# Essays

## Journal Impact Factor: "the poor man's citation analysis" and alternative approaches

**Werner Marx**
*Max Planck Institute for Solid State Research, Heisenbergstr.1, D-70569 Stuttgart, Germany; w.marx@fkf.mpg.de*
**Lutz Bornmann**
*Division for Science and Innovation Studies, Administrative Headquarters of the Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany; bornmann@gv.mpg.de*

**Abstract** The Journal Impact Factor has a number of drawbacks preventing its use for assessment of separate journal articles and individuals. With that in mind, most experts would endorse the San Francisco Declaration on Research Assessment (DORA), which highlights the appropriate use of bibliometric indicators for quantitative research assessments. To curb the problem of skewed citations, an alternative, normalised metric is proposed. Percentiles, or percentile rank classes method, is particularly useful for normalisation. It is also advisable to use specific percentile rank classes and to assess individual scientists with $P_{top\ 10\%}$ or $PP_{top\ 10\%}$ indicators.

**Keywords** Bibliometrics; research evaluation; alternative metrics.

In the process of quantitative (bibliometric) research evaluation, citation analysis may be erroneously replaced by the use of the journal impact factor (JIF).[1] This is unacceptable, since the JIF is merely an impact measure for scholarly journals. It was originally proposed to help librarians distinguish influential journals of interest to their readership, but not to evaluate a single paper in a journal or research performance of a scientist.[2]

Experts in bibliometrics are well aware that the JIF has a number of drawbacks preventing its use for research assessment. Most importantly, the distribution of citations to a journal's articles is often highly skewed since a large number of citations go to a few items in the journal. As a result, citation rates are influenced by a small fraction of highly cited items. The JIF's timeframe (two years) is often too short for comprehensive evaluation of a journal performance in slowly developing disciplines. Adjustment of citation behaviour for disciplines, cross-disciplinary comparisons and comparisons of journals publishing predominantly certain types of articles (eg reviews, original research papers) are impossible with the use of JIF.

Anthony van Raan once noted that "if there is one thing every bibliometrician agrees, it is that you should never use the journal impact factor to evaluate research performance for an article or for an individual — that is a mortal sin".[3]

**References** - continued from page 61

1 Kennedy C. The Development and Use of Stereo Photography for Educational Purposes. *Journal of the Society for Motion Picture Engineers* 1936;26:3-17. doi: 10.5594/J01280

2 Spottiswoode R, Spottiswoode NL, Smith C. Basic Principles of the Three Dimensional Film. *Journal of the Society for Motion Picture and Television Engineers* 1952;59:249-286. doi: 10.5594/J01778

3 Joe Levine, IEEE, personal communication, 23 September 2011.

4 Jay Morreale, SID, personal communication, 1 September 2011.

5 John Dennis, NSA, personal communication, 1 September 2011.

*6 The Associated Press Stylebook and Briefing on Media Law*. New York: Associated Press; 2013. ISBN 978-0-917360-57-2.

*7 The Yahoo! Style Guide*. New York: Yahoo! Inc; 2010. p. 483. ISBN 978-0312569846.

8 Kathy Sheehan, SPIE, personal communication, 29 May 2013.

9 Mark Fihn, Veritas et Visus, personal communication, 28 May 2013.

10 Woods AJ, Merritt JO, Fisher S, Bolas M, Benton S, Holliman NS, et al. (eds). *Selected SPIE/IS&T papers on DVD-ROM: Stereoscopic Displays and Applications 1990-2009: A Complete 20-Year Retrospective - and The Engineering Reality of Virtual Reality 1994-2009 (CDP51)*. Bellingham (WA): SPIE; 2010. ISBN: 978-0-8194-7659-3

11 Strunk W, White EB. *The Elements of Style*. 4th ed. Boston, Massachusetts: Allyn & Bacon; 2000. ISBN: 0-205-30902-X

12 *Stereo World*. 38(6) (May/June 2013). Portland, Oregon: National Stereoscopic Association (NSA); 2013. Available at http://www.stereoworld.org/ [Accessed 19 June 2013].

13 *Information Display*. 9(3) (May/June 2013). Campbell, California: Society for Information Display (SID); 2013. Available at http://informationdisplay.org/ [Accessed 19 June 2013].

14 *IEEE Spectrum*. 50(5) (INT) (May 2013). New York: Institute of Electrical and Electronics Engineers (IEEE); 2013. Available at http://spectrum.ieee.org/ [Accessed 19 June 2013].

15 *SPIE Professional*. 8(2) (April 2013). Bellingham, Washington: SPIE; 2013.

16 *It Is Innovation (i3)*. 1(3) (May 2013). Arlington, Virginia: Consumer Electronics Association (CEA); 2013. Available at http://www.ce.org/i3 [Accessed 19 June 2013].

