The science of progress and the progress of science

With increasing demands for science to provide value, how do we assess research to ensure that short-term gains do not undermine long-term goals?

Philip Hunter

Scientific ‘progress’ is under increasing scrutiny in the light of growing demand from governments, funding agencies and businesses to see a return on their investment in research. Along these lines, the assessment of research has become increasingly important to policy-makers in deciding how to distribute finite amounts of money. However, this scrutiny itself can change the direction of research and carries the risk of emphasizing short-term gains at the expense of important long-term goals. Fortunately, those involved in scrutinizing science and studying its processes and progress—bundled under the term ‘meta-science’—are extending their assessment beyond the short-term to embrace environmental, societal and even philosophical aspects of science. The aim is to integrate work on measuring the scientific output of institutions, research groups and individuals with more fundamental aspects of scientific discovery, building on the work of philosophers of science such as Karl Popper and Thomas Kuhn.

...scrutiny itself can change the direction of research and carries the risk of emphasizing short-term gains at the expense of important long-term goals

It seems futile to attempt to predict or understand the impact of great scientific advances by using short-term assessments of research output. In fact, the

best framework for assessing both major advances and ongoing innovation probably takes a long view and is not narrowly focused on immediate output, as measured in publications or citations. However, meta-science must also resolve the tension between the disparate objectives of scientists, the public and funders, which are often resolved over different timescales. A good example of these conflicting viewpoints concerns fishing policy: research indicates that stocks can only be saved by temporarily imposing draconian limits on catches, but the short-term impact might be economic devastation for local communities in return for the promise of a sustainable fishing industry over a longer period.

Various countries have begun to establish frameworks that seek to reconcile these different demands on research within their funding programmes, with Australia among the pioneers in this area. In 2011, Les Rymer, from the Group of Eight Limited, which represents eight leading Australian universities, was lead author on a report for the Group on ‘Measuring the Impact of Research’ (http://www.go8.edu.au/_documents/go8-policy-analysis/2011/go8backgrounder23_measimpactresearch.pdf). Rymer was given the task of analysing the cross-fertilization between different disciplines, as well as between the strands of research within a field, in pursuit of long-term goals, such as a cure for Alzheimer disease. The report also sought compromise between different economic goals, ranging from immediate impact to long-term competitive advantage. “There are many intangible benefits of research which are

It seems futile to attempt to predict or understand the impact of great scientific advances by using short-term assessments of research output

nevertheless real and of value—including on national reputation and attractiveness as a place to learn, work and invest,” Rymer wrote in the report. He also acknowledged the challenge of determining the best methods for evaluating research across many dimensions, from individual projects to whole institutions and nations. This challenge has been taken up in various projects around the world, stimulated by a growing demand for new methods of research assessment.

Interest in new methods of research assessment began in 2005 with the invention of the *h*-index by American physicist Jorge Hirsch [1]. The *h*-index attempts to measure the cumulative impact of an individual or team, but has proven somewhat controversial. Despite this, its appeal is that it provides a single number by which to compare people and groups, and it provides a starting point for further refinement. As such, refinement of the *h*-index has been seized on by some agencies and institutions for use as a tool in assessing grant and tenure applications. One such refinement has been to try to introduce some predictive capability, as the *h*-index per se is merely a measure of how well a scientist or group has done in the past.

However, the current *h*-index is a poor predictor of its own future value, as it deliberately does not take account of information—such as the journal a paper is published in—that might help to determine the number of citations that a published paper will attract. If applied to hiring decisions, the *h*-index would therefore disadvantage young scientists in particular; although as Werner Marx, from the Max Planck Institute for Solid State Research in Stuttgart, Germany, pointed out, the *h*-index is of little help in predicting the future performance of any scientist. Marx has applied bibliometrics—the analysis of scientific literature—to the history of science, and noted that “the most recent one to two publication years can hardly be evaluated because of the time delay of citations as a basic phenomenon”.

