

Bibliometric indicators: opportunities and limits

Christopher W. Belter

See end of article for author's affiliation

DOI: <http://dx.doi.org/10.3163/1536-5050.103.4.014>

Keywords: Bibliometrics, Program Evaluation, Research Personnel/Statistics and Numerical Data, Research/Organization and Administration, Peer Review, Research

Evaluating scientific research has always been difficult. The peer-review process, which has been the mainstay of science evaluation for nearly a century, takes time, expertise, and no small amount of resources to do properly. But several trends in scientific research have made this process even more challenging. The sheer number of scientific publications produced per year has been growing at an exponential rate for over fifty years [1, 2] and has shown no sign of slowing down anytime soon. These publications are also growing increasingly technical and specialized, making qualified reviewers more and more difficult to find. Finally, the glut of researchers in the biomedical pipeline combined with the recent recession have resulted in a larger number of researchers competing for a shrinking pool of available research funds [3]. Evaluating scientific research in this context is becoming not only increasingly difficult, but also increasingly important to ensure that the right researchers receive promotions and funding to continue their work.

In this environment, a number of review boards, institutions, and even countries are turning to bibliometrics to facilitate the review process. Bibliometrics is the quantitative analysis of publications. It essentially extracts data from publications and analyzes that data in various ways to answer questions about the research that those publications represent. It is a method of studying the producers, processes, and evolution of research using research publications as a proxy for research. As such, the field of bibliometrics encompasses a wide variety of approaches and methods, but it has become best known for its attempts to measure the impact of scientific research through the use of various bibliometric indicators like the impact factor [4] and the H-index [5]. Because these indicators are perceived to be more objective than peer review, because they can be calculated with far less time and effort than peer review, and because there is some evidence that these indicators agree with peer judgment, reviewers and policy makers are increasingly using these

indicators in addition to, and in some cases in place of, peer review of research impact.

Although the use of bibliometric indicators can provide a valuable supplement to the peer-review process, these indicators are all too often taken out of context and applied without a full understanding of the bibliometric research on which they are based. As a result, they are frequently used to measure things that they were not intended to measure or to make comparisons they are not actually capable of making. This article provides a short introduction to the basic ideas behind these indicators and discusses ways that they can be used responsibly to minimize the biases of peer review. For more extensive and technical overviews of this topic, see Haustein [6] and Mingers [7].

WHAT BIBLIOMETRIC INDICATORS CAN AND CANNOT MEASURE

All bibliometric indicators build on the idea that we can measure the impact of a paper by counting the number of other papers that have cited it. Citations, the theory goes, act as a vote of confidence or a mark of influence from one paper to another. The fact that one paper cites another is an indication that the cited paper has had some influence, or impact, on the paper citing it. Counting the number of citations received by a paper, then, allows us to measure the impact that paper has had on science as a whole.

Extrapolating outward from there, counting citations to a set of papers—by a single author, institution, or even an entire country—allows us to measure the impact that set of papers has had on scientific research. More citations equal more impact.

The problem with this idea is that acknowledging influence is only one of the many reasons that authors cite other papers [8]. Authors cite other papers for all kinds of reasons: to refer to a particular methodology, to point out examples of other work done on the same topic, to reinforce a point they

make in the text, to give credit to their mentors or experts in the field, or even to discuss examples of flawed methods or misleading results. Current bibliometric indicators cannot account for this variety; they count all citations equally, regardless of the actual reason for the citation. As a result, we cannot say for certain that a highly cited paper is actually highly influential. What we probably can say, following the lead of Eugene Garfield, AHIP, FMLA, one of the founders of bibliometrics, is that highly cited papers are highly useful to authors for writing other papers [9]. What those papers are useful for, however, is not clear.

This means that citation counts measure a very specific definition of “impact.” Citation counts only measure the impact, or usefulness, of papers to the authors of other papers; they do not measure the impact of those papers on anything else. There is no way to tell purely from a paper’s citation count whether the paper reported a breakthrough in biomedical understanding, an advance in clinical practice that significantly improved patient outcomes, a particularly useful method of analyzing data, or a timely survey of the existing literature.

