

Editorial

Impact fact-or fiction?



The EMBO Journal (2013) 32, 1651–1652. doi:10.1038/emboj.2013.126; Published online 17 May 2013

The Journal Impact Factor dominates research assessment in many disciplines and in many countries. While research assessment will always have to rely to some extent on quantitative, standardized metrics, the focus on this single measure has gone so far as to hamper and distort scientific research. The Declaration on Research Assessment (DORA), signed by influential journals, funders, academic institutions and individuals across the natural sciences, aims to raise awareness and to redress the use of non-objective research assessment practices.

When Eugene Garfield was considering which scientific journals to include in the Science Citation Index (SCI) more than half a century ago, he rightly noted that the importance of a journal does not necessarily relate to its size. Citations to articles in a journal appeared to provide a quantitative means to assess the interest of the scientific community in a journal. To obtain up to date information, Garfield considered only citations made in the last full year, and, to control for publication volumes, he simply divided by the number of articles. At the time, changing the assessment window did not appear to matter too much (Garfield, 2006) and Garfield settled on considering citations to papers from the preceding 2 years for his new bibliometric tool: the *Journal Impact Factor (JIF)* was born. The aim of the *JIF* was to select appropriate journals for SCI, although it soon led to the first 'Journal Citation Reports' (JCR) in 1969.

The *JIF* quickly became a tool favoured by librarians to ensure their collections contained all the key journals, not merely large journals. As the number of journals grew, authors increasingly referred to the JCR to find an appropriate home for their papers beyond the few journals in their immediate area that they were well acquainted with. Molecular biology expanded rapidly in those years and journal ranking became an apparently useful tool to filter out the relevant literature to read. As journal numbers skyrocketed, the *JIF* was published to three decimal places, according to Garfield a move to avoid listing journals with identical *JIFs* (Garfield, 2006). Garfield himself considers this practice questionable, realizing that it suggests a misleading level of objectivity.

Citation-based metrics of course emerged well before quantifiable webaccess, and arguably they remain a more meaningful measure, as a citation will usually indicate that an author has read a piece of work and found it sufficiently important to refer to it in a scholarly discourse (although often the literature analysis remains at the level of a review, which then also tends to be cited instead of primary papers).

Global research output has skyrocketed to over 1 million biomedical papers annually. No field has grown faster than molecular biology, as the techniques have spread to all walks of biology. Consider the tumour suppressor p53. After the foundation of the field in the early 1980s, only a handful of papers were published annually until the 1990s, when the field exploded to reach steady state for the last decade at well over 3000 papers annually. Not even experts can keep pace with the literature at this rate.

As fields mature, papers become increasingly specialized—that means harder to assess by colleagues working outside the immediate field, who are also increasingly pressed for time in evaluating an exploding, globalized research community. Faced with ever more applications for positions from around the world in fields peripheral to their expertise, who can blame those charged with research assessment to defer to journal name or *JIF* as a proxy for quality? However, this is not a reason to continue with 'business as usual', as an objective research assessment infrastructure is essential to support an optimal research enterprise.

The *JIF* is a limited measure even for the role for which it was developed: to compare journals. To name but two key problems, there are field-specific differences in citation rates and it conflates citations to primary research papers and the reviews literature. Consider how much more limited its reach is when assessing the research performance of individuals! Research assessment by funders and institutions must not push the burden of their decision making to journal editors. The role of editors is to publish the papers that are most appropriate for their journals—and no more.

When use of the *JIF* as a proxy for the quality of a specific paper spills over to authors, editors and referees, conference organizers and prize committees, a dangerous circle in the research ecosystem is closed that gravitates around the *JIF* as the single pseudo-objective and -universal quality criterion. In an era where institutions set specific minimal *JIF* thresholds for their researchers, and where funding—and in some countries personal financial bonuses—is awarded based on journal name, it is not surprising that researchers feel overbearing pressure to publish a paper at all costs in a journal with a name or *JIF* that translates into sufficient academic credit. One may not be surprised if occasionally research and publication ethics are deemed to play a subservient role. The breadth, depth and thoroughness of a study may affect the *JIF* less than the nature of the claims made, the editorial scope of the journal and, indeed, the journal name itself. In times of tight funding, a measured, careful scholarly publishing culture may find it harder to compete with a more cavalier 'the end justifies the means' approach to publishing.

The *JIF* is limited in other ways. Individual papers in most journals exhibit a very broad spectrum of citation rates. Rather than being a sign of editorial incompetence, this can indicate that a journal does not merely aim to enrich for high citation potential. However, as a result the *JIF* may well be very different from the citation rate to a specific article. Conversely, as the *JIF* represents the average citations, not the median, highly citing single papers may single handedly inflate the *JIF*. Furthermore, reviews tend to accumulate more citations because of their broad scope, in particular, when reference lists of primary research papers are limited. Indeed, the academic standing of journals and individuals alike can be buttressed by the prolific publication of reviews.

