

Beware the impact factor

The journal impact factor is a good predictor of the quality of journals as measured by citations to primary research articles. It is, however, a poor indicator of citations to specific papers or of the future performance of individual researchers.

When it comes to scientific measures, the journal impact factor wins both in terms of broadest use and as the most loathed metric^{1,2}. Indeed, the fact that it is simple to understand — it is roughly the average number of citations that primary research papers published in two consecutive years gather in the following year — makes it all too easy to point out its shortcomings: the metric also includes citations to non-primary content (such as reviews and news articles); for many fields, citations accumulate slowly and thus the two-year time window seems too short; and the average number of citations per paper can be skewed by a few highly cited ones³, of which high-impact journals have a big share. Many feel that these limitations favour highly selective and multidisciplinary journals disproportionately.

Here we argue that these limitations are irrelevant. Figure 1 shows that, for a sample of 100 journals across the spectrum of science and engineering, the 2011 impact factor correlates well with the five-year median of citations to primary research papers published in 2008–2012. It is important to stress that the values for the median — which corresponds to the minimum number of citations received by half of the papers, and thus is robust to outliers and variations in the shape of the distribution — do not include citations to non-primary content and have a time window of five years.

That citation averages (such as the impact factor) and medians correlate is not surprising if one considers that the shape of the citation distributions may be comparable across journals, as the similarities between the usual two-year and the less-known five-year impact factors suggests⁴. What is perhaps unexpected is the robustness of the impact factor as a predictive metric: citations to non-primary content and the apparently too short two-year time window have little effect on the overall correlation. Still, it is interesting to note that the largest deviations from the linear fit in Fig. 1 correspond to medical journals, some of which produce a disproportionate amount of non-primary content (such as *The Lancet* and *The Journal of the American Medical Association*) or to journals that have significantly altered the yearly amount of primary content during the five-year time frame for which the median is calculated.

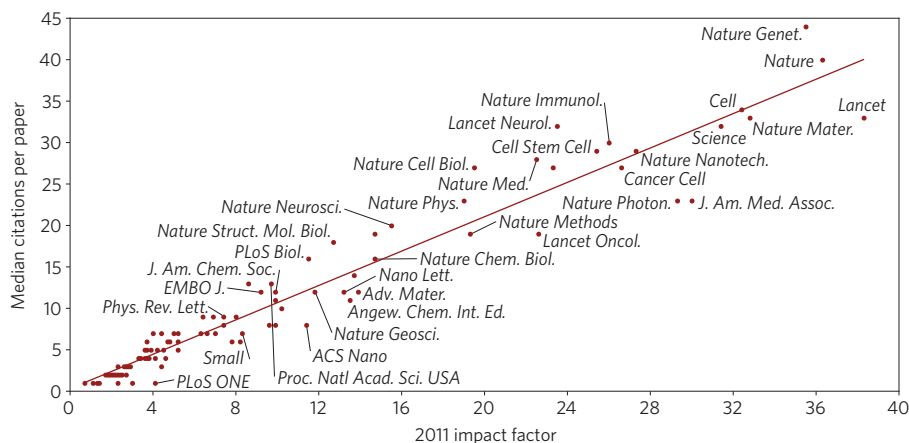


Figure 1 | A journal's impact factor is a good predictor of its five-year median of citations to primary research articles. The data and linear fit ($r^2 = 0.94$) correspond to a sample of 100 journals launched before 2008. The five-year median values are of citations (as of 5 January 2013) to research papers (that is, excluding reviews, news, editorial material and other non-primary research articles) published in 2008–2012. The specific median values and slope of the linear fit (here 1.04) depend on the citation time window (here 1 January 2008 to 5 January 2013), impact-factor year and data source (here Thomson Reuters Web of Science). Journals included in the sample span the physical and chemical sciences, the biological and medical sciences, the earth and environmental sciences, and engineering.

As a case in point, the median number of citations for *PLoS ONE* is 1 whereas its 2011 impact factor is 4.1, largely because since 2008 it has increased its output more than six-fold⁵ (from less than 3,000 papers in 2008 to about 19,000 in 2012). The impact factor, being a lagging indicator with a narrower time window, has yet to reflect this.

It is therefore clear that but for outliers^{6,7} the impact factor is an appropriate measure of journal quality according to citations. And it is also beyond question that the impact factor does not generally correlate to the performance of individual researchers or to citations to individual papers^{2,8,9}. As with any statistical measure, it is unsafe to use it as a proxy for an unrepresentative subset of the original sample. It would thus be unwise, for instance, to rate scientists on the basis of the total number of papers weighted according to the impact factor of the journal where they have been published. A simple exercise proves the point: pick a few scientists and rank the papers they published five years ago in decreasing order of citations alongside the impact factor of the corresponding journal in that year. The odds are that, if there is any

correlation at all, this is weak or the outliers are plentiful.

As Fig. 1 shows, half of the papers published by *Nature Materials* in the past five years have received more citations than at least half the papers published in most other journals (that is, any journal with a lower impact factor). The median and its predictor the impact factor are therefore quality signals that are valid for comparisons between journals publishing on similar scientific topics. Yet beware of those who use them instead of article-level metrics¹⁰ when assessing a small subgroup of papers or authors. Impact factors should have no place in grant-giving, tenure or appointment committees. □

References

1. <http://occamstypewriter.org/scurry/2012/08/13/sick-of-impact-factors>
2. Cameron, B. D. *Portal Libraries Acad.* **5**, 105–125 (2005).
3. *Nature Mater.* **10**, 645 (2011).
4. Campanario, J. M. *Scientometrics* **87**, 189–204 (2011).
5. *Nature Mater.* **11**, 353 (2012).
6. Thomson Reuters; <http://go.nature.com/yPsfuh>
7. Campanario, J. M. *J. Am. Soc. Inf. Technol.* **62**, 230–235 (2011).
8. Rostami-Hodjegan, A. & Tucker, G. T. *Br. J. Clin. Pharmacol.* **51**, 111–117 (2001).
9. Epstein, R. J. *J. Invest. Med.* **52**, 531–536 (2004).
10. *Nature Mater.* **11**, 907 (2012).