doi:10.2533/chimia.2009.66 Chimia 63 (2009) 66–68

The h Index – Help or Hype?

Veronika R. Meyer*

Abstract: Three years ago a bibliometric index for the qualification of a person's scientific output was proposed by Hirsch, the so-called h index. This is an integer number which combines the number of papers of an author and the number of citations they gathered. Thus the h index is an indicator for both the productivity and the impact of a scientist. This paper presents the properties of the h index and the great attention it attracted within a short time. Numerous other indices, claimed to be better than the original, were proposed in the meantime. These developments are discussed critically.

Keywords: Bibliometry · h Index · Hirsch index · Index inflation

Introduction

It is one of the wonders of the human mind that gifted people are able to materialize the things that are in the air. Goethe wrote 'Werther', John Lennon composed 'Imagine', Einstein explained space and time – and people were fascinated. A paper published in the November 2005 issue of *PNAS, the Proceedings of the National Academy of Sciences of the United States of America*, found rapid attention and passionate discussion in many fields of science and the humanities. Jorge E. Hirsch, a physicist of the University of California at San Diego, proposed 'An index to quantify an individual's scientific research output'[1] (this is in fact the title of his paper). Although Hirsch's proposal was not the first attempt to rank scientists in an objective way, namely by a number, he obviously presented an idea that was both convincing and controversial. By end of November 2008, his paper was cited not less than 210 times, as listed by the *ISI Web of Knowledge*.[2] It also provoked quite a number of scientists to come up with other proposals to measure the 'scientific importance' of a person.

The summary of Hirsch's paper is short and brings his idea to the point: "I propose the index h, defined as the number of papers with citation number ≥h, as a useful index to characterize the scientific output of a researcher." In other words, the h index is identical to the highest rank r, when papers are listed according to decreasing number of citations they received, where the paper on this rank has h = r citations. Or even simpler: It is the highest number h of papers of a person that were cited at least h times. If there is no paper in the list with an equal number of citations as according to the rank then h is <r. The h index combines both the productivity of an individual (his or her number of papers) and the attention earned by the scientific community (the number of citations these papers received over time).

The meaning of these statements is explained here with my own case. With the function 'Author Finder' and the name 'meyer v r' a total of 74 papers is listed in the *ISI Web of Knowledge* (Nov. 28, 2008). These papers can be sorted by date, first author, relevance,[3] source title, publication year, or times cited. With the latter function the most-cited paper is given rank 1; it was cited 113 times. The numbers of citations drop rapidly in the list. The second-ranking paper received 45 citations and the third one 33. The papers with rank from 10 to 14 were all cited 14 times, the one with rank 15 was cited 13 times. Therefore rank and number of citations cross at 14 which is my actual h index. See the Fig. for a graphical representation of rank and citation numbers.

Obviously a high h index is the characteristic of a scientist who is writing many papers that are cited frequently. A list of living chemists with h indices of 50 and higher was recently published by *ChemistryWorld*.[4] It consists of 520 names with G. M. Whitesides at the top with a most impressive h index of 140. This means that he is the author or co-author of 140 scientific papers which all received 140 or more citations (plus a long list of publications which were cited less than 140 times each). According to Hirsch the advantage of his index is the fact that it neither over-estimates the most-cited papers of a researcher which may be 'lucky strikes' nor does it consider the numerous papers of a prolific person which do not find much attention by the scientific community.

The impact of Hirsch's paper was remarkable. By the end of the following year, 2006, it was already cited 21 times (a number almost every scientist dreams of). The h index is a striking number once one has understood its definition. Therefore it is now offered by mouseclick by the *ISI Web of Knowledge* as the function 'Create Citation Report' and everybody can easily find out his or her index, at least in principle. Another database which allows this is *Scopus* whereas *SciFinder Scholar* or *Google Scholar* do not offer the h index calculation.

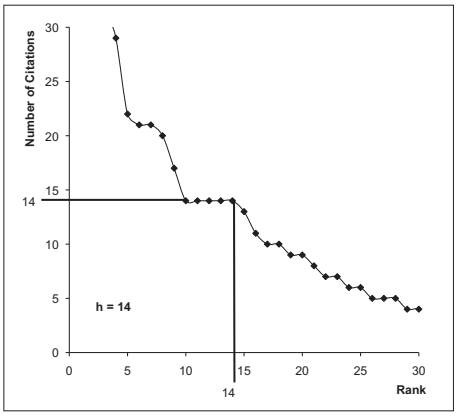

Fig. Detail of the relationship between publication rank and number of citations gathered by V. R. Meyer

The Features of the h Index …

– It is an integer. The higher the better.

