

Why Do P&T Committees Keep Hiring the Journal Impact Factor?

David W. Lewis
June 2019

© 2019 David W. Lewis. This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

The way the title of this article asks a question that might seem odd. P&T committees along with others seeking to gage academic merit and accomplishment use the journal impact factors (JIF), but we don't usually think of them as hiring the JIF. Asking the question this way is comes from Clayton Christensen's notion of the "job to be done".¹ Christensen contends that focusing on what job a person, in this case the members of a committee, is trying to accomplish and what products or services is hired to do that job provides insights that would otherwise not be discovered. So, what is the job that P&T committee members have that they hired the JIF to do. The answer is at one level obvious – they need to evaluate faculty from many disciplines whose expertise they often are not qualified to judge. As Per O. Seglen, puts it, "Committees tend, therefore, to resort to secondary criteria like crude publication counts, journal prestige, the reputation of authors and institutions, and estimated importance and relevance of the research field, making peer review as much of a lottery as of a rational process."² When push comes to shove, they turn to the JIF.

The JIF provides a clear and easy to get measure and does the job do be done. If the faculty member being evaluated publishes in journals with high JIFs they are worthy. If their publications are in journals with lower JIFs they are not as qualified. This is probably though to simple. The job is a little more complex. The job is not just to evaluate faculty based on the publications, but to do so when many of the publications are recent and other measures of article quality, such as the number of citations, are not available. This is the beauty of the JIF. It can be used to judge recently published articles. The other ways to doing this job involves reading and evaluating the work themselves or asking outside experts to provide an opinion.

The first is very time consuming and rarely does a P&T committee member feel qualified to make these judgements. The second alternative relies on trusting unknown experts. Given these alternative using the JIF seems expedient, even if you clearly understand its flaws.

And the JIF has no end of flaws. Let's start by reviewing how the JIF works. The JIF is calculated as follows:

In any given year, the impact factor of a journal is the number of citations, received in that year, of articles published in that journal during the two preceding years, divided by the total number of "citable items" published in that journal during the two preceding years.³

The calculation is usually based on the *Web of Science* database, though Elsevier's *SCOPUS* database recently offer an alternative, but similar, ranking. The JIF is usually calculated to the third decimal point, a level of precision that is entirely unjustified, just conveys a sense of exactitude that is comforting. So, for example, the JIF for *Nature* in 2017 was 41.577 and for *PLOS One* it was 2.766. It seems clear *Nature* is a much better journal than *PLOS One* and so an article pushed in *Nature* is worth more. Case closed. Job done. The fact that the JIF as critics and that it is not a perfect metric is understood, but so it goes.

The criticisms of the JIF are many. It has clearly been demonstrated and it is widely recognized that the JIF is deeply flawed. Per O. Seglen has a nice list of these flaws:

- Journal impact factors are not statistically representative of individual journal articles
- Journal impact factors correlate poorly with actual citations of individual articles
- Authors use many criteria other than impact when submitting to journals
- Citations to "non-citable" items are erroneously included in the database
- Self citations are not corrected for
- Review articles are heavily cited and inflate the impact factor of journals
- Long articles collect many citations and give high journal impact factors
- Short publication lag allows many short term journal self-citations and gives a high journal impact factor
- Citations in the national language of the journal are preferred by the journal's authors
- Selective journal self-citation: articles tend to preferentially cite other articles in the same journal
- Coverage of the database is not complete
- Books are not included in the database as a source for citations

- Database has an English language bias
- Database is dominated by American publications
- Journal set in database may vary from year to year
- Impact factor is a function of the number of references per article in the research field
- Research fields with literature that rapidly becomes obsolete are favoured
- Impact factor depends on dynamics (expansion or contraction) of the research field
- Small research fields tend to lack journals with high impact
- Relations between fields (clinical v basic research, for example) strongly determine the journal impact factor
- Citation rate of article determines journal impact, but not vice versa⁴

All of these flaws are well documented, but let's just take the first. Seglen looked at distribution of the citations to articles and how these articles contributed to the citations for the journal and found:

The cumulative curve shows that the most cited 15% of the articles account for 50% of the citations, and the most cited 50% of the articles account for 90% of the citations. In other words, the most cited half of the articles are cited, on average, 10 times as often as the least cited half. Assigning the same score (the journal impact factor) to all articles masks this tremendous difference—which is the exact opposite of what an evaluation is meant to achieve. Even the uncited articles are then given full credit for the impact of the few highly cited articles that predominantly determine the value of the journal impact factor.⁵

Seglen's article is over 20 years old and many others have documented the problems he enumerated. Beyond the concerns expressed by Seglen there are concerns about unscrupulous manipulation of JIFs.⁶ So we have not only a deeply flawed measure, but the likelihood of at least occasional fraud.

So why do P&T committees keep hiring the JIF? Because no other exiting solution to the job they need to do – evaluate the quality of faculty work, especially recent work – as easily and quick as the JIF.

If we want to change this behavior, we need to find a way to get the evaluation job done that is as easy and quick as hiring the JIF. It also needs to have an aura of authority, like the one provided by the JIF's three decimal place. Currently there are no alternatives that can compete against the JIF. The alternatives are complex, hard to explain, or time consuming to use, so they are unlikely to be hired.

I would offer two possible approach to creating alternatives to the JIF that could be offered to P&T committees.