The number of citations received by a paper cannot measure whether or not the research reported by that paper improved people’s health. It can only measure whether or not the paper was useful to other authors for writing their own papers. This is a form of impact, to be sure, but not necessarily the one that reviewers think they measure.

COMMON MISTAKES

In addition to the confusion about what citations measure, there is also confusion about how bibliometric indicators actually work, leading evaluators to make mistakes when using them. Perhaps the most common mistake is using a journal’s impact factor to measure the impact of an article published in that journal. It turns out that any journal’s impact factor is primarily determined by citations received by a small fraction (10%–30%) of the articles published in that journal. That is, a few papers in that journal receive an extremely high number of citations, while the vast majority of the papers receive few to no citations [10, 11]. There is no way to tell if a paper published in a journal, even a journal like *Nature* or *Science*, will actually be highly cited. In fact, many papers published in high-impact journals like *JAMA* receive fewer citations than articles in many low-impact journals. To be clear, the impact factor may be a valid measure of a journal’s

citation impact; it is not, however, a valid measure of an article’s citation impact.

Another common mistake is to compare the values of common bibliometric indicators like the impact factor or an author’s H-index across disciplines. It turns out that there are substantial differences in the number of citations available among disciplines. There are simply more publications, and more citations, in a discipline like molecular biology than in a discipline like nursing. As a result, papers in molecular biology tend to be more highly cited than papers in nursing. This, in turn, means that the average impact factor or H-index in molecular biology is higher than the same indicator in nursing. The same is true for other disciplines: Citation counts and bibliometric indicators mean different things in different disciplines. An impact factor of 2.43 might be extremely high in one discipline but relatively low in another. As a result, comparing most bibliometric indicators across disciplines is like comparing apples to oranges: It is essentially meaningless.

A third common mistake that evaluators make is that they fail to take time into account. Citations not only take time to accumulate, but they also continue to accumulate over time. Studies have shown that papers need at least two to three years after publication to accumulate enough citations for bibliometric indicators to be reliable [12, 13]. This means that the most recent papers included in any institutional evaluation using bibliometric indicators should be at least two years old. On the other hand, citations continue to accumulate after this initial time period, meaning that older papers tend to be more highly cited than younger ones because they have had more time to accumulate citations. As a result, any list of the most highly cited papers published in a discipline or journal over a given time period will be biased toward older papers. This also means that bibliometric indicators for individual authors will always be biased toward older authors. This is because younger authors do not have as many publications as older authors and because their publications have not had as much time to be cited.

OPPORTUNITIES

So, with all of the limitations and mistakes associated with bibliometric indicators, why would anyone use them? The short answer is because the gold standard, peer review, has its own problems. In addition to the time and cost it takes to perform, the peer-review process rarely produces consistent or reproducible results [14]. The recommendations of

one peer-review panel may directly contradict the recommendations of another, even when they are reviewing the same grant application or paper. Reviewers are also subject to conscious and unconscious forms of bias, which can seriously affect their judgment and ultimate recommendations [15]. Finally, in evaluations of individuals or institutions with hundreds to thousands of publications, peer reviewers cannot evaluate all of them, so they may read only a handful of papers and ignore the rest. These biases mean that although peer review remains the gold standard for evaluating scientific research, it is not without its flaws.

Bibliometric indicators, when used responsibly, can reduce these biases. It just so happens that the strengths of bibliometric indicators exactly correspond to the weaknesses of peer review. Bibliometric indicators can be calculated for an entire publication set and represent the collective judgment of a broad segment of the scientific research community, rather than that of the selected individuals chosen for the review panel. Bibliometric indicators can also be more transparent and reproducible than peer review. They can also help guide the review process by pointing out irregularities in the publication or citation records that the reviewers might wish to focus on during their review. As a result, the bibliometrics research community recommends that bibliometric indicators be used as a complement to, not a replacement for, informed peer review when evaluating scientific research [16, 17]. Each method balances the weaknesses of the other.