– It is not sensitive to extremes but in fact it is of astonishing robustness.[5] It is not influenced by a few highly cited papers on the top ranks or by a long 'tail' of publications which did not attract much attention.

– It cannot be higher than the number of papers published by a certain author. If somebody wrote, *e.g.* no more than 20 papers, his or her h index will never be higher than 20 even if the least important publication would get 50 citations over time.

*Correspondence: PD Dr. V. R. Meyer
Empa, Swiss Federal Laboratories for Materials Testing and Research
Lerchenfeldstrasse 5
CH-9014 St. Gallen
Tel.: +41 71 274 77 87
Fax: +41 71 274 77 88
E-mail: veronika.meyer@empa.ch

– It increases with time as long as a certain author continues to be cited. Citations may well increase when a scientist ceases to publish and even after death. The h index of an individual cannot decrease.

– Hirsch showed that it has good predictive power for careers (at least in physics),[6] thus making it a *help* for tenures, promotions or funding. However, it is not clear whether this claim is also true for other scientific fields and whether it is more predictive than simpler data such as the total number of citations gathered.

– It is one of the features that can be directly found in the *ISI Web of Knowledge* or in *Scopus*.

… and its Weak Points (?)

– The h index is biased in favour of researchers with long careers;[7] it cannot be used to compare the merits of young scientists and 'old hands'. (In fact, it was not put forward for such comparisons.)

– It depends on specific fields of science as already demonstrated by Hirsch with a comparison of physics and the life sciences; individuals publishing in the latter field tend to reach higher h indices than physicists.[1] Publication and citation habits are not identical in the various disciplines. Normalization factors, although not trivial to calculate, have been published: If the h index of a physicist is not corrected (because it emerged from this field), then that of a chemist needs to be multiplied by 0.92 and or even by 0.44 for scientists in molecular biology and genetics to be comparable.[8] On the other hand, mathematicians may multiply their h index by 1.83.

– Books and book chapters are not included in the *ISI Web of Knowledge*. They might be on top of the list of an individual thanks to a large number of citations, thus their influence on the h index may be small but there are other opinions.[9] The *ISI Web of Knowledge* includes articles, meeting abstracts, reviews, letters, and notes as default; in addition, h indices based on any of these types, solely or in combination, can be obtained easily.

– The h index cannot distinguish between scientists with highly different citations/rank patterns, *i.e.* an individual with many highly influential papers and another person with only a few or no papers with rank < h.[10] It is possible to achieve an h index of 20 with no more than 20 papers or to get it only after a long career with 100 papers.

– The total numbers of scientists, of papers, and of citations are growing. Concerning the latter, it is much easier today to find papers and to copy-paste them into one's own publication. As a result, there is a danger of excessive and uncritical quoting. All three points together may lead to the fact that it is easier today to climb up the 'Hirsch ladder' than it was 50 years ago. A comparison of scientists over decades may be misleading.

– Long papers are cited more often than short ones; Ball has shown that this fact is at least true for astronomy[11] but why should the other disciplines of science be immune to such a pitfall although length and importance are not correlated? Thus, a scientist who tends to write lengthy papers will probably gain a higher h index than a colleague who is able to bring the results of research to the point.

– Self-citation can be a problem, and it is even possible to increase one's own h index by quoting the right paper of one's own list of publications. In my opinion self-citation should not be counted for the calculation of the h index, however, the *Web of Knowledge* does not offer this possibility. (There is an additional hard nut to crack: Self-citations are defined as the fact that V. R. Meyer mentions a V. R. Meyer paper in one of her papers. Yet, if my co-author R. Däppen adds our joint top-list paper to the references of one of his publications, this fact does not count as a self-citation with regard to my case although he is not an outside person who found our publication in the course of a literature search.) Hirsch held that the effect of self-citation on the h index is very small, if at all.[11] Engqvist and Frommen agree, and Batista *et al.* state that "it is hard to inflate".[12] On the other hand, Zhivotovsky and Krutovsky[13a] show that "Self-citation can inflate h-index" (this sentence is the title of their paper) and propose to exclude self-citation from the calculation of the index, an opinion shared by Iglesias and Pecharromán[8] or Vinkler.[13b]

– Scientists with identical last names and first name initials are not distinguished by the *ISI Web of Knowledge*. It is up to the users to find out which papers belong to the list of a certain person. Full first names are only shown after additional mouseclicks; therefore, one should consolidate the list with some labour. If first and second names are identical it is necessary to take a closer look at the topics of the papers and at the affiliations. A search for P. Vogel, a member of the Advisory Board of *Chimia*, yields too many papers if it is only based on initial and last name. Besides Pierre Vogel, a synthetic organic chemist, there is also Peter Vogel (oncology) or Petr Vogel (particle physics), to name but two. On the other hand, a famous 'true individual', as seen by the *ISI Web of Knowledge*, is Nobel prize winner J. M. Lehn.

