
Letter to the Editor

Self-citation can inflate *h*-index

LEV A. ZHIVOTOVSKY,^{a,b} KONSTANTIN V. KRUTOVSKY^c

^a N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow (Russia)

^b Morrison Institute for Population and Resource Studies, Stanford University, Stanford (USA)

^c Department of Forest Science, Texas A&M University, Texas (USA)

Sir,

A new citation index, *h*-index, for rating the research output of individual scientists recently proposed by HIRSCH [2005] seems an objective measure of individual research achievement that is transparent, unbiased and easily measurable (see GLÄNZEL [2006] for review). The *h*-index represents the highest number of papers a scientist has published that have each received at least that number of citations: for example, a researcher with an *h* of 50 has written 50 papers that have each had at least 50 citations. The *h*-index depends on both the number of a scientist's publications, and their evaluation by the corresponding scientific community. However, the *h*-index does not separate someone's citation by his or her peers from self-citation, which refers to citing one's own publications. Self-citation can potentially affect bibliometrics, citation index and science policy, and it remains a controversial issue (see GLÄNZEL & AL., [2006] for review).

Received August 2, 2006

Address for correspondence:

LEV A. ZHIVOTOVSKY

N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences

3 Gubkin Str., Moscow 119991, Russia

E-mail: levazh@gmail.com

0138–9130/US \$ 20.00

Copyright © 2008 Akadémiai Kiadó, Budapest

All rights reserved

We studied the impact of self-citation on the h -index to see whether and how severe self-citation may bias the value of the h -index. To address this problem we modeled the behavior of the h -index as a function of the number of years of publishing and the number of citations, and then compared the values of the h -index based only on self-citation with the values based on peer citation that excluded self-citation.

We simplified our model in order to get a contrast qualitative output showing the influence of those parameters. We assumed that a given person has authored or coauthored n published papers a year for T years and that peers acknowledged each of his or her paper by c citations each year regardless of when the paper was published. Therefore, $n(T-t)$ papers published t years ago or earlier have been cited at least ct times. If these two quantities are equal to each other, their common integer value is h .

Equating these gives $t = nT/(n+c)$, and thus roughly $h \approx \frac{nc}{n+c}T$. Hence, in case of peer citation by scientific community, the h -index directly correlates with the time of publishing, and depends symmetrically on both the annual number of papers and their impact.

In case of self-citation, we assumed that in each new paper the author or coauthor cited s of his or her own previously published papers randomly chosen (we ignored a 'border effect' of the earliest publications). Therefore, each of $n(T-t)$ papers published t years ago will get on the average $s/(T-t) + s/(T-t+1) + \dots = s \cdot \left(\sum_{i=1}^T 1/i - \sum_{i=1}^{T-t} 1/i \right) \approx s \ln(T) - s \ln(T-t)$ self-citations (the formula follows from an approximation of a harmonic series by the natural logarithm plus Euler's constant). Equating both parts and noting that $n(T-t) \approx h$ gave $h/s + \ln(h/s) = \ln(Tn/s)$, where Tn is the total number of published papers. An approximate solution of this equation is $h \approx s \cdot \ln(N/s) \cdot \frac{\ln(\ln(Tn/s))}{1 + \ln(Tn/s)}$.

For example, if someone has published four papers a year for 20 years, and each of them has been cited by peers five times annually, then $h=45$. In contrast, if someone has published for the same 20 years with the same efficiency of four papers a year, and each of his or her paper contained five self-citations of his or her own previously published papers, then $h=10$. Therefore, the modest self-citation does not contribute much in the h -index compared to peer citations. However, if the same person has published ten papers a year for 40 years, and if each paper contained 10 self-citations, then only due to self-citations alone the h -index would reach 26. Although self-citations are generally ageing faster than foreign citations [GLÄNZEL & AL., 2004] and thus contribute moderate bias in the values of citation indexes at their top range, we can conclude that *self-citation may falsely indicate the self-citer as being an outstanding accomplished scientist while this would not necessarily be the case.*

We believe that rational self-citation is a necessity and well-justified. However, although it requires more detailed analysis, the results of our preliminary study demonstrate that while modest self-citation does not greatly affect the h -index in the short run, excessive self-citation can inflate the h -index in the long run despite the h -index is positively affected by aging (earlier papers) unlike self-citation. Furthermore, excessive self-citation may be a common case for some scientific fields in some closed scientific societies considering situations when early in his or her career someone becomes in charge of a large group of slowly rotated researchers who are inclined to share authorship with their leader and cite his or her papers, especially if this person is on editorial boards of journals where these papers are submitted or is in charge of distributing fund grants. This calls for exclusion of self-citations from the h -index, and likely also from any other citation indexes, to achieve a more objective evaluation of the published research impact of individual scientists. It would be useful anyway, if both indexes are calculated and compared, one that is based on all citations and another excluding self-citations.

References

- GLÄNZEL, W. (2006), On the h -index – A mathematical approach to a new measure of publication activity and citation impact, *Scientometrics*, 67 (2) : 315–321.
- GLÄNZEL, W., THIJIS, B., SCHLEMMER, B. (2004), A bibliometric approach to the role of author self-citations in scientific communication, *Scientometrics*, 59 (1) : 63–77.
- GLÄNZEL, W., DEBACKERE, K., THIJIS, B., SCHUBERT, A. (2006), A concise review on the role of author self-citations in information science, bibliometrics and science policy, *Scientometrics*, 67 (2) : 263–277.
- HIRSCH, J. E. (2005), An index to quantify an individual's scientific research output, *Proceedings of the National Academy of Sciences of the United States of America*, 102 (46) : 16569–16572.