



Comments on the Use of the Journal Impact Factor for Assessing the Research Contributions of Individual Authors

Alexander I. Pudovkin*

Institute of Marine Biology, Russian Academy of Sciences, Vladivostok, Russia

Keywords: Eugene Garfield, science citation index, bibliometric indicators, journal impact factor, research performance

The journal impact factor (JIF) is probably the best known invention of Eugene Garfield. Originally the JIF was introduced at the Institute for Scientific Information in early 1960s as an “in-house” index to assist in deciding whether to cover a journal in the *Science Citation Index* (Garfield and Sher, 1963; Garfield, 2006). The JIF, a measure of the short-term average citation rate of papers published in a journal, permitted Garfield to make a fair comparison of journals that publish hundreds of papers a year with “mastodons,” like the *Journal of Biological Chemistry*, which publish thousands of papers each year. Only later did the JIF acquire its present importance for librarians, journal publishers, and scientists. Along with the popularity of the JIF has come criticism. The literature critical of the JIF and its uses is huge and cannot be considered here, but a few general themes are worth attention.

It is often pointed out that the distribution of citations to papers in a journal is highly skewed, so an arithmetic average, used in the JIF calculation, is a poor representation of central tendency or typical performance: a few highly cited papers may significantly shift the JIF score upward from the median value. Other critics note that unscrupulous editors may manipulate the types of articles a journal publishes to game the JIF calculation, which counts all citations to a journal’s content but divides this number by citable items only (articles and reviews). Highly cited editorials, for example, would raise the JIF. There are also many reports of journal editors requiring authors to add citations to recent papers in the journal and thereby increase the JIF artificially. All three are valid complaints, but there are reasonable rebuttals to each. Although the distribution of citations to papers published in a journal is indeed skewed, there is a very high correlation between the JIF and the citation score of the median paper (Anonymous, 2013; Garfield and Pudovkin, 2015). This means that the JIF reflects not the citedness of a few top cited papers, but rather the citedness of the bulk of the journal’s papers. The citation skewness itself does not seem an important obstacle for using JIF in assessment of papers as noted by Ludo Waltman and Vincent Traag of CWTS, Leiden University, in theoretical terms using a variety of arguments in an as-yet unpublished article (Waltman and Traag, 2017).

Manipulations of the numerator/denominator numbers have been documented only occasionally and for only a few journals. Also, Clarivate Analytics monitors JIF data and suspends journals exhibiting “anomalous” patterns in their citation profiles (Anonymous, 2017). In general, JIF values are quite stable through time. Also, the traditional 2-year and the 5-year JIFs are quite similar.

The most intense criticism of the JIF is reserved for its use to characterize the research impact of individual papers and individual scientists (Seglen, 1989, 1992, 1994, 1997; DORA Declaration, 2012; Hicks et al., 2015; Wilsdon et al., 2015; Zhang et al., 2017). It is repeatedly emphasized that it is a great mistake to consider the JIF value as a proxy for citedness of individual papers in a journal. From this, it is argued that it is wrong to consider the JIF as a proxy for the influence of an author. Well, the JIF is certainly not a proxy for the citation score of a paper. But the skewness of the citation distribution is not the issue here (see Waltman and Traag, 2017). If the distribution would have been

OPEN ACCESS

Edited by:

David A. Pendlebury,
Clarivate Analytics, United States

Reviewed by:

Henk F. Moed,
Sapienza Università di Roma, Italy
Vincent Traag,
Leiden University, Netherlands

*Correspondence:

Alexander I. Pudovkin
aipud@mail.ru

Received: 13 December 2017

Accepted: 16 January 2018

Published: 02 February 2018

Citation:

Pudovkin AI (2018) Comments on the Use of the Journal Impact Factor for Assessing the Research Contributions of Individual Authors. *Front. Res. Metr. Anal.* 3:2. doi: 10.3389/frma.2018.00002

symmetrical or even normal, the citation scores of individual papers would differ from the JIF value simply because of the intrajournal variation of paper citedness: about half of the papers would be more cited than would be expected from the JIF value and the other half would be less cited. For instance, the number of citations for the document type “articles” (regular research reports) published in *Nature* in 2010 (by November 2017) range from 18 to 3,240, the third and first quartiles being 96 and 326 [data from Web of Science (WoS)]. Thus, the citation frequency of a paper may be and mostly is quite different from the average citedness.