“There are many intangible benefits of research which are nevertheless real and of value—including on national reputation and attractiveness as a place to learn, work and invest...”

A group at the Northwestern University of Chicago, USA, has attempted to address this weakness of the *h*-index [2]. Their basic idea is to take account of a researcher's past history in addition to the *h*-index to predict better what might be his or her future *h*-index. This work has so far been confined to the realm of neuroscience, and further research would be required to extend it to other disciplines, according to one of the paper's authors, Konrad Kording, who runs the Bayesian Behaviour Lab at the university. Within neuroscience, however, the study has met its objective: “Our compound index predicts the *h*-index much better than the *h*-index itself, and explaining far more variance makes it far more useful,” Kording commented. “Still, it predicts far less than the entire variance and can undoubtedly be improved.”

Kording accepts that his paper has attracted substantial criticism, but notes that much of it is aimed at the focus on the *h*-index itself, rather than the attempt to incorporate compound information to predict future performance. For example, some critics cited an earlier study that found that multiple authorship on many papers means the *h*-index is fundamentally

biased [3]. As such, authors who have only a small role on a variety of papers can accumulate an *h*-index that overstates their cumulative contribution.

This criticism relates to another complaint: when drilling down to individual scientists, the *h*-index becomes statistically invalid, not just because of the distortions by multiple authors, but also because the numbers involved are too small for significant results. “There has been debate over whether measurement of *h*-index at the individual level is ethical,” commented Loet Leydesdorff, a Dutch sociologist and cyberneticist at the University of Amsterdam in the Netherlands, who has worked extensively on the sociology of innovation. “As soon as one uses statistics, the *N* of publications in the denominator becomes so small that one is probably no longer entitled to say much.”

Leydesdorff conceded that the analysis of publications and citations is useful for identifying broadly whether an individual is worth considering for a grant or job, but noted that it is not much use in making finer distinctions within a pool of chosen candidates. “In a study on different ways of comparing rejected to awarded applications with Lutz Bornmann and Peter van den Besselaar, we found that the grantees were better in terms of publications and citations when compared with the rejected applicants as a whole group, but when we discarded the tail of the distribution, the ‘best rejected’ are often significantly better than the grantees,” Leydesdorff explained (<http://www.leydesdorff.net/meta-evaluation/index.htm>). “Perhaps this suggests it is too difficult to distinguish between ‘good’ and ‘excellent’ for committees,” he concluded.

If applied to hiring decisions, the *h*-index would therefore disadvantage young scientists...

However, Leydesdorff's collaborator, Lutz Bornmann, a sociologist of science at the Max Planck Society in Munich, Germany, has not given up on improving the *h*-index. “The *h*-index can be used for the assessment of scientists, if the scientists are active in the same field and do have the same ages or at least similar ‘academic’ ages,” he said. “If this is not the case, one should use field and time-normalized indicators [4]. These kinds of indicators are becoming standard in bibliometrics.”

A tougher issue is how to measure the wider impact of research as a whole, rather than the individual impact or contribution of any one person or group. This is a considerably bigger challenge, according to Marx, because of the potential to inadvertently stifle emerging research. “The far more serious matter is that citation data could favour mainstream research and disadvantage new up and coming research fields,” Marx said. “Citation counts are determined by many factors beyond quality, such as language of the publishing journal, the citation performance of the cited references, and possibly the size of the relevant community. In large publication sets, such biases are averaged out. But until now, not enough research has been undertaken to make sure that niche research is treated adequately and is not disadvantaged by bibliometric indicators.”

