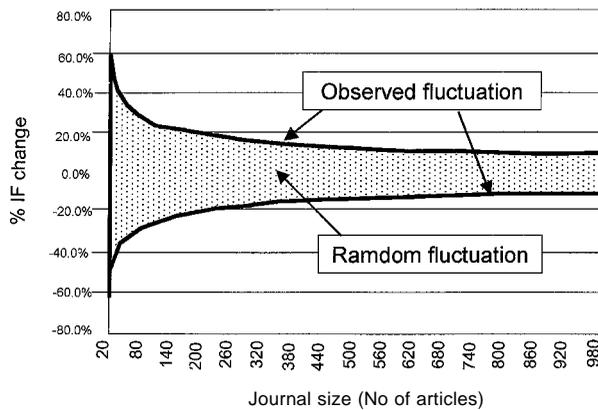


Fig. 4b.- Impact Factor Fluctuation vs Journal Size.

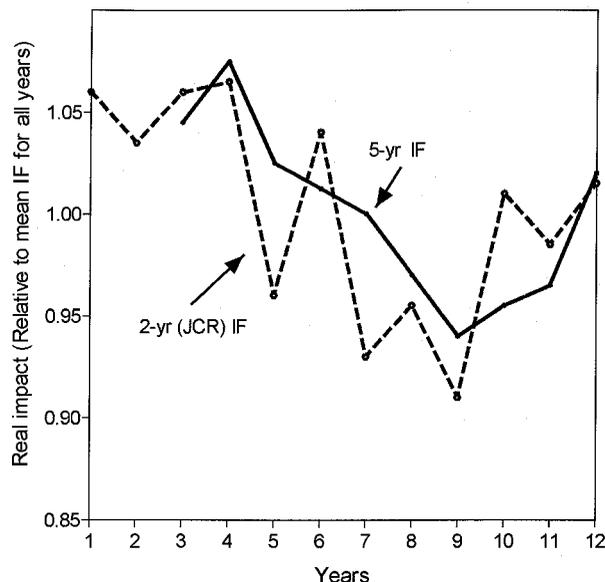


Fig. 5.— Impact Factor Measurement Window Fluctuations. Average 200+ Chemistry Journals.

#### *Looking through different windows*

Expanding the size of the measurement window from the two years of the standard JCR impact factor can iron out some of the statistical variations. The effects of doing this are illustrated in Figure 5. Here the average two and five-year impact factors for around 200 chemistry journals have been plotted against time. The two-year impact factors show considerable variability, jumping up and down in value each year. The five-year measures, however, while still showing changes over time, present a much smoother curve. A measure that is often used in evaluating a journal or laying claims to its importance is the rank it has by impact factor amongst other journals in *its* subject area. However, dramatic changes in rank can occur simply by changing the time frame of measurement. For example, of 30 chemistry journals examined, 24 changed in rank by up to 1 to 11 positions when changing from a two-year to a five-year impact measurement.

#### **Why does Impact Variability Matter?**

The previous section has shown how easily impact factors are affected by a host of conditions which do not directly impinge upon their principal use, a measure of the impact of publishing in a particular journal, but which sensibly limit how they can be applied. It is clearly inappropriate to use them to rank all types of journals in all subject areas, and even comparing the same type of journal in the same subject category (e.g., all letters journals in condensed matter physics) still leaves the results subject to the statistical factors. Year-to-year variability will still be very high for small titles and emphasized by the “official” JCR impact factor.

Given these considerations, even the journal rankings by subject area produced by ISI should be treated with care. As a rule of thumb journals with impact factors that differ by less than 25% belong together in the same rank. The use of the absolute value of an impact factor to measure quality should be strongly avoided, not only because of the variability discussed above, but also because the long-term average trends in different fields vary. In Figure 5 it would be foolish to suggest that chemistry research being done in year 9 was worse than any other year. Equally foolhardy is to penalize authors for publishing in journals with impact factors less than a certain fixed value, say, 2.0, given that for the average sized journal this value could vary between 1.5 and 2.25, without being significant.

The use of journal impact factors for evaluating individual scientists is even more dubious, given the statistical and sociological variability in journal impact factors.

