Evaluation of scientific impact: insights and incites

Jason H. T. Bates1 and Peter D. Wagner2
1Pulmonary/Critical Care Division, University of Vermont, Burlington, Vermont; and 2Medicine and Bioengineering, University of California, San Diego, La Jolla, California

SUCCESS IN SCIENTIFIC RESEARCH requires an unreasonable willingness to work hard under often frustrating circumstances, particularly when results are elusive and funding is scarce. The appeal of being recognized for one’s achievements can thus provide powerful motivation to keep going in the meantime. In particular, because the true benefits of scientific research can be such a long time coming, short-term assessments of research performance based on notoriety within the scientific community provide both personal reward and professional advancement. But this raises some troubling questions. How strong is the correlation between a short-term assessment of research value and the ultimate impact the research has on the progress of scientific advancement? Is there even a correlation at all?

Achievement of notoriety in science begins with the dissemination of one’s work to a wide audience. Accordingly, the coin of the realm for scientific productivity is the published journal paper. For most of the history of modern science reputations were manifest rather nebulously within the collective wisdom of the scientific community; famous scientists were somehow just known to have published great papers. The same applied to the scientific journals themselves; particular scientific communities somehow came to recognize that certain journals represented their prestige forums. The Journal of Applied Physiology is a case in point; until about the mid 1990s one could keep up with the field of, say, respiratory mechanics just by reading the Journal of Applied Physiology, perhaps supplemented with a handful of other journals.

The arrival of the internet search engine has changed all that by providing immediate access to any paper on any topic regardless of where it appears. This can be a good thing, because, for example, it reduces the chances of high quality work languishing in obscurity simply because of its author’s choice of venue. It also allows for the ready calculation of citation metrics such as the impact factor (IF), which are quantitative and objective and provide rapid feedback. Citation metrics would also seem to avoid the personal biases inherent in the opinions of whatever cabal has managed to gain control over a particular field of science. Despite these apparent advantages, however, the power of the internet to deliver citation metrics on demand has ushered in a new scientific era that is also deeply troubling, especially when these metrics are used to guide resource distribution and career advancement.

The justification for using citation metrics to judge scientific impact is founded on two key notions: 1) impact is reflected in notoriety and 2) notoriety can be captured in the number of times a paper has been cited. These notions are deeply flawed in some important ways, most notably because a paper may receive numerous citations for reasons that have little to do with its influence on the advancement of knowledge. For example, review papers provide convenient “one stop shopping” for scientists wishing to cite appropriate prior work in their own publications, but even the best reviews merely synthesize advances that were actually made by other scientists. Despite this, reviews still tend to garner more citations than do the original studies; according to Web of Science, 10 of the 20 top IFs in 2013 were assigned to journals that include the word “reviews” in their title. High citation numbers are also often associated with papers that provide reference values for quantities of wide utility. This may be useful, but it can hardly be said to constitute an important scientific advance compared with, say, the discovery of a new phenomenon or the development of a novel enabling methodology.

The lionization of citation metrics is also problematic in that they depend greatly on the number of people working in a particular field. Web of Science lists the highest 2013 IF in the field of mathematics at 3.1, whereas the highest IF in the field of immunology is 41.4, with another 62 immunology journals having higher IF values than any mathematics journal. Does this mean that immunology is 62 times more important than mathematics, or could it be that only a select minority of scientists are clever enough to do mathematical research? The very fact that we can pose this question illustrates the absurdity of the situation.

Of the various citation metrics that have been proposed to gauge scientific impact, none is as widely used nor as controversial as the IF. The problem with the IF is that it is geared toward evaluating the “hotness” of an area by taking only recent citations into account, but this area may soon go cold as further work by others shows it to be unimportant. Papers that invite contradiction by being controversial, or even just plain wrong, can also contribute disproportionately to the IF. There is thus great danger in assigning too much significance to only one metric. In 2010, the Journal of Applied Physiology ranked 11th out of 77 physiology journals in terms of IF, but 3rd in terms of citation half-life (1). This would seem to indicate that while the Journal of Applied Physiology tends not to make an immediate splash, its papers play a lasting role in the scientific enterprise. Which is more important? Is this even a question worth asking? In fact, one can argue that citation metrics have little to do with intrinsic scientific quality and may even fail to recognize it completely. For example, in what was perhaps the greatest “miracle year” in the history of science, 1905, Einstein published four papers that each changed the course of physics. This work was largely ignored by the physics community for several years, thereby contributing little to the IF of Annalen der Physik.

Perhaps most concerning about citation metrics is that they are open to abuse. Journals use the IF to vie for the affections of the scientific community and naturally attempt to improve it by gaming the system. Exploiting the citability of review articles is a particularly effective way to increase the IF, and the scientific community eagerly plays along by seeking the

Address for reprint requests and other correspondence: J.H.T. Bates, Univ. of Vermont College of Medicine, 149 Beaumont Ave., Burlington, VT 05405-0075 (e-mail: jason.h.bates@uvm.edu).
reflected glory of being in highly cited company. Citation metrics can also be manipulated by individual scientists. For example, you can maximize the value of your own $H$-index by making sure to cite every one of your prior papers in each new submission [for those interested in this noble pursuit, the formula after $n$ papers is that $H$ will be at least $n/2$ if $n$ is even and at least $(n - 1)/2$ if $n$ is odd].

The truth is that judging cutting-edge research is extremely difficult and in many respects can only be done properly by peers who have the necessary years of training and experience. Peer review, of course, is not without the limitations of a system that relies on professional discipline to keep personal bias at bay and certainly needs to be strictly monitored for conflict of interest. Peer review must also be based on carefully articulated evidence of research productivity rather than vacuous opinion, but at least it has the potential to save us from the intolerable totalitarianism that would result from slavery to numbers. The situation is perhaps reminiscent of Winston Churchill’s famous statement concerning democracy as being “the worst form of government, except for all the others.” Nevertheless, citation metrics are apparently here to stay, and they arguably have a useful role to play in the impartial evaluation of research provided they are used in a balanced fashion without undue reliance on any one metric in particular. We must not forget, however, that citation metrics are no substitute for the carefully considered opinions of arms-length peers who have the capacity to weigh all the intangibles that go into an informed assessment of research impact.

**DISCLOSURES**

No conflicts of interest, financial or otherwise, are declared by the author(s).

**AUTHOR CONTRIBUTIONS**

Author contributions: J.H.B. drafted manuscript; J.H.B. and P.D.W. edited and revised manuscript; J.H.B. and P.D.W. approved final version of manuscript.

**REFERENCE**