Spanish Journal of Medicine

REVIEW ARTICLE

Author-level metrics: *h*-index and beyond

José M. Porcel and Leyre Liesa*

Department of Internal Medicine, Hospital Universitari Arnau de Vilanova, IRBLleida, Universidad de Lleida, Lleida, Spain

Abstract

Author-level metrics are usually employed for academic promotion and research funding. The *h*-index is a way of measuring scientists' productivity and impact on their field, determined by the number of publications and the number of times those publications have been cited. However, the *h*-index calculation does not capture the influence of factors such as research topics, article types, highly cited items, self-citations, number and position of authors, and academic career length. Nonetheless, variants of *h*-index that address some of these limitations correlate widely with their original metric, are not available in bibliographic databases and, overall, add little for measuring research productivity.

Keywords: Authorship. Hirsch index. Bibliometrics. Internal medicine.

Introduction

Assessing academic productivity of clinical researchers is an evolving issue with the search for the ideal index still going on. The *h*-index is considered the mainstream author-level metric due to its simple calculation, but it has also received criticisms. In fact, assessment of individual researchers should consider a broad range of bibliometric measures, not only a single indicator. Definition of *h*-index, its strengths and weaknesses, as well as variants and alternatives to this standard bibliometric indicator will be addressed. This article is intended to be a concise introduction to the author-level metrics that evaluate scholarly works.

Total publications and citations

The total number of publications reflects the author's productivity and can be obtained by sourcing information from databases, such as Web of Science (WoS), Scopus, or MEDLINE. Unfortunately, neither the type of articles (original, review, editorial, research letter, letter to the Editor, case report, etc.) nor their quality are taken into consideration by this metric. Therefore, many low-quality items (e.g., letters and case reports) can boost publication record numbers.

On the other hand, the total number of citations, also available at WoS and Scopus, is a simple measure of the impact of research that reflects the interest in the cited items. However, the number of citations is influenced by the specific research topic (i.e., it is easier for a scholar to receive citations if he/she write on a topic on which large number of articles are published) and does not consider the quality of the citing journal. Consequently, these two crude quantitative measures (total number of publications or citations) are not generally used in academia and or in the evaluation of research policy results[1].

Visual abstract available at https://www.spanishjmed.com/frame_esp.php?id=74

***Correspondence:**
José M. Porcel
E-mail: jporcelp@yahoo.es

Date of reception: 09-08-2022
Date of acceptance: 05-10-2022
DOI: 10.24875/SJMED.22000011

Available online: 09-12-2022
Span J Med. 2022;2(4):87-91
www.spanishjmed.com

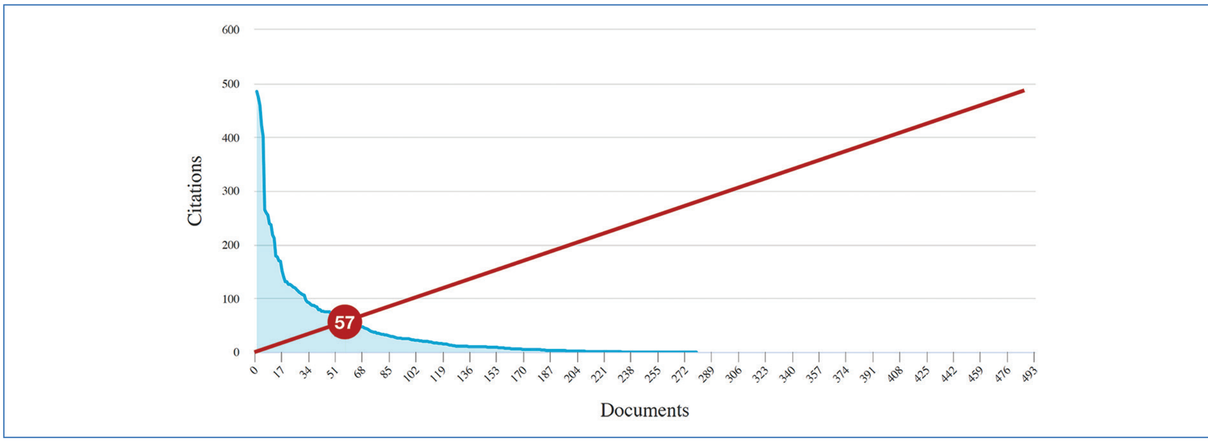

Figure 1. Graphical representation of an author's *h*-index (57).