Combining bibliometric indicators and peer review results in more fair, balanced, and accurate assessments of scientific research. With nothing less than the future of scientific research at stake in these assessments, it is vital that we get them right. Using all of the tools available to us, and using them properly, seems like the least we can do.

REFERENCES

- de Solla Price DJ. Quantitative measures of the development of science. *Archives Internationales d'Histoire des Sciences*. 1951;4(14):85–93.
- de Solla Price DJ. *Little science, big science*. New York, NY: Columbia University Press; 1963.
- Alberts B, Kirschner MW, Tilghman S, Varmus H. Rescuing US biomedical research from its systemic flaws. *Proc Natl Acad Sci USA*. 2014 Apr 22;111(16):5773–7. DOI: <http://dx.doi.org/10.1073/pnas.1404402111>.
- Garfield E. The history and meaning of the journal impact factor. *JAMA*. 2006 Jan 4;295(1):90–3. DOI: <http://dx.doi.org/10.1001/jama.295.1.90>.
- Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA*. 2005 Nov 15;102(46):16569–72. DOI: <http://dx.doi.org/10.1073/pnas.0507655102>.
- Haustein S, Larivière V. The use of bibliometrics for assessing research: possibilities, limitations and adverse effects. In: Welpel IM, Wollersheim J, Ringelhan S, Osterloh M, eds. *Incentives and performance*. Springer International Publishing; 2015. p. 121–39. DOI: http://dx.doi.org/10.1007/978-3-319-09785-5_8.
- Mingers J, Leydesdorff L. A review of theory and practice in scientometrics. arXiv: 1501.05462v2. 2015.
- Bornmann L, Daniel HD. What do citation counts measure? a review of studies on citing behavior. *J Doc*. 2008;64(1):45–80. DOI: <http://dx.doi.org/10.1108/00220410810844150>.
- Garfield E. Is citation analysis a legitimate evaluation tool? *Scientometrics*. 1979 May;1(4):359–75. DOI: <http://dx.doi.org/10.1007/bf02019306>.
- Seglen PO. Why the impact factor of journals should not be used for evaluating research. *BMJ*. 1997 Feb 15;314(7079):498–502. DOI: <http://dx.doi.org/10.1136/bmj.314.7079.497>.
- Redner S. Citation statistics from 110 years of *Physical Review*. *Phys Today*. 2005 Jun;58(6):49–54.
- Abramo G, Cicero T, D'Angelo CA. Assessing the varying level of impact measurement accuracy as a function of the citation window length. *J Informetrics*. 2011 Oct;5(4):659–67. DOI: <http://dx.doi.org/10.1016/j.joi.2011.06.004>.
- Wang J. Citation time window choice for research impact evaluation. *Scientometrics*. 2013 Mar;94(3):851–72. DOI: <http://dx.doi.org/10.1007/s11192-012-0775-9>.
- Bornmann L. Scientific peer review. *Annu Rev Inf Sci Technol*. 2011;45(1):197–245. DOI: <http://dx.doi.org/10.1002/aris.2011.1440450112>.
- Lee CJ, Sugimoto CR, Zhang G, Cronin B. Bias in peer review. *J Am Soc Inf Sci Technol*. 2013 Jan;64(1):2–17. DOI: <http://dx.doi.org/10.1002/asi.22784>.
- Moed H. The future of research evaluation rests with an intelligent combination of advanced metrics and transparent peer review. *Sci Public Policy*. 2007 Oct;34(8):575–83.
- van Raan AFJ. Advanced bibliometric methods as quantitative core of peer review based evaluation and foresight exercises. *Scientometrics*. 1996 Jul/Aug;36(3):397–420.



AUTHOR'S AFFILIATION

Christopher W. Belter,
christopher.belter@nih.gov,
 NIH Library, National Institutes
 of Health, Bethesda MD