The \( h \)-index is an index that attempts to measure both the productivity and citation impact of the published body of work of a scientist or scholar (an author-level metric). The index is based on the set of the scientist's most cited papers and the number of citations that they have received in other publications. The index can also be applied to the productivity and impact of a scholarly journal\(^\text{[1]}\) as well as a group of scientists, such as a department or university or country.\(^\text{[2]}\) The index was suggested in 2005 by Jorge E. Hirsch, a physicist at UCSD, as a tool for determining theoretical physicists' relative quality\(^\text{[3]}\) and is sometimes called the \textit{Hirsch index} or \textit{Hirsch number}.\n
### Definition and purpose

The index is based on the distribution of citations received by a given researcher's publications. Hirsch writes:

A scientist has index \( h \) if \( h \) of his/her \( N_p \) papers have at least \( h \) citations each, and the other \( (N_p - h) \) papers have no more than \( h \) citations each.

In other words, a scholar with an index of \( h \) has published \( h \) papers each of which has been cited in other papers at least \( h \) times.\(^\text{[4]}\) Thus, the \( h \)-index reflects both the number of publications and the number of citations per publication. The index is designed to improve upon simpler measures such as the total number of citations or publications. The index works properly only for comparing scientists working in the same field; citation conventions differ widely among different fields.

### Calculation
Formally, if $f$ is the function that corresponds to the number of citations for each publication, we compute the $h$ index as follow. First we order the values of $f$ from the largest to the lowest value. Then, we look for the last position in which $f$ is greater than or equal to the position (we call $h$ this position).

For example, if we have a researcher with 5 publications A, B, C, D, and E with 10, 8, 5, 4, and 3 citations, respectively, the $h$ index is equal to 4 because the 4th publication has 4 citations and the 5th has only 3. On the contrary if the same publications have 25, 8, 5, 3, and 3, then the index is 3 because the fourth paper has only 3 citations.

$$f(A)=10, f(B)=8, f(C)=5, f(D)=4, f(E)=3 \rightarrow h\text{-index}=4$$
$$f(A)=25, f(B)=8, f(C)=5, f(D)=3, f(E)=3 \rightarrow h\text{-index}=3$$

If we have the function $f$ ordered in decreasing order from the largest value to the lowest one, we can compute the $h$ index as follows:

$$h\text{-index}(f) = \max_i \min(f(i), i)$$

The $h$ index can be seen as the Sugeno integral (a type of fuzzy integral).[5] Then, the most common index of number of citations of an author can be seen as a Choquet integral of the same function $f$.

The Hirsch index is equivalent to the Eddington index to evaluate cyclists. The $h$-index serves as an alternative to more traditional journal impact factor metrics in the evaluation of the impact of the work of a particular researcher. Because only the most highly cited articles contribute to the $h$-index, its determination is a simpler process. Hirsch has demonstrated that $h$ has high predictive value for whether a scientist has won honors like National Academy membership or the Nobel Prize. The $h$-index grows as citations accumulate and thus it depends on the "academic age" of a researcher.

## Input data

The $h$-index can be manually determined using citation databases or using automatic tools. Subscription-based databases such as Scopus and the Web of Knowledge provide automated calculators. Harzing's Publish or Perish program calculates the $h$-index based on Google Scholar entries. In July 2011 Google trialled a tool which allows scholars to keep track of their own citations and also produces an $h$-index and an i10-index.[6] In addition, specific databases, such as the INSPIRE-HEP database can automatically calculate the $h$-index for researchers working in high energy physics.

Each database is likely to produce a different $h$ for the same scholar, because of different coverage.[7] A detailed study showed that the Web of Knowledge has strong coverage of journal publications, but poor coverage of high impact conferences. Scopus has better coverage of conferences, but poor coverage of publications prior to 1996; Google Scholar has the best coverage of conferences and most journals (though not all), but like Scopus has limited coverage of pre-1990 publications.[8][9] The exclusion of conference proceedings papers is a particular problem for scholars in computer science, where conference proceedings are considered an important part of the literature.[10] Google Scholar has been criticized for producing "phantom citations," including gray literature in its citation counts, and failing to follow the rules of Boolean logic when combining search terms.[11] For
example, the Meho and Yang study found that Google Scholar identified 53% more citations than Web of Knowledge and Scopus combined, but noted that because most of the additional citations reported by Google Scholar were from low-impact journals or conference proceedings, they did not significantly alter the relative ranking of the individuals. It has been suggested that in order to deal with the sometimes wide variation in $h$ for a single academic measured across the possible citation databases, one should assume false negatives in the databases are more problematic than false positives and take the maximum $h$ measured for an academic.[12]

**Results across disciplines and career levels**

Hirsch suggested (with large error bars) that, for physicists, a value for $h$ of about 12 might be typical for advancement to tenure (associate professor) at major research universities. A value of about 18 could mean a full professorship, 15–20 could mean a fellowship in the American Physical Society, and 45 or higher could mean membership in the United States National Academy of Sciences.[13]

The London School of Economics found that (full) professors in the social sciences had average $h$-indices ranging from 2.8 (in law), through 3.4 (in political science), 3.7 (in sociology), 6.5 (in geography) and 7.6 (in economics). On average across the disciplines, a full professor in the social sciences had an $h$-index about twice that of a lecturer or a senior lecturer (associate professor), though the difference was the smallest in geography.[14]