17 *3rd Dimension*. 8(2) (May 2013). Temple, Texas: Veritas et Visus; 2013. Available at http://www.veritasetvisus.com/3rd_dimension.htm [Accessed 19 June 2013]

18 Google Trends [Internet]. Mountain View, California: Google Inc.; 2013. [updated 30 May 2013; cited 30 May 2013]. Available at www.google.com/trends

19 Eric Kurland, 3-DIY, personal communication, 28 May 2013.

20 Lenny Lipton, Leonardo IP, personal communication, 8 Nov 2011.

21 Lipton L. *Foundations of the Stereoscopic Cinema – A Study in Depth*. New York: Van Nostrand Reinhold Company; 1978.

He called such evaluation 'the poor man's citation analysis'.[1] With that in mind, most experts would endorse the San Francisco Declaration on Research Assessment (DORA),[4] which aims to turn the authors' attention to appropriate bibliometric indicators fit for quantitative research assessments. DORA has attracted a lot of comments and responses, including a statement from Thomson Reuters reiterating the inappropriateness of JIF as a measure of the quality of individual articles and encouraging authors to choose publication venues based on factors not limited to the JIF.[5] Nonetheless, it is unlikely that alternative and more appropriate citation metrics will soon gain recognition as research assessment tools outside the community of bibliometricians.

Comparing citation counts to individual journal articles is more informative than weighing JIF values of the journals containing these articles. Unfortunately, the meaning of these citation metrics is not widely understood. For bibliometricians, citation analysis is the impact measurement of individual scholarly items based on citation counts. Citation impact is just one aspect of the article's 'quality', which complements its accuracy and originality. Since a clear definition of the scientific quality does not exist, no all-in-one metric has yet been proposed. At the same time, it is well known that the citation-based data correlate well with research performance (quality) asserted by peers. A prime example of the latter is the UK research assessment exercise ratings, which proves that citations can be used as a proxy for measuring research performance, provided the indicators and measurements are designed and approved by bibliometricians.

Proposal of a new bibliometric indicator usually stems from empirical observations. One is that the differences in average citation counts in various disciplines depend on the activity and productivity of the contributors. Citation rates are time-dependent: the older the publication, the more likely it is highly cited. Comparing citation counts in various disciplines and at different time points is incorrect, unless there is a proper standardisation or normalisation. Normalisation is possible by using reference sets,[6] which assess the citation impact of comparable publications. The reference sets contain publications that were published in the same year and subject category. The arithmetic mean of the citations for all publications in a reference set is calculated to specify the expected citation impact.[7] This enables to calculate the Relative Citation Rate (RCR) - the observed citation rate of an article divided by the mean expected citation rate.

As with the JIF, the calculation of RCR has an inherent disadvantage related to the lack of normalisation of citations for subject category and publication year. To curb the problem of skewed citations, an alternative, normalised metric should be used. Percentiles, or percentile rank classes method is particularly useful for the normalisation.[8]The percentile of an article gives an impression of the impact it has achieved in comparison to similar items in the same publication year and subject category. Unlike the RCR, percentiles are not affected by skewed distributions: highly-cited items do not receive excessively high weight.

The relative ease of the percentiles' calculation is one of their advantages. All publications in a given year and subject category provide the reference set. The citations of these publications are the yardstick. The publications are sorted by citation numbers and are broken down into percentile ranks ranging between 0 and 100. The percentile of a publication is its relative position within the reference set: the higher the rank, the more citations the publication has. For example, a value of 90 indicates that the publication belongs to the 10% of most-cited ones. A value of 50 is the median level, which means an average impact.

The publication set for the percentiles methods ranges from single articles to publication records of an individual scientist or an institution. The percentiles for a certain publication set can be analysed by different methods.[9] Along with the percentiles, it is possible to focus on specific percentile rank classes, and particularly on the assessment of individual scientists with $P_{top\ 10\%}$ or $PP_{top\ 10\%}$ indicators.[10] Both indicators count the number of successful publications normalised for publication year and subject category. $P_{top\ 10\%}$ is the number and $PP_{top\ 10\%}$ is the proportion of publications that belong to the top 10% most-cited ones.

Given the advantages of the percentiles and related $PP_{top\ 10\%}$ the Leiden Ranking and SCImago Institutions Rankings have already incorporated these metrics in the global rankings of academic and research institutions.

## References


1 Adam D. The countinghouse. *Nature* 2002;415(6873):726–729. doi:10.1038/415726a.

2 Bornmann L, Marx W, Gasparyan AY, Kitas GD. Diversity, value and limitations of the journal impact factor and alternative metrics. *Rheumatology International* 2012;32(7):1861–1867. doi: 10.1007/s00296-011-2276-1.

3 Van Noorden R. Metrics: A profusion of measures. *Nature* 2010;465(7300):864–866. doi: 10.1038/465864a.

4 San Francisco Declaration on Research Integrity. Available at www.ascb.org/SFdeclaration.html [Accessed 1 June 2013].

5 Thomson Reuters Statement Regarding the San Francisco Declaration on Research Assessment.Available at http://researchanalytics.thomsonreuters.com/statement_re_sfdra/ [Accessed 1 June 2013].

6 Vinkler P. The evaluation of research by scientometric indicators. Oxford, UK: Chandos Publishing, 2010.

7 Schubert A, Braun T. Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics* 1986;9(5-6):281–291.

8 Bornmann L, Marx W. How good is research really? Measuring the citation impact of publications with percentiles increases correct assessments and fair comparisons. *EMBO Reports* 2013;14(3):226–230. doi: 10.1038/embor.2013.9.

9 Bornmann L. How to analyze percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes, and top-cited papers. *Journal of the American Society for Information Science and Technology* 2013;64(3):587–595. doi: 10.1002/asi.22792.

10 Bornmann L. A better alternative to the h index. *Journal of Informetrics* 2013;7(1):100. doi: 10.1016/j.joi.2012.09.004.