However, the fact that a paper is published in a high-impact journal offers some evidence of its putative quality and potential importance. Similar opinion is expressed by Kurtz and Henneken (2017). Take the case of a relatively poorly cited paper appearing in *Science* or *Nature*. Such a paper was evaluated by at least two reviewers, who are specialists, in a process overseen by a knowledgeable editor. If the paper was approved for publication, it was judged for quality and importance by at least three informed, critical persons. For these high-impact journals, the ratio of submitted/accepted manuscripts is very high. This results in strong competition among the submitted papers, which allows these journals to exercise requirements of higher scientific merit of the manuscripts and the quality of presentation. The sieve of strict reviewing and editing by high-impact journals aims to ensure that papers published are of very high quality. Few will doubt that selection of a paper for publication in *Nature* or *Science* is *prima facie* evidence of quality regardless of its eventual actual citedness.

Garfield was ambivalent about using the JIF as an indicator of the impact of individual papers or authors. He wrote: “It would be more relevant to use the actual impact (citation frequency) of individual papers in evaluating the work of individual scientists rather than using the journal impact factor as a surrogate. The latter practice is fraught with difficulties, as Seglen and others have pointed out” (Garfield, 2001). Privately he told me that JIF values of papers published by an author may, however, be representative of an author’s professional standing. This indicator should not be the only measure considered, but rather it should be combined with other bibliometric indices (total number of citations, Hirsh index, etc.), and these always in combination with the expert opinion of peers. The JIF may be especially useful in consideration of recent publications, which have not yet had much time to accumulate citations (Abramo et al., 2010). Bornmann and I expressed this opinion (Bornmann and Pudovkin, 2017), and this was fully supported by Garfield, who reviewed our manuscript.

I would like to consider the well-known papers by Seglen (1989, 1992, 1994, 1997) and a recent review of Seglen’s conclusions using a much larger data set (Zhang et al., 2017). The conclusion of the recent paper is “None of our findings are contrary to the understanding that JIFs should not be used as performance measures of individual researchers and their publications.” However, I think that this conclusion is not supported enough by the actual data presented. In Figure 13 in the paper by Seglen (1994), reproduced as Figure 1 in the paper by Zhang et al., one can see that for highly cited authors, there is a strong correlation between the mean article citedness and the impact factor of the journals in which they were published. For less cited

authors, there is a weaker correlation in the lower range of the JIF (0.5–4.0). For higher values of JIF (4.5–8.0), the citedness does not grow with an increase of JIF. A similar though more complex pattern is presented in the study by Zhang et al., in which authors interpret the data as evidence of poor correlation between paper citedness and the JIF.

I suggest another interpretation. In his 1994 paper, Seglen studied the performance of 16 researchers from biomedical fields (a very heterogeneous category). Some subfields of biomedicine, like molecular genetics or immunology, are characterized with high citation intensity or density. In other subfields, like general biology or taxonomy, citation intensity is much less. For highly cited authors, who probably work in the subfields of high citation intensity, there is a strong correlation mentioned above between the paper citedness and the JIF (although not quite linear). The less cited authors probably worked in fields exhibiting less citation density. Thus, their most cited papers, even when published in high-impact factor journals are cited with lesser frequency, one more characteristic for their specialty subfield. Hence, their citedness reaches a plateau, clearly seen in the paper by Seglen (1994). The same interpretation may be valid for the recent paper revisiting his conclusions (Zhang et al., 2017). Spearman correlation between paper citedness and the JIF for the pooled data in the latter paper is 0.55, which is not negligibly small.

As stated above, the main argument against the use of the JIF for characterization of individual papers (and scientists) is a lack of strong correlation between the JIF and citedness of the individual papers published in them. The authors expressing this view implicitly believe that citedness of the papers is a measure of its scientific merit. However, many would argue that the best available criterion of the merit of a research paper is the opinions of peers. In this context, it is interesting to note that there is no strong correlation between citedness of papers and their appraisal by peers: the correlation is quite modest, 0.45 (Bornmann and Leydesdorff, 2013), not much different from the correlation between citedness and IF just mentioned (Seglen, 1994; Zhang et al., 2017).

I argue that the impact factor of journals in which a paper was published *can* be used as an indication of the paper’s scientific merit, along with other bibliometric indicators, raw or normalized. Prognostic value of the normalized JIFs in differentiation of poor and high performers is shown by Bornmann and Williams (2017). The importance of normalization of JIFs was specifically discussed in the study by Pudovkin and Garfield (2004). For instance, the median JIF in the WoS specialty category “Agronomy” is only one quarter of that in “Biochemistry & Molecular Biology.” That is why Garfield and I suggested rank normalization to compare journals across subject categories. Percentile normalization of JIFs was implemented in the JCR in 2015. Thus, considering scientists of different specialties, one should use both the original (“raw”) JIF and the normalized JIF percentile.