It is here that performance measurement and deeper philosophical considerations begin to overlap. Marx has collaborated with Bornmann to develop conditions for evaluating research that are not prejudiced against emerging fields or new ideas. In doing so, they have countered the philosophy developed by Kuhn, according to whom science is a darwinian process in which the progression of ideas is governed by natural selection. “Kuhn did not deny that there is progress in science, but he denied that it is progress towards anything,” Marx explained. “He often used the metaphor of biological evolution: scientific progress, for him, was like evolution as described by Darwin; a process driven from behind, rather than pulled towards some fixed goal to which it grows ever closer. For [Kuhn], the natural selection of scientific theories is driven by problem solving. When, during a period of normal science, it turns out that some problems can't be solved using existing theories, then new ideas proliferate, and the ideas that survive are those that do best at solving these problems. Kuhn recognized that Maxwell's and Einstein's theories are better than those that preceded them, in the same way that mammals turned out to be better than dinosaurs [...] but when new problems arise, they will be replaced by new theories that are better at solving those problems, and so on, with no overall improvement.”

Marx rejects this notion: “All this is wormwood to scientists like myself, who

think the task of science is to bring us closer and closer to objective truth." Such thinking led Marx and Bornmann to develop their Anna Karenina Principle (AKP), which provides a set of tests for assessing a scientific development or proposal. It is named after Leo Tolstoy's famous first sentence in his novel *Anna Karenina*, "Happy families are all alike; every unhappy family is unhappy in its own way."

Tolstoy meant that for a family to be happy, several crucial conditions must all be fulfilled, including good health of all family members, acceptable financial security and mutual affection. Similarly, for the AKP to hold in assessing new theories, all its conditions must be fulfilled, such as

peer review of grant proposals and manuscripts; citation of publications; and new scientific discoveries. Among the 11 criteria assessed by AKP, one is that a new theory should have explanatory power. This accords with the view of Karl Popper, a contemporary of Kuhn. Popper and Kuhn disagreed on this point, according to Brad Wray, a specialist in the philosophy of science at the State University of New York, USA: "Kuhn believed that revolutionary changes of theory interrupted the 'normal' practice of science, and that these revolutionary changes often involved a loss in explanatory power," Wray explained. "That is, some things that we could explain before the revolution could not

"The far more serious matter is that citation data could favour mainstream research and disadvantage new up and coming research fields"

be explained after, at least not by the new theory. Popper, on the other hand, believed that scientists only accept a new theory if it preserved the successes of its predecessor. So central to Popper's view is that scientific knowledge is essentially cumulative."

Wray believes that it is too soon to confirm whether Popper was correct. "I do not think that we are ready yet to leave Kuhn behind. [...] I think Kuhn has provided us

with a general framework for understanding how science works, but to fully understand Kuhn's view we need to look at his later work, *The Road Since Structure*. Importantly, in the papers in this book, Kuhn emphasizes the importance of the creation of new scientific specialties."

Indeed, Wray observes the emergence of common ground after Popper's and Kuhn's deaths. "A lot of the work in the social epistemology of science aims to understand how the social structure of science contributes in positive ways to the growth of knowledge," he said. "Some of the more interesting work in this area is

In the real world of competition for research money [...] a certain degree of darwinism is inevitable

now being done in Europe," he added. "I see a lot of this work as extending the general Kuhnian image and project, though not all of these authors see their work that way."

Kuhn's legacy, Wray argues, is to convince philosophers that the social structure of science has an important role in progress; an idea that does not contradict the view of Marx and Bornmann that science should pursue an ideal or goal, rather than being subjected to natural selection. In the real world of competition for research money, however, a certain degree of darwinism is inevitable. Yet, meta-science increasingly provides a long-term perspective and helps to identify ideas and research of potential promise that would otherwise be overlooked.

CONFLICT OF INTEREST

The author declares that he has no conflict of interest.

REFERENCES

1. Hirsch J (2005) An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* **102**: 16569–16572
2. Acuna D, Allesina S, Kording KP (2012) Future impact: predicting scientific success. *Nature* **489**: 201–202
3. Schreiber M (2008) To share the fame in a fair way, h_m modifies h for multi-authored manuscripts. *New J Phys* **10**: 040201
4. Bornmann L (2012) A better alternative to the h index. *J Informetr* **7**: 100

Philip Hunter is a freelance journalist in London, UK.

EMBO reports (2013) **14**, 601–604; published online 14 June 2013; doi:10.1038/embor.2013.83