The first option would be to develop a composite based on a number of metrics. The JIF might be one of these measure but it would not stand alone. The *Metrics Toolkit* developed by Robin Champieux, Heathr Coates and Stacy Konkiel is an excellent source for alternative metrics.⁷ It cites 16 metrics for journal articles, for example, the Altmetric Attention Score, article downloads, blog or Facebook mentions, the Publons Score, PeerPub Comments, and Wikipedia citations. JIF makes the list as measure for journals. Though the Metrics Toolkit provides good guidance for faculty members whose job to be is preparing a strong case for promotion and tenure. Using it to do the P&T committee job to be done is likely a non-starter. It has too many alternatives and not definitive judgements. Using some combination of the measures from the Metrics Toolkit to create a composite score for journal articles. For example, a three part measure might include the normalize Altmetric Attention Score, normalize downloads for the past year or two, and the total number of blog or Facebook mentions, the Publons Score, PeerPub Comments, and Wikipedia citations all added together and normalized in some way. I would expect that a measure of this sort would need to be developed on a campus, school, or department level, and that it would be calibrated over time. The measure needs to be simple and easy to access. It might include three or four numbers, but not 16, though combining different measures into a single number is likely OK. Comparing such a composite with the JIF would provide an opportunity to determine if its flaws have been misleading in the past. Such a measure, if it is quick and easy to use, has a good chance of getting hired by P&T committees over the JIF.

The second alternative is to create an article level metric that is capable of determining the likelihood of a recently published article becoming important later. Fortunately, such a system exists, though it is not be used for this purpose. Meta is a machine learning system that has ingested all of the bio-medical literature on its way to being a new type of search engine. The developers claim that, "Large-scale trials conducted by Meta in partnership with industry demonstrated that Bibliometric Intelligence out-performed tens of thousands of human editors by a factor 2.5x at predicting article-level impact for new manuscripts, prior to publication. It also performed 2.2x better than the same group of editors at identifying "superstar articles" - those that represent the top 1% of high-impact papers, prior to publication."⁸ Meta, or other machine learnings systems like it, should be able to provide a standard measure of an article's future impact. It wouldn't need have the precision of three decimals, rather a measure that predicts whether the article will be

in the top 10%, and then a prediction on which quarter of the universe it is likely to end up in. An article level metric of this sort would likely be hired over the JIF for the job that P&T committees need to have done.

To conclude, P&T committees will continue to hire the JIF for the job of evaluating faculty work. If a more accurate measure is developed that is as easy and quick to use as the JIF, then it is like to be hired. Other approaches are likely to remain unemployed.

Notes

¹ The Jobs to Be Done theory is covered at length in his 2016 book, Clayton M. Christensen, Taddy Hall, Karen Dillon, and David S. Duncan, *Competing Against Chance: The Story of Innovation and Customer Choice*, New York, NY: Harper Business, 2016. A shorter version can be found in Clayton M. Christensen, Taddy Hall, Karen Dillon, and David S. Duncan, "Know Your Customer's 'Jobs to Be Done'," *Harvard Business Review*, 94(9): 54-60 September 2016. The classic example of the job to be done concerns milkshakes and makes the point very effectively see: Carmen Nobel, "Clay Christensen's Milkshake Marketing," *Harvard Business School Working Knowledge*, February 11, 2011, <https://hbswk.hbs.edu/item/clay-christensens-milkshake-marketing>

² Per O Seglen. Why the Impact Factor of Journals Should Not Be Used for Evaluating Research," *BMJ* 314: 498-502 February 15, 1997, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2126010/pdf/9056804.pdf>

³ "Impact Factor," *Wikipedia*, https://en.wikipedia.org/wiki/Impact_factor (accessed June 25, 2019).

⁴ Per O. Seglen. Why the Impact Factor of Journals Should Not Be Used for Evaluating Research," *BMJ* 314: 499 February 15, 1997, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2126010/pdf/9056804.pdf>

⁵ Per O. Seglen. Why the Impact Factor of Journals Should Not Be Used for Evaluating Research," *BMJ* 314: 499 February 15, 1997, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2126010/pdf/9056804.pdf>

⁶ See for example: Ben R Martin, "Editors' JIF-Boosting Stratagems - Which Are Appropriate and Which Not?" *Research Policy* 45(1): 1-7 February 2016, DOI: <https://doi.org/10.1016/j.respol.2015.09.001>; Phil Davis, "How Much Citation Manipulation Is Acceptable?" *Scholarly Kitchen*, May 30, 2017, <https://scholarlykitchen.sspnet.org/2017/05/30/how-much-citation-manipulation-is-acceptable/>; Jerry A. Jacobs, "Manipulated Journal Rankings?" *Inside Higher Ed*, July 1, 2016, <https://www.insidehighered.com/views/2016/07/01/examination-whether-academic-journal-rankings-are-being-manipulated-essay>, and Allen W. White and Eric A. Fong, "Coercive Citation in Academic Publishing," *Science* 335(6069): 542-543 February 3, 2012, DOI: <https://doi.org/10.1126/science.1212540>

⁷ *Metrics Toolkit*, developed by Robin Champieux, Heathr Coates and Stacy Konkiel, <https://www.metrics-toolkit.org>

⁸ Aries Marketing, "Artificial Intelligence Integration Allows Publishers a First Look at Meta Bibliometric Intelligence," October 17, 2016, <https://www.ariessys.com/views-press/press-releases/artificial-intelligence-integration-allows-publishers-first-look-meta-bibliometric-intelligence/> See also: Liu Yang, Shankar Vembu, Amr Adawi and Ofer Shai, *Bibliometric Intelligence: Enabling Editors Through Machine Learning*, Meta Whitepaper, November 17, 2016.