h-index

The *h*-index is preferable to metrics that measure only a researcher's number of publications or citations received by these publications. It was created in 2005 by Jorge Eduardo Hirsch, an Argentine-American professor of physics at the University of California (San Diego, USA)[2]. *h*-index is an intersection of productivity (number of papers published) and recognition (number of citations to these papers). It was defined by Hirsch as follows: "A scientist has index *h* if *h* of his or her number of papers (N_p) have at least *h* citations each and the other (N_p - *h*) papers have ≤ *h* citations each"[2]. That is, *h* equals the number of papers that have received at least *h* citations (Fig. 1). Hence, if a scientist has currently an *h*-index of 60 (for example, its inventor Dr Hirsch), then that means he/she has 60 papers with at least 60 citations each. Both productivity and impact are required for a high *h*-index; neither a few highly cited papers nor a long list of papers with only a handful of (or no) citations will yield a high *h*-index. Hirsch's original work from 2005 has been cited 6126 times according to WoS (accessed July 31, 2022) and 6508 times according to the Scopus database (accessed July 31, 2022).

The *h*-index can be obtained through databases such as WoS, Scopus, and Google Scholar. Overall, Google Scholar returns significantly higher *h*-index scores than Scopus or WoS; the latter being the most stringent. The underlying reason is that these databases have varying coverage of publications and citations[3].

What is a good *h*-index is a matter of opinion. In general, an *h*-index of 20 characterizes a successful scientist, 40 an outstanding scientist, and 60 or more a truly exceptional one. However, this is a broad generalization and actual figures may vary enormously among different disciplines or fields and are influenced by the length of a research career. The median *h*-index of 195 laureates with the Nobel Prize in Physiology and Medicine between 1901 and 2009 was 43 (WoS) at the time of the award[4]; yet for 42 laureates from 2001 to 2017, it raised to 72.5[5]. This discrepancy may arise from the fact that scientific publications and citations grow steadily over time[6].

In 2021, nine highly cited Spanish researchers in Clinical Medicine were recognized in WoS due to their scientific production over the past decade (*h*-indexes accessed on July 31, 2022): Jesús Fernando San Miguel (Hematologists from Clínica Universitaria de Navarra, *h*-index = 120), Josep Tabernero (Oncologist from Hospital Vall d'Hebron de Barcelona, *h*-index = 114), Jordi Bruix (Hepatologist from Hospital Clinic de Barcelona, *h*-index = 113), José Luis Zamorano (Cardiologist from Hospital Universitario Ramón y Cajal, Madrid, *h*-index = 104), Enriqueta Felip (Oncologist from Hospital Vall d'Hebron de Barcelona, *h*-index = 95), Joaquin Bellmunt (Oncologist from Hospital del Mar, Barcelona; *h*-index = 86), Luis Paz-Ares (Oncologist from Hospital Universitario 12 de Octubre, Madrid, *h*-index = 86), Javier Cortés (Oncologist from International Breast Cancer Center, Barcelona, *h*-index = 80), and Maria Victoria Mateos (Hematologist from Hospital Clínico de Salamanca, *h*-index = 79). Table 1 shows current *h*-indexes of Presidents of the Sociedad Española de Medicina Interna (SEMI) over the past 30 years.

The *h*-index, however, does not capture the full history of a scientist's contributions and has several caveats[6]:

Table 1. *h*-indexes of Presidents of the Sociedad Española de Medicina Interna (SEMI)*

President (period)	*h*-index, WoS	*h*-index, Scopus
Jaume Guardia Massó (1990-1992)	38	40
Jaime Merino Sánchez (1992-1994)	14	16
José Manuel Martínez-Vázquez (1994-1996)	20	21
Ciriaco Aguirre Errasti (1996-1998)	25	18
Blas Gil Extremera (1998-2000)	30	24
Miquel Vilardell Tarrés (2000-2002)	51	53
Angel Sánchez Rodríguez (2002-2004)	25	15
Miguel Angel González de la Puente (2004-2006)	7	8
Ramon Pujol Farriols (2006-2008)	30	31
Pedro Conthe Gutiérrez (2008-2010)	24	20
Javier Garcia Alegría (2010-2012)	15	16
Pilar Roman Sánchez (2012-2014)	21	33
Emilio Casariego Vales (2014-2016)	14	14
Antonio Zapatero Gaviria (2016-2018)	20	22
Ricardo Gómez Huelgas (2018-2020)	26	28
Jesús Díez Manglano (2020-2022)	18	17

*Accessed 31 July, 2022. Note that for some authors *h*-indexes are approximate because they may have signed with different names and affiliations (some databases, such as Scopus, allow requests to merge authors).