– The name problem is also critical with regard to people who change their name during their career, *e.g.* due to marriage, with regard to double-names (very common in Spanish but also with married women), with regard to names with special characters (Wüthrich, Wuthrich or Wuethrich?), and with regard to the number of given names. In my case one paper is not listed under V. R. Meyer because the R. was not considered by the main author. It is a publication which has received 19 citations so far. However, a search under V. Meyer yields a total of 287 papers and it is difficult for a third person to dig out this lost gem.

The Index Inflation

The h index is a striking proposal although its definition is arbitrary. In his seminal paper, Hirsch himself discusses the possibility of other figures of merit such as the total number of papers or of citations gathered.[1] He rejects them all with the exception of m (now called m quotient), with m = h/(time span in years from the first published paper till the present).[14] It allows persons with different 'scientific age' to be compared.

By mentioning other possible indices, Hirsch opened the door to a discipline which may be called 'index science' (since its protagonists claim the scientific character of their ideas). A remarkable number of new proposals were published within short time (Table, probably not complete). In all cases the authors supply evidence in support of their 'better' idea with deep investigations of publication and citation lists as well as with statistics about the output of 'outstanding' and/or 'average' scientists. Most of these indices have recently been discussed by Bornmann *et al.*, followed by the introduction of their own m index.[15] The latter is a true example of the current index inflation! The topic has become a *hype*.

Antonakis and Lalive[16] maintain that the h index gives too much weight to quantity (number of papers) instead of quality. Two scholars from the same scientific field may both have a h index of 15 but one of them received a total of 1000 citations whereas the other one had 2000; the latter had a higher impact to science, therefore his ranking should be higher. As a consequence Antonakis and Lalive propose the 'index of quality and productivity', IQp, a complicated function of the number of citations, number of papers, the impact factors of the journals which publish the papers of the author in question (in order to correct for the different fields of science with differing citation habits), and the scientific age of this author. They claim that IQp correlates well with expert ratings of 'greatness' – it seems as if they have tried to define the ultimate index. However, IQp is much less attractive than all the other proposed indices due to its intricate manner of calculation; in addition, it is not straightforward to understand. And, last but not least, is it possible to coin

Table. Bibliometric indices and other figures of merit that have been proposed recently

Property	Name	Reference
Productivity vs. Impact	Number of papers	trivial
	Number of citations earned	trivial
Mean impact	Mean number of citations earned per paper	[1]
Combination of productivity and impact	h index	[1]
	PI index	[13b]
	h_α index	[17]
Most cited papers are considered	g index	[18]
	h(2) index	[19]
	R index	[20]
	A index	[21]
	m index	[15]
	h_w index	[22]
	g_α index	[17]
Self-citations are not considered	h_s index	[23]
Time dependence	m quotient	[1]
	h index sequence	[24]
	h rate	[7b]
	AR index	[20]
Number of coauthors is considered	h_I index	[12b]
	h_m index	[25]
The ultimate index, see text	IQ_p	[16]

a single number to describe the life-work of a person?

Looking through Critical Glasses

Some scientists do not hide their critical position:

"One can speculate that this bare-bones, one-dimensional way of ranking scientists may not be completely unrelated with the predominantly Anglo-Saxon habitude of classifying base-ball and cricket players by their batting average... Physicists are particularly fond of models involving a minimum of parameters... However, one ought to remember in this connection that Nature has no obligation to us of being simple..."[8]

However, perhaps it is not a typical Anglo-Saxon idea although Hirsch is a professor at the University of California in San Diego. Two French scientists recently reported about their four-month stay at the University of California in Berkeley (at the Departments of Anthropology and Chemistry, respectively):

"... Nor did we meet many people who believe that citation indices are a satisfactory measure for excellence. Many agreed with our view that such statistics ... are often used as a cheap substitute for a serious evaluation. The authorities in France and in Switzerland appear to be running behind when they emphasize citation indices and 'bibliometrics': the US is already moving beyond. Wisely."[26]

The h index can even give rise to a special type of humour. In order to increase your h by 30% or more you should not write any books or book chapters (because they are not considered for the calculation), you should not engage yourself in teaching or administration tasks; but inventing new indices or exotic terms may be advantageous.[27] As a contrast, a serious opinion is the following one:

"Finally, what do h indices really tell us? Of course, it is distasteful to reduce a lifetime's work to a number. Some scientists make huge contributions through their mentoring and generosity with ideas, skills, and time. Without them, academia would collapse: a department solely comprised of relentless publishers would be a joyless place for eager students."[7a]

However:

"Our impression is that h and m contain small kernels of truth."[7a]

A Personal View

The longer I was reading the 'index science' literature the more I became convinced that the above-mentioned quotations hit the nail on the head, including the last one claiming that the original h index may "contain a small kernel of truth". The h index may indeed be a *help* besides numerous other facts to be considered in the case of awarding a tenure position. However, its weight should be small compared to the truly important features of a person: devotion to a certain scientific field, enthusiasm, teaching ability (for a university position), and a great reservoir of ideas.

