

Journal metrics: what's in a number?

Metrics of research assessment should be applied with judgment and clarity to avoid undermining scientific efforts.

Research aims to be impactful by expanding our knowledge and benefiting society. Despite the proliferation of metrics for scientific output at the journal, article and researcher level, 'impact' remains ill-defined, and its assessment is often controversial, as exemplified by the use, and misuse, of the most well-known bibliometric, the journal impact factor (JIF).

Released by Clarivate Analytics, the JIF measures the mean number of citations received in a particular year by papers published in a journal over the two previous years. Thus it can be a useful measure of journal citation rates, and inasmuch as citability alone can be such an indicator, it might also tell us something about how impactful the cumulative publication output of a journal might be — but important caveats cannot be overlooked. A journal's impact factor has little meaning as an absolute number and is only informative when compared to other journals. Such comparisons may be misleading, for example, when the journals serve different disciplines characterized by distinct citation patterns; have different publication outputs and therefore different spreads of citation distributions; or publish content that serves altogether distinct purposes in the scientific record — most notably, the JIF conflates primary research and review articles. Being an arithmetic mean the JIF allows for a few very highly cited papers to skew the citation distribution, underrepresenting the typically longer tail of less cited papers. Indeed the 2-year median citation score that we provide on the Nature Research journals metrics website to complement the JIF, and which is not subject to such outliers, is lower for all journals listed (<http://go.nature.com/2arq7OM>). It is quite telling that skewed distribution curves are observed for the 2015 JIF of multiple journals, with 65–75% of the articles having fewer citations than the JIF (see Lariviere, V. *et al.* Preprint at <http://doi.org/bmc2;2016>), and this holds true for 67% of the *Nature Cell Biology* 2016 JIF citable content when the same methodology is applied. But caution should be exercised when assigning significance to the number of citations — by limiting citable items to a two-year period the JIF effectively overlooks papers of which the impact might not be readily apparent, for example, because follow-up work is lengthy, or because they are pioneer studies that will contribute to new lines of research years later. Looking at the 5-year JIF could be more informative in that respect. Measuring citations across the publication output of a journal also brushes over field-specific citation trends — slow-burn or mature research areas may have lower citation rates than rapidly advancing ones, but that is not to say that the former are not important.

Although the JIF was developed to assess citation frequency at a journal level, this limited metric has been widely used out of context as an indicator of researcher performance or the quality of individual papers. Its shortcomings and misuse have long been bemoaned in the community, including in editorials of *Nature Cell Biology* and the *Nature* journals (see <http://go.nature.com/2t2Ng30>). The undue emphasis on this single metric spurred efforts such as the San Francisco Declaration

on Research Assessment (DORA; <http://go.nature.com/2vbr8jt>), which in December 2012 pledged to move away from relying on the JIF for evaluating research output and outlined a set of recommendations to institutes, funders and publishers. The policies of *Nature Cell Biology* had long been aligned with the spirit of DORA (see *Nat. Cell Biol.* **16**, 1; 2014) and earlier this year we reasserted our commitment to these principles by becoming formal signatories (*Nature* **544**, 394; 2017). We have also welcomed further initiatives such as the Leiden Manifesto for Research Metrics (*Nature* **520**, 429–431; 2015), which in 2015 proposed a set of ten principles for research evaluation, as well as a move from funders around the world, including the Research Councils UK, the European Molecular Biology Organization (EMBO), the Canadian Institutes of Health Research and Australian Research Council, to remove the influence of the JIF when assessing grant applications.

Recent years have also seen the development of multiple metrics aiming to capture distinct aspects of scientific output, each with its own limitations. The Eigenfactor Score measures the number of citations received by a journal over five years, but weighs their origin, with highly cited journals influencing the score more. It also increases with publication output, but when divided by the percentage of articles published in the journal it provides the Article Influence Score, which is considered roughly analogous to the 5-year JIF. The Immediacy Index calculates the average number of times an article is cited in its publication year, but as for the JIF this does not mean that less rapidly cited articles are less important. Separate from these, the *h*-index attempts to measure productivity and citability of individual scientists, but continuously increases with age and does not account for different author contributions to a paper — more prolific, older authors will have higher scores. To provide a more complete view we list several metrics together with information on peer-review performance, such as times from submission to decisions and publication, for the Nature Research journals (<http://go.nature.com/2arq7OM>).

And yet the pull of the JIF remains strong, as it consistently comes up in surveys about how researchers decide where to submit their work. Nevertheless, more qualitative aspects such as quality of the peer review, interactions with editors, and a journal's readership and overall reputation also weigh heavily, and these cannot be captured by mathematics. The simplicity of ascribing importance through numbers when deciding where to publish, what research to fund, or which researcher to hire may be appealing, especially for scientists, who are data-driven by inclination and training. But this reductionist approach holds many pitfalls — a major one being that numbers do not allow for nuance. There is no one-measure-fits-all approach to evaluate a journal's impact, a researcher's achievements, or a paper's significance. Combining appropriate metrics with more qualitative indicators, such as scientific rigour as revealed through the peer review and post-publication evaluation, and broader contributions to a field, including policy and scientific practice, could yield more complete assessments.