Science & Society

EMBO reports

The journal Impact Factor and alternative metrics

A variety of bibliometric measures has been developed to supplant the Impact Factor to better assess the impact of individual research papers

Lutz Bornmann[1] & Werner Marx[2]

Journal impact factors (JIFs) have become a widely used tool to judge the quality of scientific journals and single publications. JIFs are calculated by the scientific division of Thomson Reuters and published annually in the Journal Citation Reports (JCR). At first, the JCR's origin was guided by the needs of librarians who wanted to use a quantitative method to select journals for their holdings. Approximately 11,000 academic journals are currently listed in the JCR and the JIF has become one of the most important indicators in evaluative bibliometrics. Although this metric was never designed for evaluating papers or individuals, rather for evaluating journals as a whole, the availability of the JIFs has turned it into a common tool for evaluating research. It is especially common in Europe to use JIFs as a basis for making decision on research grants, hiring, and salaries. However, JIFs are not statistically representative of individual papers and correlate poorly with their actual citations. A study of six economics journals showed "that the best article in an issue of a good to medium-quality journal routinely goes on to have much more citations impact than a 'poor' article published in an issue of a more prestigious journal" [1].

There is a growing unease within the scientific community, among journal publishers and within funding agencies, that the widespread misuse of JIFs to measure the quality of research—with profound impact on researchers' careers—is detrimental for science itself. The San Francisco Declaration on Research Assessment (DORA), initiated by the American Society for Cell Biology together with editors and publishers, calls for moving away from using JIFs to evaluate individual scientists or research groups and developing more reliable ways to measure the quality and impact of research. Various funding agencies have also begun to discourage the use of JIFs in their funding decisions and instead ask applicants to submit only their most relevant papers in contrast to papers published in high-impact journals. Here we discuss the JIF and its various shortcomings for evaluating individual publications or researchers and various alternatives, in particular the new Relative Citation Ratio (RCR) that is now being used by the US National Institutes of Health (NIH).

> *"Various funding agencies have also begun to discourage the use of JIFs in their funding decisions and instead ask applicants to submit only their most relevant papers..."*

Thomson Reuters calculates the JIF by taking the number of journal publications within a 2-year window and summing up their citations over the following year. The number of citations is then divided by the number of citable items. However, this calculation is plagued by errors and inconsistencies, particularly because the numerator counts all citations to all types of publications, while the denominator considers only the number of the so-called citable documents. The San Francisco Declaration on Research Assessment therefore recommends against using "journal-based metrics, such as impact factors, as a surrogate measure of the quality of individual research papers, to assess an individual scientist's contributions, or in hiring, promotion, or funding decisions" (http://www.ascb.org/dora). A report from the International Mathematical Union states: "While it is incorrect to say that the impact factor gives no information about individual papers in a journal, the information is surprisingly vague and can be dramatically misleading" [2].

Even the use of JIFs for their original purpose—to indicate a journal's performance within a narrow subject category—should be carried out with caution. JCR reports JIFs on a scale with three decimals, which allows them to order journals by rank. The three decimals create, however, the impression of data precision, which cannot be expected for bibliometric data. For example, Moed [3] performed a comprehensive study on the accuracy of cited references in the Web of Science, which is the basis for calculating the JIF: Matching 22 million cited references to their target paper, he found missed match for 7.7% of papers. According to the Leiden Manifesto for research metrics, "the journal impact factor is published to three decimal places to avoid ties. However, given the conceptual ambiguity and random variability of citation counts, it makes no sense to distinguish between journals on the basis of

1 Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Munich, Germany. E-mail: bornmann@gv.mpg.de
2 Max Planck Institute for Solid State Research, Stuttgart, Germany
DOI 10.15252/embr.201642823 | Published online 28 June 2016

very small impact factor differences. Avoid false precision: only one decimal is warranted" [4].

> *"Even the use of JIFs for their original purpose – to indicate a journal's performance within a narrow subject category – should be carried out with caution"*

Many journals in the JCR have a comparably low number of publication and citations, which can lead to large variations of the JIFs over the years. For example, *Physics of Life Reviews* has published between 10 and 14 citable items between 2007 and 2014 and the JIFs vary between 2.545 and 9.478.

The citation window of the standard JIF is very short with only one year following the publication years. This contradicts conventions in bibliometrics to use citation windows of at least 3 years. Citation impact needs time to accrue.

There are not only strong field-dependent citation cultures, but also subfield-dependent citation habits, which makes the comparison of JIFs from different subject categories meaningless, and a comparison between subfields highly questionable. Journals assigned to subject categories such as chemistry or physics may not be comparable with journals assigned to materials science. Journals from various subfields in materials science—such as biomaterials versus textiles—are hardly comparable.