– It does not consider issues of multiauthorship. That is, the author's position in the paper is not considered, so there is no extra credit for being the first (second or last) author, which usually indicates a greater role in the investigation. Similarly, some scientists may achieve a high *h*-index simply, because they co-author papers with other highly productive researchers, but either they occupy a less relevant position in the article or, even worse, they are "guest/gift" authorships (the author is listed solely as a gesture of respect or as an attempt to make a paper appear more credible than it is) or "coercion" authorships (a person in a position of authority uses this authority to compel another author to include him/her on a manuscript). These last two categories of authorship are contemplated as outright unethical practices.
– Comparing scientists from different disciplines or even with different research topics within the same discipline is problematic due to disparity in citation counts. For example, a researcher on human immunodeficiency virus (HIV) is more likely to achieve higher *h*-indexes than a researcher on pleural diseases, because in the past 40 years, 7 times more has been published on the former subject than on the latter. In fact, according to Expertscape (https://expertscape.com/), the world's number one and two HIV experts have WoS/Scopus *h*-indexes (accessed on July 31, 2022) of 91/103 (Kenneth H. Mayer) and 96/95 (Robert Siliciano), whereas *h*-indexes of the two top ranked experts in pleural diseases are 38/39 (Najib M. Rahman) and 39/41 (José M. Porcel), respectively.
– The *h*-index does not properly credit authors who publish few but highly influential papers. For example, a scientist A with 20 papers cited 20 times each would have an *h*-index of 20, whereas a scientist B with ten documents which were cited 200 times each would only achieve an *h*-index of 10. Scientist B publishes fewer documents, but their impact is much higher than those of scientist A. In other words, scientist A publishes twice as many articles as scientist B, although with a much lower impact. Despite this, according to the *h*-index, scientist A would be regarded as much more successful than scientist B.
– The *h*-index is influenced by the length of a scientist's career or lifetime citedness, which is disadvantageous to early career researchers that usually won't have a very high *h*-index. One rule that is widely accepted is that a respectable *h*-index score would be at least equal to the number of years that a scientist has put into his or her work.
– Self-citation practices may increase the *h*-index, although some databases (e.g., Scopus) allow calculation of the *h*-index after removing self-citations.
– The *h*-index cannot decline even if a scientist does not publish any paper after a number of active years of publication. This metric can only increase over time, so it is not able to differentiate between active and inactive researchers. In fact, the *h*-index score of deceased authors commonly raises giving the false impression of growing productivity.

h-index variants and other metrics

Variants of the *h*-index try to circumvent the shortcomings of this metric though, in general, they are regarded as superficial enhancements. Several dozen variants of *h*-index have been proposed, although only some significant ones will be commented on.

Co-authorship normalized metrics

The h_m-index was proposed by Schreiber in 2008 and is determined in a similar way to the *h*-index, but counting the number of papers fractionally based on the number of authors[7]. A weight is assigned to each publication as the inverse of the number of authors (e.g., 1 author paper = 1, 2 author paper = 0.5, and 3 author paper 0.33). The h_m-index would be the effective number of papers that have been cited h_m or more times. If a researcher only publishes sole author papers, h-index equals to h_m-index, but, in all other cases, h_m-index will be less than *h*-index. This alternative metric, however, does not distinguish between different author positions and is not offered by the standard databases (i.e., it must be calculated manually).

Other co-authorship-adjusted metrics are the 'profit (p)-index' (which accounts for number of co-authors and the sequence of authors on the paper)[8] and the *h*-fac index (which ponders positively the commitment and participation of the first author)[9].

Evaluating highly cited scientists

The *g*-index, which was introduced by Leo Egghe, credits authors of highly cited papers[10]. To compute this index, the citations are considered in a descending order. The resultant score is the largest number of top "*g* articles" receiving together at least g^2 citations. Hence, a *g*-index of 10 indicates that the top ten publications of an author have received at least a total of 100 citations (10^2). The *g*-index is always higher than the *h*-index and is particularly helpful for comparing researchers with identical *h*-index. For example, an author with 10 published articles, three of which are cited 60, 30 and 10 times (100 in total), will have a g-index of 10 and an *h*-index of 3. The *g*-index can be calculated on the Harzing Publish or Perish website (https://harzing.com/resources/publish-or-perish) using data from Google Scholar or subscription citation databases.

Another complement to the *h*-index for evaluating highly cited scientists or for comparing the scientific output of scientists having an identical *h*-index is the so-called *e*-index, but its calculation is complex[11].

