Evaluating individual researchers' performance

Lutz Bornmann
Division for Science and Innovation Studies, Administrative Headquarters, Max Planck Society, Hofgartenstr. 8, 80539 Munich, Germany; bornmann@gv.mpg.de

Werner Marx
Max Planck Institute for Solid State Research, Heisenbergstraße 1, 70569 Stuttgart, Germany; w.marx@fkf.mpg.de

Abstract In 2005 the *h* index was proposed to reflect individual researchers' output in terms of publications and citations. However, this intergral indicator is not normalized for age and subject category, and therefore comparisons between researchers differing in academic age and professional background are impossible. To overcome the limitations of the *h* index, we propose alternative metrics, which are based on the percentile approach.

Keywords Bibliometrics, credentials, *h* index, percentile approach, $P_{top\ 10\%}$, $PP_{top\ 10\%}$

Researchers rarely receive scientific awards and prizes. Successful researchers are usually awarded for scientific discoveries and achievements in the distant past, and there is always a dilemma of quantitatively evaluating implications of research.

In natural and life sciences bibliometric indicators have become indispensable tools for measuring individual researchers' performance. These indicators quantitatively reflect the weight of publications and citations. Ideally, all publications and their citations should be included in the evaluation. However, it is hardly possible to reliably assess the citability of the most recent publications, which need some time to collect citations. Depending on the subject category, citations of a publication reach their peak level within two to three years and thereafter steadily decrease. This is particularly true for chemistry, physics, biology, biomedicine, and clinical medicine.[1]

Concrete steps for bibliometric evaluation of individual researchers have been made in the past decade. In 2005 the *h* index was proposed to reflect individual researchers' output in terms of publications and citations in one integral indicator.[2] Before 2005 the number of publications and citations had been counted separately. The *h* index rapidly attracted much attention and became a subject for research.[3] It is now displayed as an individual indicator in Web of Science, Scopus and Google Scholar.

The *h* index has its advantages and disadvantages. The main advantage is that it is easy to calculate, provided there is a list of papers and their citations in descending order. However, the index is not normalized for age and subject category, and therefore comparisons between researchers differing in academic age and professional background are impossible. To overcome the limitations of the *h* index, quite recently alternative metrics have been proposed.[4] The new indicators are based on a percentile approach (a relative scale between 0 and 100) where only publications with a percentile less than or equal to 10 are counted. It takes into account the number of a researcher's publications belonging to the 10% most-cited publications in a certain field and publication year ($P_{top\ 10\%}$). The percentile approach assesses the impact of a publication compared to others in the same subject category and publication year. Using the distribution of citation frequencies (sorted in descending order), all publications in the same subject category and year as the publication in question are broken down into percentile ranks. The lower the percentile rank of a publication, the more citations it received (in InCites, Thomson Reuters). The percentile of a publication in question is determined using the distribution of percentile ranks across all publications. For example, a value of 10 means that the publication in question is among the 10% most-cited publications; the other 90% of publications have less impact.

The percentile approach of standardizing citations allows comparing the impact of publications from different subject categories and publication years. Another advantage of the new approach is that it does not set an arbitrary threshold to determine publications with a high citation impact. Earlier the arbitrary use of the *h* index was criticized,[5] and the development of thresholds based on empirical analysis and subject category was suggested.[6]

$P_{top\ 10\%}$ can be defined as the number of highly cited or excellent publications which belong to the 10% most cited publications in a certain subject category and publication year.[7] Eugene Garfield's words of wisdom are appropriate to recall here: "A highly cited work is one that has been found useful by a relatively large number of people, or in a relatively large number of experiments".[8]