Journals within a specific JCR subfield are often different with respect to their publications. Owing to the differing citation characteristics of document types such as research articles, letters, commentaries, and reviews, different journal types are not comparable among each other.

Some high-impact journals such as *Nature* and *Science* publish not only research papers, but also a large number of editorials and news articles. These items may be well cited, but they are not counted in the denominator, which leads to a substantial overestimation of their JIFs. Another problem for reliable JIF calculations comes from citing different versions of the same journal, such as *Angewandte Chemie* (AC) that is being published in the original German edition, and in an International English language edition since 1962. As some authors cite papers published in AC with reference to both the German and the International edition, citations to AC are counted twice, thus artificially inflating the JIF by about 15%.

The publication of several versions of the same manuscript during a two-stage publication process can also limit the validity of the JIF. The open-access journal *Atmospheric Chemistry and Physics* (ACP) for instance first publishes submissions on the ACP Web site in *Atmospheric Chemistry and Physics Discussions* (ACPD) before the final manuscript is published in ACP after peer review. This could also inflate the JIF if citations to papers in ACP and papers in ACPD enter into the numerator, but only ACP papers enter into the denominator. In this specific case, however, Thomson Reuters distinguishes between the two different editions and calculates a correct JIF for ACP.

Meanwhile, additional indicators for measuring the impact of a journal have been established and included in the JCR. The Eigenfactor Score puts stronger emphasis on citations coming from highly cited journals than those coming from less cited journals. The Article Influence is calculated by dividing a journal's Eigenfactor Score by the number of papers in the journal. Both measures consider the journal's papers over the first 5 years after publication and exclude journal self-citations.

The most important additional indicators in the JCR are the Cited Half-Life and the Citing Half-Life. The Cited Half-Life is defined by Thomson Reuters as: "the number of years, going back from the current year, that account for 50% of the total citations received by the cited journal in the current year" (http://science.thomsonreuters.com). The Citing Half-Life is defined as "the number of journal publication years, going back from the current year, that account for 50% of the total citations given by the citing journal in the current year" (http://science.thomsonreuters.com). The Cited Half-Life reflects how long the papers are remembered within the scientific community. The Citing Half-Life reflects the citation practice of the journal's authors concerning other papers. From the point of view of a journal's papers, the Cited Half-Life can be regarded as passive—performed by colleagues, mostly publishing in other journals—whereas the Citing Half-Life is active since it is done by the authors of the journal's papers. The Cited Half-Life provides information about how a journal's papers are remembered by the community (i.e., their long-term impact) and can be seen as more significant than the Citing Half-Life.

In addition to the alternative metrics provided in the JCR, Elsevier and online databases such as Index Copernicus and VINTI are publishing other metrics. The most prominent metric is the SCImago Journal Rank (SJR) indicator "that ranks scholarly journals based on citation weighting schemes and eigenvector centrality" [5]. SJR assigns different values to citations depending on the importance of the journals.

> *"The RCR is a new approach used to normalize citations on the cited-side, because it relies on co-citations to generate the reference set"*

The *h* index for journals was introduced as a robust alternative indicator advantageously supplementing journal impact factors and is calculated in the same way as the *h* index for individual scientists. A number of complementary indices and alternatives have been put forward, but a meta-analysis showed high inter-correlations between JIF, *h* index, and different variants. It does not matter which indicator is used in journal evaluation.

A research group affiliated with the NIH developed the Relative Citation Ratio (RCR) as an alternative to JIF for measuring the impact of single publications [6]. It is rooted in the long-standing bibliometric tradition of using field-normalized indicators to measure citation impact instead of bare citation counts. In bibliometrics, two methods exist for calculating field-normalized citation counts: cited-side and citing-side normalization. For cited-side normalization, the citation counts of a paper are compared with the citation counts of papers in a reference set that were published in the same subject category and publication year. For citing-side normalization, each citation of a paper is weighted by the citation density of the corresponding citing

paper's subject category. The idea behind citing-side normalization is that the number of references reflects the citation density of the subject category in which the citing paper was published.

The RCR is a new approach used to normalize citations on the cited-side, because it relies on co-citations to generate the reference set. All papers co-cited with the paper in question are considered to represent the subject category of the paper and therefore its reference set. Stefano Bertuzzi, executive director of the American Society for Cell Biology, "applauds the NIH for moving away from the journal impact factor (JIF). He wrote that the metric 'evaluates science by putting discoveries into a meaningful context. I believe that the RCR is a road out of the JIF swamp'" [7].