Much has been said and written about these limitations of the *JIF*, and recently a number of alternative citation-based journal level metrics have emerged (such as h-index, SCImago, EigenFactor or the 5-year IF). Some are weighed to test the longevity of the findings, the network-level cross-referencing of

papers, or they compensate for field-specific citation dynamics. More recently, webaccess, social media activity and file sharing have started to give citation-based metrics a run for their money. The attraction is of course the reduced lag phase, which makes article level-based metrics more useful. Nevertheless, these developments have not dented one bit the predominance of the '2 previous years' papers cited last year' formula. The attraction of the *JIF* is that it appears to be predictive of future performance of a journal or a researcher, that it tends to correlate somewhat with a researcher's anecdotal perception of journal quality, and that it entices to comparisons between disciplines. Metrics informed research assessment is of course much easier with such a single, apparently universal, measure than a compendium of analytical tools, which only rarely show consistent patterns.

Using the *JIF* to obtain a rough indication of journal performance is not unreasonable, but given the obvious limitations, this exercise can rapidly turn into comparing apples with oranges and nobody should kid themselves that anything beyond the integers represents meaningful differences. However, when the *JIF* is applied to evaluate individuals, it becomes a dangerously inaccurate tool. At minimum, citation-based metrics should be used to assess the performance of individual papers. The inherent lag phase of course renders it of limited value in assessing the potential of researchers. Research assessment will always tend to prefer prospective metrics. To be clear, the *JIF* is not inherently flawed and Thomson Reuters is certainly not to blame for the state of affairs. The scientific community itself has to avoid misusing the *JIF* and needs to develop more robust research assessment. In the future, article-based metrics based on online usage and sharing may become a meaningful approach, but only to compare individuals in a similar field and research environment. At present, however, article-based metrics also still have to be taken with a big pinch of salt. For one, there are reasons other than quality or importance to access, share and cite an article.

Scientific journals play an important role in the research ecosystem, and journal editors struggle just as much as researchers to ensure that their *JIF* will continue to attract papers of a quality expected for their journal. Some have resorted to enrich their journals with papers in high citing subjects or increase the ratio of reviews to primary research papers. Few even game the system by encouraging citations to journal content or by publishing papers that accumulate citations while not being classed as citable by ISI. Journals find themselves in a similarly competitive environment to researchers and concerted action at journal level to enhance the *JIF* can influence the whole research ecosystem, especially when this is mirrored at the level of the researcher who feels pressure to pursue citation-rich research.

A group of editors of a number of scholarly journals met in December to discuss the Impact Factor. Remarkably, the discussion switched within minutes from discussing potential problems to what to do about the problems. That discussion took longer, but it resulted in two notable developments:

The group decided to send a letter to Thomson Reuters to argue for three enhancements to the *JIF*:

1. a separate *JIF* for reviews and for primary research papers to avoid conflating these two disparate article types
2. the need for transparency; in particular, release of the data used to calculate the *JIF* and declaration of the non-citable content and citations to that content.
3. publication of a median-based *JIF* alongside the mean to visualize biases by individual high citing papers.

No reply to this letter had been received at the time of going to press.

Realizing that the most effective way to change a tight ecosystem is to effect change at the level of all of the layers of the system, the group agreed to reach out to all the stakeholders: funding agencies, research institutions, scientists, journals and the organizations that supply metrics. The San Francisco Declaration on Research Assessment (DORA, <http://www.ascb.org/SFdeclaration.html>) includes the following recommendations:

1. Avoid using journal metrics to judge individual papers or individuals for hiring, promotion and funding decisions.
2. Judge the content of individual papers and take into account other research outputs, such as data sets, software and patents, as well as a researcher's influence on policy and practice.
3. Balance the Impact Factor with other metrics and reduce emphasis on the *JIF* in journal promotion. Article-level metrics are more specific than journal-based metrics.
4. Declare detailed authorship contributions.
5. Avoid limits on reference lists and remove reuse and access limitations. Where appropriate, cite the primary literature.
6. Open data used to calculate metrics.
7. Account for article types in reporting metrics; define what constitutes inappropriate manipulation of metrics.
8. Promote and teach best practice focussing on the value and influence of specific research outputs.

These recommendations were made to ensure that the research paper will remain a central mode for sharing research findings in the future, although it also extends to other important research outputs, such as data sets.

Note that it would be implausible to envisage a metrics-free research community, and DORA instead argues that limitations of the *JIF*, in particular, should be considered with due care. Note also that the statement is not arguing against reviews, it does not wholesale ban acknowledgement that the *JIF* exists, nor does it presuppose a complete lack of format requirements for references—instead, it argues for measured policies designed to avoid an overreliance on the *JIF* in the assessment of individuals and institutions.

The EMBO Journal and EMBO support DORA. A remarkable spectrum of signatories has accumulated, ranging from major research institutes and funders to concerned researchers across the spectrum from applied mathematics to medicine. We hope that many other individuals, academic institutions, funders and policymakers endorse the statement to show that use of the *JIF* should be restricted largely to what it was designed for: a ballpark assessment of a journal's significance. Additional metrics may emerge that will enhance research assessment, but until that time, anyone who is serious about research assessment would be well advised to read the papers themselves or to consult expert referees. DORA is open for signing by individuals and institutions!

Bernd Pulverer

Chief Editor
The EMBO Journal and Head of Scientific Publications, EMBO,
 69117 Heidelberg, Germany
 E-mail: bernd.pulverer@embo.org

Reference

Garfield E (2006) The history and meaning of the journal impact factor. *JAMA* **295**: 90–93