My own 30-year experience as a member of committees for recruitment of new research staff and for regular reviews at the Institute of Marine Biology, Russian Academy of Sciences is relevant. I know that committee members usually only look through the list of publications and rarely even read the titles of

papers; however they do pay attention to the journals in which the papers of the applicant were published. Thus, the judgment of the committee members is based mostly on the journal titles. The impact factor of the journals (“raw” and percentile) would be much more informative. I believe such a situation is quite common for the developing countries like China, India, and throughout Latin America and Eastern Europe. The mere fact that a paper is published in a journal covered by WoS or Scopus is considered in some countries a substantial evidence of the quality of the paper, regardless of the impact factor of the journal or the actual citedness of the paper.

I believe that the main role of bibliometric indicators is in screening out poor performers rather than identifying and discriminating among “champions.” Moed (2017) has recently expressed a similar view in recommending the use of citation data for determining “minimum performance standards.” To gauge differences in excellence using the indicators is unreasonable for many reasons. I will list only a few:

- (1) Number of authors on a paper. A team of authors can produce more papers than a single author, so a member of a research team may easily co-author many papers a year, while a single researcher will produce fewer. Using fractional credit with the weight of $1/n$ (n being the number of authors) seems a poor solution: it makes participation in multiauthor research “bibliometrically” less rewarding. Possibly a better weight would be $1/\sqrt{n}$. In any case, explicit statements of each author’s contributions would be required to address author credit for papers and citations in a precise way (Sauermaun and Haeussler, 2017).
- (2) Difference in citation density in various fields and subfields. A key problem is the delineation of fields and subfields to obtain valid baseline measures to assess relative performance (Waltman and van Eck, 2013; Ruiz-Castillo and Waltman, 2015; Ioannidis et al., 2016).
- (3) Citations differ in importance. The weight of a citation mentioning one paper among many others in a literature review is not large. However, if a paper is specifically cited as a reason for performing the research presented, then the weight of the citation is significant. Citation context, including analysis of sentiment in the text of the citing passage (a “cintance”),

has long been desired but only recently explored, thanks to greater availability of full text in a digital form (Moravcsik and Murugesan, 1975; Ding et al., 2014).

So, gauging the differences in performance among highly productive authors is difficult because of the imperfections and limitations of citation-based indicators, including JIFs. However, identifying poor performers is a relatively simple task: they do not publish in high-impact journals, and their papers are poorly (if at all) cited.

Bibliometric theoreticians seem to operate very far from the actual problems of those managing research institutions and conducting research assessments. Simple appeals not to use the JIF in the evaluation of individual scientists are ignored by those managing the merit rating procedures. JIFs are used as proxy measures worldwide. What is the alternative? It is as Garfield said: to use instead actual citedness of authors and their papers. That is reasonable, even to be encouraged, but it is much more time consuming. Besides, as I argue, citedness is not the ultimate evidence of the quality of a paper (Bornmann and Leydesdorff, 2013), especially for recent publications. To base a merit rating on the opinion of peer reviewers would be preferable, but it is not practical for regular assessment procedures since, generally, no enough peers with the necessary expertise are available when needed and there is usually no enough time to devote to evaluation. Summing up my opinion, it is reasonable and justified to use JIFs for merit rating of individual scientists; this is an easy and quick procedure. For a thorough analysis of individual performance, it would be preferable to use a set of bibliometric indices (total citation number, Hirsh index, JIFs) and peers’ reviews.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

ACKNOWLEDGMENTS

The author appreciates help and valuable advice of Mr. David Pebblebury (Claricate Analytics).