**The Numerator/Denominator Problem**

The formulation of the impact factor also leads to some unfortunate calculation effects. As it is a ratio, clear and unambiguous definitions of the items counted for the top and bottom of the fraction are essential. As is illustrated in Figure 6, the published impact factors are the ratio between the number of citations to all parts of the journal and the number of papers. But what exactly counts as a paper? Do letters to the editor count? Or editorials? Or short abstract papers? ISI classify papers into a number of different types (articles, reviews, proceedings papers, editorials, letters to editor, news items, etc). Only those classified as “articles” or “reviews” and “proceedings papers” are counted in the denominator for the impact factor calculation, whereas citations to all papers (including editorials, news items, letters to the editor, etc) are counted for the numerator. This can lead to an exaggerated impact factor (average cites per paper) for some journals compared to others. While there are valid, practical reasons for this approach, discrepancies can occur where, as a result, some journals are more favored than others.

If a very strict definition of impact factor is used, where citations to only selected article types are divided by the number of those selected article types, considerable differences can emerge from the published impact factors. Figure 6 shows this effect for a number of journals in medicine, physics and neuroscience. About 40% of medicine journals have published impact factors that are 10% greater than

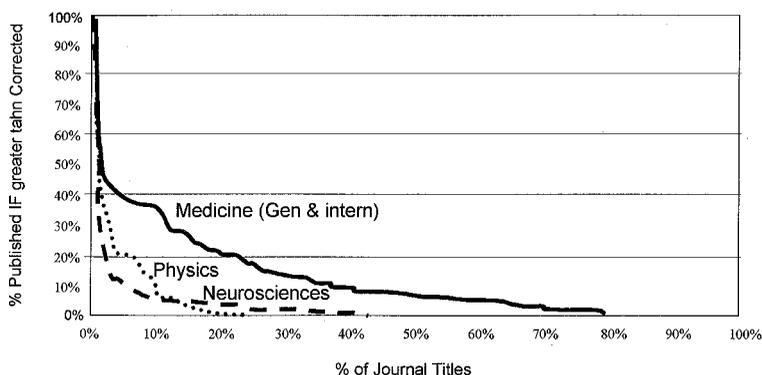


Fig. 6.– Published versus Corrected Impact Factors.  
**Published IF:** [Citations of All Article types]/ [N° of articles (selected articles types only)]. **Corrected IF:** [Citations to selected Article types]/[N° of articles (selected articles types only)].

the strictly calculated ones, and 5% of these journals have differences as great as 40% or more. In physics, about 7% of journals have published values that are 20 % more than those strictly calculated; while in neuroscience very few of the journals differ significantly.

The problem is caused by the difficulties in classifying papers into article types and deciding which article types to include in the impact factor calculation, particularly when publishing practices can vary across disciplines. This is particularly problematic in medicine where letters to the editors (which are not "Letters papers" in the sense used elsewhere in science) or editorials or news items can collect significant numbers of citations. This so-called numerator/denominator problem is yet another example of why considerable care needs to be taken when using impact factors.

## Conclusions

This pamphlet has shown that impact factors are only one of a number of measures for describing the "impact" that particular journals can have in the research literature. The value of the impact factor is affected by the subject area, type and size of a journal, and the "window of measurement" used. As statistical measures they fluctuate from year to year, so that great care needs to be taken in interpreting whether a journal has really "dropped (or risen)" in quality from changes in its impact factor. Use of the absolute values of impact factors, outside of the context of other journals within the same subject area, is virtually meaningless; journals ranked top in one field may be bottom in another. Extending the use of the journal impact factor from the journal to the authors of papers in the journal is highly suspect; the error margins can become so high as to make any value meaningless. Professional journal types such as those in medicine frequently contain many more types of source item than the standard research journal. Errors can arise in ensuring the right types of article are counted in calculating the impact factor.

Citation measures, facilitated by the richness of ISI's citation databases, can provide very useful insights into scholarly research and its communication. Impact factors, as one citation measure, are useful in establishing the influence journals have within the literature of a discipline. Nevertheless, they are not a direct measure of quality and must be used with considerable care.

- - - -

*In research, the usefulness of error is that it leads to more research, and this is what the word tells us ... In order to get anything right, we are obliged first to get a great many things wrong.*

En la investigación, lo útil del error es que conduce a más investigación, y eso es lo que la misma palabra nos dice ... Para encontrar la verdadera solución, primero tenemos que equivocarnos muchas veces.

Lewis Thomas

*The Youngest Science, Notes of a Medicine Watcher*, New York: Viking, 1983, p 82