Scientists 'bad at judging peers' published work,' says new study

8 October 2013

Are scientists any good at judging the importance of the scientific work of others? According to a study published 8 October in the open access journal PLOS Biology (with an accompanying editorial), scientists are unreliable judges of the importance of fellow researchers' published papers.

The article's lead author, Professor Adam Eyre-Walker of the University of Sussex, says: "Scientists are probably the best judges of science, but they are pretty bad at it."

Prof. Eyre-Walker and Dr Nina Stoletzki studied three methods of assessing published scientific papers, using two sets of peer-reviewed articles. The three assessment methods the researchers looked at were:

- Peer review: subjective post-publication peer review where other scientists give their opinion of a published work;
- Number of citations: the number of times a paper is referenced as a recognised source of information in another publication;
- Impact factor: a measure of a journal's importance, determined by the average number of times papers in a journal are cited by other scientific papers.

The findings, say the authors, show that scientists are unreliable judges of the importance of a scientific publication: they rarely agree on the importance of a particular paper and are strongly influenced by where the paper is published, overrating science published in high-profile scientific journals. Furthermore, the authors show that the number of times a paper is subsequently referred to by other scientists bears little relation to the underlying merit of the science.

As Eyre-Walker puts it: "The three measures of scientific merit considered here are poor; in particular subjective assessments are an error-prone, biased and expensive method by which to assess merit. While the impact factor may be the most satisfactory of the methods considered, since it is a form of prepublication review, it is likely to be a poor measure of merit, since it depends on subjective assessment."

The authors argue that the study's findings could have major implications for any future assessment of scientific output, such as currently being carried out for the UK Government's forthcoming Research Excellence Framework (REF). Eyre-Walker adds: "The quality of the assessments generated during the REF is likely to be very poor, and calls into question whether the REF in its current format is a suitable method to assess scientific output."

PLOS Biology is also publishing an accompanying Editorial by Dr Jonathan Eisen of the University of California, Davis, and Drs Catriona MacCallum and Cameron Neylon from the Advocacy department of the open access organization the Public Library of Science (PLOS).

These authors welcome Eyre-Walker and Stoletzki's study as being "among the first to provide a quantitative assessment of the reliability of evaluating research", and encourage scientists and other to read it. They also support their call for openness in research assessment processes. However, they caution that assessment of merit is intrinsically a complex and subjective process, with "merit" itself meaning different things to different people, and point out that Eyre-Walker and Stoletzki's study "purposely avoids defining what merit is".

Dr Eisen and co-authors also tackle the suggestion that the impact factor is the "least bad" form of assessment, recommending the use of multiple metrics that appraise the article rather than the journal ("a suite of article level metrics"), an approach that PLOS has been pioneering. Such metrics might include "number of views, researcher
bookmarking, social media discussions, mentions in
the popular press, or the actual outcomes of the
work (e.g. for practice and policy)."

**More information:** Eyre-Walker A, Stoletzki N
(2013) The Assessment of Science: The Relative
Merits of Post-Publication Review, the Impact
Factor, and the Number of Citations. *PLoS Biol*
11(10): e1001675. [DOI: 10.1371/journal.pbio.1001675]

Provided by Public Library of Science
APA citation: Scientists 'bad at judging peers' published work,' says new study (2013, October 8)