To compare the number of $P_{top\ 10\%}$ with an expected value, it is possible to calculate the proportion of $P_{top\ 10\%}$ in a researcher's publication set ($PP_{top\ 10\%}$). Such a comparison with an expected value is impossible in case of the *h* index. The expected value of $PP_{top\ 10\%}$ is 10%. In other words, if a random sample of publications of a researcher is to be picked from the Web of Science or another database, it is expected that 10% of these will belong to the 10% most cited publications in a certain subject category and publication year.[7] $PP_{top\ 10\%}$ has already been regarded as the most important indicator in the Leiden Ranking of universities by the Centre for Science and Technology Studies (Leiden University, The Netherlands).[9]

In principle all indicators of individual researchers should be age-normalized. In the same way as Jorge Hirsch proposed the m quotient for the *h* index,[2] we propose the number of years as an active researcher ($P_{top\ 10\%}$ quotient) to normalize $P_{top\ 10\%}$ for age. As a good example, the cumulative impact of a researcher's publications can be weighed by correcting for years since completion of his/her doctoral

studies.[10] When the number of years as an active researcher is counted, the $P_{top\ 10\%}$ quotient is therefore corrected not just for the publication year and subject category, but also for age.

Finally, we would like to point out that the evaluation of an individual's performance should not be based on a single indicator. Bibliometric indicators should be complemented by other parameters (eg academic duties, mentoring, patents, speaker invitations, international contracts, awards).[11] Such a complex approach is especially important for humanities, communication and social sciences, where the use of the current bibliometric indicators is unsatisfactory.[11]

References

1 Abramo G, Cicero T, D'Angelo CA. Assessing the varying level of impact measurement accuracy as a function of the citation window length. *Journal of Informetrics* 2011;5(4):659–667. doi: 10.1016/j.joi.2011.06.004

2 Hirsch JE. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102(46):16569–16572. doi: 10.1073/pnas.0507655102

3 Bornmann L, Daniel H-D. The state of *h* index research. Is the *h* index the ideal way to measure research performance? EMBO Reports 2009;10(1):2–6. doi: 10.1038/ embor.2008.233

4 Bornmann L. How to analyse percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes and top-cited papers. *Journal of the American Society for Information Science and Technology* 2013;64(3):587–595. doi: 10.1002/asi.22792

5 Waltman L, van Eck NJ. The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology* 2012;63(2):406–415. doi: 10.1002/asi.21678

6 Kreiman G, Maunsell JH. Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience* 2011;5:48. doi: 10.3389/fncom.2011.00048

7 Bornmann L, de Moya Anegón F, Leydesdorff L. The new Excellence Indicator in the World Report of the SCImago Institutions Rankings 2011. *Journal of Informetrics* 2012;6(2):333–335. doi: 10.1016/j.joi.2011.11.006

8 Garfield E. Citation indexing - its theory and application in science, technology, and humanities. New York, NY, USA: John Wiley & Sons, Ltd, 1979.

9 Waltman L, Calero-Medina C, Kosten J *et al*. The Leiden Ranking 2011/2012: data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology* 2012;63(12):2419–2432. doi: 10.1002/asi.22708

10 Nosek BA, Graham J, Lindner NM *et al*. Cumulative and career-stage citation impact of social-personality psychology programs and their members. *Personality and Social Psychology Bulletin* 2010;36(10):1283–1300. doi: 10.1177/0146167210378111

11 Sahel JA. Quality versus quantity: assessing individual research performance. *Science Translational Medicine* 2011;3(84):84cm13. doi: 10.1126/scitranslmed.3002249

12th EASE General Assembly and Conference

13th – 15th June 2014, University of Split, School of Medicine, Croatia

Ana Marusic has kindly arranged the use of facilities at the School of Medicine of the University of Split and we are very grateful to the University authorities for hosting us.

Entertainment:
School of Medicine Men's Chorus

Welcome reception:
Garden of the Mestrovic Gallery, Split

Conference dinner:
Villa Rosina, Split

Organizing Committee: Joan Marsh, Arjan Polderman, Rod Hunt, Ana Marusic, Pero Vidan, Dalibora Behmen