Review of: "Measuring researchers' success more fairly: going beyond the H-index"

Stephen Haines[1]

1 University of Minnesota

Review of: Measuring researchers' success more fairly: going beyond the H-index

The author takes on a subject of importance to those whose productivity may be evaluated by the H-Index. Because of its automated simplicity, the H-index is often used as part of the evaluation of academic productivity but, as the author points out, can be used for other purposes such a limiting access to applications for research funding in Italy.

The author's fundamental concern, that the H-index is a flawed measure of academic productivity, is not new and is widely acknowledged to be correct. As other reviewers have pointed out, a more detailed review of alternative measures would be useful to the reader.

The author proposes to construct a better alternative measure, but even in the abstract he enumerates the reasons that this is not just a difficult task, but an impossible one. He sets criteria that cannot be met: clear reasoning, easy math, consensus among publishers. Other important problems are the need for the measure to be automated so that administrators can use it easily, the ability to validly compare academic productivity over a wide range of disciplines (the Dean's dream) and, therefore, the dependence on existing published databases of publication, research grants and other fundamental products of academic work.

The author's proposal has several insurmountable flaws. It is based on citation counts. He mentions the problem of compliance with internationally recognized authorship criteria. Other reviewers have pointed out that there are both positive and negative citations but all count equally in the citation count and the problem of self-citation. Other issues include differing citation policies among journals and the wide variety of reasons for one article citing another.

My own most highly cited paper that doesn't represent "team science" is an historical review. So far it has relieved at least 280 (the number goes up every year) authors of having to do a detailed historical review of the topic, but I would not rate it as a great, novel contribution to science. These citations should not rate as highly as some of my less frequently cited papers.

Then there is the problem of multiple authors. Formoso does not present a useful solution to this problem. The study of many difficult scientific problems require the work of large scientific teams. We do not yet have a reasonable method for granting recognition to those who are not the team leaders of such large efforts. According to Scopus, my most highly cited work (7914 citations as of April 3, 2023) is the result of reporting by a list of contributors that had to be published as a

1 1/2 page (in small font) appendix to the main article. My work involved in that contribution was much more important and difficult than that involved in the above mentioned historical review.

The author's proposal to distribute the total number of citations of an article in a linearly declining way depending on position of authorship in the publication has no emperic evidence to support it as reasonable and it fails completely as the number of authors increases. In addition, he mentions the problem that some disciplines choose to list all but the first and last authors alphabetically, which eliminates entire disciplines from the valid use of this method.

Which brings up the problem of comparisons across disciplines. It is a simple fact that opportunities for publication differ across disciplines. There are far fewer opportunities to publish in small disciplines like neursourgery or pediatric neurology compared to large disciplines such as cardiology or infectious diseases. This disadvantage holds true even when the number of people practicing the discipline is taken into account. (See Do TH, Miller C, Low WC, Haines SJ. A Proof of Concept for Applying the Radicchi Index (hf) to Compare Academic Productivity and Scientific Impact Among Medical Specialties. Neurosurgery. 86(4):593-603, 2020 04 01. UI: 31232431 – an example of self-citation intended to decrease the length of this review.) If this problem of inappropriate weighting of numbers of citations is not taken into account, the measure will inherently disadvantage scientists in smaller disciplines and may have an even greater negative effect on those whose work is primarily interdisciplinary.

This discrepancy also negatively influences the use of the Impact Factor to weight the value of citations. Other reviewers have pointed out problems related to the use of the impact factor in efforts to evaluate productivity.

So where does all this negativity get us? It should put a damper on any effort to develop a widely generalizable single measure of scientific productivity. Multiple measures will be necessary, each of which can address a few of these problems. Recognizing that administrators will always look for easily automated ways of evaluating academic productivity, they should be provided with multiple tools, validated for specific circumstances, and trained to use them wisely, being aware of built in biases and the fact that they should only be a part of an very difficult and important evaluation process.