REFERENCES

- Abramo, G., D’Angelo, C. A., and Di Costa, F. (2010). Citations versus journal impact factor as proxy of quality: could the latter ever be preferable? *Scientometrics* 84, 821–833. doi:10.1007/s11192-010-0200-1
- Anonymous. (2013). Beware the impact factor. *Nat. Mater.* 12, 89. doi:10.1038/nmat3566
- Anonymous. (2017). *Title Suppression from Journal Citation Reports*. Available at: <http://wokinfo.com/media/pdf/jcr-suppression.pdf>
- Bornmann, L., and Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: a comparative study using data from InCites and F1000. *J. Informetr.* 7, 286–291. doi:10.1016/j.joi.2012.12.003
- Bornmann, L., and Pudovkin, A. I. (2017). The journal impact factor should not be discarded. *J. Korean Med. Sci.* 32, 180–182. doi:10.3346/jkms.2017.32.2.180
- Bornmann, L., and Williams, R. (2017). Can the journal impact factor be used as a criterion for the selection of junior researchers? A large-scale empirical study based on ResearcherID data. *J. Informetr.* 11, 788–799. doi:10.1016/j.joi.2017.06.001
- Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X. L., and Zhai, C. X. (2014). Content-based citation analysis: the next generation of citation analysis. *J. Assoc. Inform. Sci. Technol.* 65, 1820–1833. doi:10.1002/asi.23256
- DORA Declaration. (2012). *The San Francisco Declaration on Research Assessment*. Available at: <http://www.ascb.org/dora/>
- Garfield, E. (2001). Impact factors, and why they won’t go away. *Nature* 411, 522. doi:10.1038/35079158
- Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA* 295, 90–93. doi:10.1001/jama.295.1.90
- Garfield, E., and Pudovkin, A. I. (2015). Journal impact factor strongly correlates with the citedness of the median journal paper. *COLLNET J. Scientometr. Inform. Manage.* 9, 5–14. doi:10.1080/09737766.2015.1027099
- Garfield, E., and Sher, I. (1963). New factors in the evaluation of scientific literature through citation indexing. *Am. Doc.* 14, 195–201. doi:10.1002/asi.5090140304

- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., and Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature* 520, 429–431. doi:10.1038/520429a
- Ioannidis, J. P. A., Boyack, K., and Wouters, P. F. (2016). Citation metrics: a primer on how (not) to normalize. *PLoS Biol.* 14:e1002542. doi:10.1371/journal.pbio.1002542
- Kurtz, M. J., and Henneken, E. A. (2017). Measuring metrics: a 40-year longitudinal cross-validation of citations, downloads, and peer review in astrophysics. *J. Assoc. Inform. Sci. Technol.* 68, 695–708. doi:10.1002/asi.23689
- Moed, H. F. (2017). *Applied Evaluative Informetrics*. Cham: Springer.
- Moravcsik, M. J., and Murugesan, P. (1975). Some result on the function and quality of citations. *Soc. Stud. Sci.* 5, 86–92. doi:10.1177/030631277500500106
- Pudovkin, A. I., and Garfield, E. (2004). Rank-normalized impact factor: a way to compare journal performance across subject categories. *Proc. Assoc. Inform. Sci. Technol.* 41, 507–515. doi:10.1002/meet.1450410159
- Ruiz-Castillo, J., and Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *J. Informetr.* 9, 102–117. doi:10.1016/j.joi.2014.11.010
- Sauerermann, H., and Haeussler, C. (2017). Authorship and contribution disclosures. *Sci. Adv.* 3, e1700404. doi:10.1126/sciadv.1700404
- Seglen, P. O. (1989). From bad to worse: evaluation by journal impact. *Trends Biochem. Sci.* 14, 326–327. doi:10.1016/0968-0004(89)90163-1
- Seglen, P. O. (1992). The skewness of science. *J. Am. Soc. Inform. Sci.* 43, 628–638. doi:10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASI5>3.0.CO;2-0
- Seglen, P. O. (1994). Causal relationship between article citedness and journal impact. *J. Am. Soc. Inform. Sci.* 45, 1–11. doi:10.1002/(SICI)1097-4571(199401)45:1<1::AID-ASI1>3.0.CO;2-Y
- Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *Br. Med. J.* 314, 498–502. doi:10.1136/bmj.314.7079.497
- Waltman, L., and Traag, V. A. (2017). *Use of the Journal Impact Factor for Assessing Individual Articles Need Not Be Wrong*. arXiv:1703.02334. Available at: <https://arxiv.org/ftp/arxiv/papers/1703/1703.02334.pdf>; <https://www.cwts.nl/blog?article=n-q2z254&title=use-of-the-journal-impact-factor-for-assessing-individual-articles-need-not-be-wrong>
- Waltman, L., and van Eck, N. J. (2013). A systematic empirical comparison of different approaches for normalizing citation impact indicators. *J. Informetr.* 7, 833–849. doi:10.1016/j.joi.2013.08.002
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., et al. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. Available at: <http://www.hefce.ac.uk/pubs/rereports/Year/2015/metrictide/Title,104463,en.html>
- Zhang, L., Rousseau, R., and Sivertsen, G. (2017). Science deserves to be judged by its contents, not by its wrapping: revisiting Seglen's work on journal impact and research evaluation. *PLoS ONE* 12:e0174205. doi:10.1371/journal.pone.0174205

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Pudovkin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.