"... one should keep in mind that bibliometric numbers are only a proxy of research quality, which measure one part of quality, namely impact or resonance"

However, Ludo Waltman recdently criticized that "the RCR metric doesn't live up to expectations" (http://www.cwts.nl/blog?article=n-q2u294&title=nihs-new-citation-metric-a-step-forward-in-quantifying-scientific-impact#sthash.w1KC3A1O.dpuf). He used a single publication as a fictitious example, which received citation impact from papers published in different subject categories to show that "publications may be penalized rather than rewarded for receiving interdisciplinary citations". New citations from a subject category with high citation density could mean that a paper's RCR decreases instead of increases. Waltman therefore does not regard the new indicator as an equitable alternative to the established field-normalized indicators already used in bibliometrics.

Bornmann and Haunschild investigated the RCR by correlating it with established field-normalized indicators: The Mean Normalized Citation Score (MNCS), a quotient composed of a paper's citations (numerator) and the average citation counts of the papers in the corresponding reference set (denominator); citation percentiles that sort papers in the reference set by their citations to rank a given paper; and the $SNCS_{(2)}$ which weighs each citation to a single paper by the number of cited references in the citing paper. Their analysis reveals that the RCR correlates highly with the established indicators [8]. It thus questions the necessity to introduce a new advanced bibliometric indicator, which is more complicated to calculate, in addition to the established alternatives.

Notwithstanding its shortcomings and the various alternatives used to measure the impact of individual papers, the JIF is still used in the scientific community as a basis for decision making in different contexts. "The JIF has reached such dominance that it influences the publication strategies of journals, hiring at institutions and even how researchers cite" [9]. However, this use of JIFs is intolerable. If an evaluation is based on bibliometric data, the citation impact of the respective papers should be determined. The ideal way of measuring citation impact in bibliometrics is using field-normalized indicators on the level of single publications. We have explained three advanced indicators; a broader overview of citation impact measures can be found in [10].

Even the use of JIFs for their originally intended aim—to compare journals—is afflicted by various shortcomings and must be carried out with caution. Meanwhile, additional indicators for measuring the impact of journals have been added to the JCR data. The Citing and the Cited Half-Life indicators provide information about how long the papers are remembered. The Eigenfactor Score and the Article Influence Score consider which journals have contributed citations. These additional journal indicators, together with the JIFs based on a 5-year time frame and the *h* index, often show a high degree of correlation among each other. Thus, JIFs are not useless, but—as Thomson Reuters states itself—"...should not be used without careful attention to the many phenomena that influence citation rates, as for example the average number of references cited in the average paper. The impact factor should be used with informed peer review" (http://thomsonreuters.com/products_services/science/free/essays/impact_factor/).

Bibliometric indicators are generally very helpful for studying the performance of individual researchers, research groups, institutions, and countries. The data is available in large databases and field-normalized indicators facilitate cross-field comparisons. However, one should keep in mind that bibliometric numbers are only a proxy of research quality, which measure one part of quality, namely impact or resonance. Two other important parts cannot be measured by citations, namely the accuracy and importance of research. This might be the reason why correlation studies between bibliometrics and expert opinions do not show a perfect relationship.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Oswald AJ (2007) An examination of the reliability of prestigious scholarly journals: evidence and implications for decision-makers. *Economica* 74: 21–31
2. Adler R, Ewing J, Taylor P, Hall PG (2009) A report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). *Stat Sci* 24: 1–28
3. Moed HF (2005) *Citation Analysis in Research Evaluation*. Dordrecht, the Netherlands: Springer
4. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I (2015) Bibliometrics: the Leiden Manifesto for research metrics. *Nature* 520: 429–431
5. Gonzalez-Pereira B, Guerrero-Bote VP, Moya-Anegon F (2010) A new approach to the metric of journals' scientific prestige: the SJR indicator. *J Informetr* 4: 379–391
6. Hutchins BI, Yuan X, Anderson JM, Santangelo GM (2015). Relative Citation Ratio (RCR): A new metric that uses citation rates to measure influence at the article level. Retrieved November 6, 2015, from http://biorxiv.org/content/early/2015/10/22/029629
7. Bloudoff-Indelicato M (2015) NIH metric that assesses article impact stirs debate. The US biomedical funding agency is

using the metric to evaluate funded research. Retrieved November 10, 2015, from http://www.nature.com/news/nih-metric-that-assesses-article-impact-stirs-debate-1.18734

8. Bornmann L, Haunschild R (2016) Relative Citation Ratio (RCR): an empirical attempt to study a new field-normalized bibliometric indicator. *J Assoc Inf Sci Technol* doi: 10.1002/asi.23729
9. Pulverer B (2015) Dora the Brave. *EMBO J* 34: 1601 – 1602
10. Waltman L (2016) A review of the literature on citation impact indicators. *J Informetr* 10: 365 – 391