Time-adjusted metrics

The *m*-quotient corrects the *h*-index for career length, thus facilitating comparisons between scientists with different periods of academic activity. It is calculated by dividing a scientist's *h*-index by the number of years that have passed since the first publication, with a score of 1 being very good indeed, 2 being outstanding, and 3 truly exceptional. Thus, in a person with 20 years of research experience, an *h*-index of 20 (i.e., *m*-quotient = 1) is very good, 40 is great (i.e., *m*-quotient = 2), and 60 (i.e., *m*-quotient = 3) is remarkable. However, the first publication is not always the start of an active career in a specific field. In a similar way, the *m*-quotient overlooks temporal interruptions in an individual career (e.g., breaks in academic publications for parental leave).

Other indexes that take into account time include the contemporary *h*-index (considers the age of an article)[12] or the timed *h*-index (compares activity at various 5-year time points)[13]. Some databases, such as Google Scholar, calculate author's metrics (*h*-index) for the past 5 years.

Other metrics

i10-index

It refers to the number of publications with at least ten citations. This simple and straightforward measure is only used by Goggle Scholar, but can be easily calculated from data provided by WoS or Scopus. Likewise, other similar indexes could be derived, such as i15, i20, i25, and i30. It can be used to highlight productivity over impact (i10) or impact over productivity (e.g., i50, i100, and i200).

Author impact factor

The author impact factor (AIF) is the analogue for authors of what the impact factor is for journals[3]. Basically, the AIF is a dynamic index which expresses the current impact of papers published by authors in recent years, and therefore, it may capture trends and variations of the performance of scholars along their careers[14]. The AIF of an author *A* for a year *t* is the average number of citations given by papers published in year *t* to papers published by *A* in a given time window (e.g., 2-5 years) before year *t*. Unfortunately, this metric has not been implemented on bibliographic portals. Erroneously, some scientists think the AIF is calculated by summing up the impact factor of each journal in which they have published an article.

Conclusions

The *h*-index is the mainstream author-level bibliometric indicator. It combines quantitative (publication counts) and impact (citation counts) data into a single whole number. However, it does not allow to compare scientists of different seniority or disciplines, nor it does consider the position of the author within the author list. Particularly, researchers working in non-mainstream areas will have lower *h*-values than those working in highly topical areas. Many variants of the *h*-index with interesting and unique attributes have been proposed which attempt to correct some of its limitations. Nevertheless, they have not attracted much attention, primarily due to their complex formula calculations and the lack of availability in common bibliographic sources.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Ethical disclosures

Protection of human and animal subjects: The authors declare that no experiments were performed on humans or animals for this study.

Confidentiality of data: The authors declare that no patient data appear in this article.

Right to privacy and informed consent: The authors declare that no patient data appear in this article.

References

1. Gasparyan AY, Yessirkepov M, Duisenova A, Trukhachev VI, Kostyukova EI, Kitas GD. Researcher and author impact metrics: variety, value, and context. J Korean Med Sci. 2018;33:e139.
2. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A. 2005;102:16569-72.
3. Porcel JM, Liesa L. The basics of journal-level metrics: what clinical researchers need to know. Span J Med. 2022;2:1-7.
4. Patel VM, Ashrafian H, Almoudaris A, Makanjuolam J, Bucciarelli-Ducci C, Darzi A, et al. Measuring academic performance for healthcare researchers with the H index: which search tool should be used? Med Princ Pract. 2013;22:178-83.
5. Schreiber WE, Giustini DM. Measuring scientific impact with the h-index: a primer for pathologists. Am J Clin Pathol. 2019;151:286-91.
6. Quaia E, Vernuccio F. The H index myth: a form of fanaticism or a simple misconception? Tomography. 2022;8:1241-3.
7. Schreiber M. To share the fame in a fair way, hm modifies h for multi-authored manuscripts. New J Physics. 2008;10:040201.
8. Aziz NA, Rozing MP. Profit (p)-index: the degree to which authors profit from co-authors. PLoS One. 2013;8:e59814.
9. Júnior FG, Godoy MF, Gregori FF. Proposal of an individual scientometric index with emphasis on ponderation of the effective contribution of the first author: h-fac índex. Rev Bras Cir Cardiovasc. 2012;27:370-6.
10. Egghe L. Theory and practice of the g-index. Scientometrics. 2006;69:131-52.
11. Zhang CT. The e-index, complementing the h-index for excess citations. PLoS One. 2009;4:e5429.
12. Sidiropoulos A, Katsaros D, Manolopoulos Y. Generalized Hirsch h-index for disclosing latent facts in citation networks. Scientometrics. 2007;72:253-80.
13. Schreiber M. Restricting the h-index to a publication and citation time window: a case study of a timed Hirsch index. J Informetr. 2015;9:150-5.
14. Pan RK, Fortunato S. Author impact factor: tracking the dynamics of individual scientific impact. Sci Rep. 2014;4:4880.