It can be assumed that the *hype* of bibliometric indices will go on for quite some time because it is easy to invent new numbers of merit and to apply them to the psychologists of Italy and to the chemists of Peru, not to forget the library scientists of Hungary. By publishing papers of this type one can 'hunt the deer'[28] and feed one's reputation. However, such research cannot really be taken seriously. The topic becomes ridiculous when the goal is to find the 'best scientist', not only of a certain field but of all science. By trying to do so some indices are calculated to four significant digits: Kurt Wüthrich has an IQp of 89.76![14] Such a number should be rounded to 90. In this regard the h index is of convincing simplicity: a likeable, coarse measure since it is an integer. But it must never be over-estimated!

Received: December 12, 2008

[1] J. E. Hirsch, *PNAS* **2005**, *102*, 16569.
[2] ISI Web of Knowledge: *http://apps.isiknowledge.com*. This website is not free of charge.
[3] Relevance is defined as "Sorts records based on a ranking system that considers how many of the search terms are found in each record".
[4] *http://www.rsc.org/images/H-index%20ranking%20of%20living%20chemists (MAR%2008)_tcm18-85867.pdf*
[5] J. K. Vanclay, *J. am. Soc. Inf. Sci. Technol.* **2007**, *58*, 1547.
[6] J. E. Hirsch, *PNAS* **2007**, *104*, 19193.
[7] a) C. D. Kelly, M. D. Jennions, *Trends Ecol. Evol.* **2006**, *21*, 167; b) Q. L. Burrell, *Scientometrics* **2007**, *73*, 19.
[8] J. E. Iglesias, C. Pecharromán, *Scientometrics* **2007**, *73*, 303.
[9] W. Glänzel, *Science Focus* **2006**, *1(1)*, 10.
[10] L. Bornmann, H. D. Daniel, *EMBO Reports* **2009**, *10*, 2.
[11] P. Ball, *Nature* **2008**, *455*, 274.
[12] a) L. Engqvist, J. G. Frommen, *Trends Ecol. Evol.* **2008**, *23*, 250; b) P. D. Batista, M. G. Campiteli, O. Kinouchi, A. S. Martinez, *Scientometrics* **2006**, *68*, 179.
[13] a) L. A. Zhivotovsky, K. V. Krutovsky, *Scientometrics* **2008**, *77*, 373; b) P. Vinkler, *J. Inf. Sci.* **2007**, *33*, 481.
[14] Irony on: As a consequence I cannot propose an m-index (such as the Meyer index) although it would not be too difficult to define some fancy number, *e.g.* m = 100h divided by the total number of papers, thus punishing researchers with long lists of publications which did not get much attention. Irony off.
[15] L. Bornmann, R. Mutz, H. D. Daniel, *J. Am. Soc. Inf. Sci. Technol.* **2008**, *59*, 830.
[16] J. Antonakis, R. Lalive, *J. Am. Soc. Inf. Sci. Technol.* **2008**, *59*, 956.
[17] N. J. van Eck, L. Waltmann, *J. Informetrics* **2008**, *2*, 263.
[18] L. Egghe, *Scientometrics* **2006**, *69*, 131.
[19] M. Kosmulski, *ISSI Newslett.* **2006**, *2(3)*, 4.
[20] B. H. Jin, L. M. Liang, R. Rousseau, L. Egghe, *Chin. Sci. Bull.* **2007**, *52*, 855.
[21] B. Jin, *Science Focus* **2006**, *1(1)*, 8.
[22] L. Egghe, R. Rousseau, *Inform. Process. Manag.* **2008**, *44*, 770.
[23] M. Schreiber, *Ann. Phys. (Leipzig)* **2007**, *16*, 640.
[24] L. M. Liang, *Scientometrics* **2006**, *69*, 153.
[25] M. Schreiber, *New J. Phys.* **2008**, *10*, 040201.
[26] A. Molinié, G. Bodenhausen, *Chimia* **2008**, *62*, 291.
[27] G. Ochs, *Nachr. Chem.* **2008**, *56*, 436.
[28] The German word *Hirsch* means *deer* in English.