Among the 22 scientific disciplines listed in the Thomson Reuters Essential Science Indicators Citation Thresholds, physics has the second most citations after space science.[15] During the period January 1, 2000 – February 28, 2010, a physicist had to receive 2073 citations to be among the most cited 1% of physicists in the world.[15] The threshold for space science is the highest (2236 citations), and physics is followed by clinical medicine (1390) and molecular biology & genetics (1229). Most disciplines, such as environment/ecology (390), have fewer scientists, fewer papers, and fewer citations.[15] Therefore, these disciplines have lower citation thresholds in the Essential Science Indicators, with the lowest citation thresholds observed in social sciences (154), computer science (149), and multidisciplinary sciences (147).[15]

Little systematic investigation has been made on how academic recognition correlates with $h$-index over different institutions, nations and fields of study. However, Hirsch estimates that after 20 years a "successful scientist" will have an $h$-index of 20, an "outstanding scientist" an $h$-index of 40, and a "truly unique" individual an $h$-index of 60. However, he points out that values of $h$ will vary between different fields.[16]

For the most highly cited scientists in the period 1983–2002, Hirsch identified the top 10 in the life sciences (in order of decreasing $h$): Solomon H. Snyder, $h = 191$; David Baltimore, $h = 160$; Robert C. Gallo, $h = 154$; Pierre Chambon, $h = 153$; Bert Vogelstein, $h = 151$; Salvador Moncada, $h =143$; Charles A. Dinarello, $h =138$; Tadamitsu Kishimoto, $h =134$; Ronald M. Evans, $h =127$; and Axel Ullrich, $h = 120$. Among 36 new inductees in the National Academy of Sciences in biological and biomedical sciences in 2005, the median $h$-index was 57.[3]

**Advantages**

Hirsch intended the $h$-index to address the main disadvantages of other bibliometric indicators, such as total number of papers or total number of citations. Total number of papers does not account for the quality of scientific publications, while total number of citations can be disproportionately affected by participation in a single publication of major influence (for instance, methodological papers proposing successful new techniques, methods or approximations, which can generate a large number of citations), or having many publications with
few citations each. The *h*-index is intended to measure simultaneously the quality and quantity of scientific output.

## Criticism

There are a number of situations in which *h* may provide misleading information about a scientist's output:[17] (However, most of these are not exclusive to the *h*-index.)

- The *h*-index does not account for the typical number of citations in different fields. It's been stated that citation behavior in general is affected by field-dependent factors,[18] which may invalidate comparisons not only across disciplines but even within different fields of research of one discipline.[19]
- The *h*-index discards the information contained in author placement in the authors' list, which in some scientific fields is significant.[20][21]
- The *h*-index has been found to have slightly less predictive accuracy and precision than the simpler measure of mean citations per paper.[22] However, this finding was contradicted by another study.[23]
- The *h*-index is a natural number that reduces its discriminatory power. Ruane and Tol therefore propose a rational *h*-index that interpolates between *h* and *h* + 1.[24]
- The *h*-index can be manipulated through self-citations,[25][26] and if based on Google Scholar output, then even computer-generated documents can be used for that purpose, e.g. using SCtgen.[27]
- The *h*-index does not provide a significantly more accurate measure of impact than the total number of citations for a given scholar. In particular, by modeling the distribution of citations among papers as a random integer partition and the *h*-index as the Durfee square of the partition, Yong[28] arrived at the formula \( h \approx 0.54 \sqrt{N} \), where \( N \) is the total number of citations, which, for mathematicians, turns out to provide a highly accurate approximation of *h*-index in most cases.

## Alternatives and modifications

*Further information: Author-level metrics*

Various proposals to modify the *h*-index in order to emphasize different features have been made.[29][30][31][32][33][34] As the variants have proliferated, comparative studies have become possible showing that most proposals are highly correlated with the original *h*-index,[35] although alternative indexes may be important to decide between comparable CVs, as often the case in evaluation processes.

- An individual *h*-index normalized by the average number of co-authors in the *h*-core has been proposed.[29] It was found that the distribution of the *h*-index, although it depends on the field, can be normalized by a simple rescaling factor. For example, assuming as standard the *hs* for biology, the distribution of *h* for mathematics collapse with it if this *h* is multiplied by three, that is, a mathematician with *h* = 3 is equivalent to a biologist with *h* = 9. This method has not been readily adopted, perhaps because of its complexity. It might be simpler to divide citation counts by the number of authors before ordering the papers and obtaining the *h*-index, as originally suggested by Hirsch.
- The *m*-index is defined as \( h/n \), where *n* is the number of years since the first published paper of the scientist;[3] also called *m*-quotient.[36][37]
- There are a number of models proposed to incorporate the relative contribution of each author to a paper, for instance by accounting for the rank in the sequence of authors.[38]
- A generalization of the *h*-index and some other indices that gives additional information about the shape
of the author's citation function (heavy-tailed, flat/peaked, etc.) has been proposed.[39]

- A successive Hirsch-type-index for institutions has also been devised.[40][41] A scientific institution has a successive Hirsch-type-index of \( i \) when at least \( i \) researchers from that institution have an \( h \)-index of at least \( i \).

- Three additional metrics have been proposed: \( h^2 \) lower, \( h^2 \) center, and \( h^2 \) upper, to give a more accurate representation of the distribution shape. The three \( h^2 \) metrics measure the relative area within a scientist's citation distribution in the low impact area, \( h^2 \) lower, the area captured by the \( h \)-index, \( h^2 \) center, and the area from publications with the highest visibility, \( h^2 \) upper. Scientists with high \( h^2 \) upper percentages are perfectionists, whereas scientists with high \( h^2 \) lower percentages are mass producers. As these metrics are percentages, they are intended to give a qualitative description to supplement the quantitative \( h \)-index.[42]

- The \( g \)-index can be seen as the \( h \)-index for an averaged citations count.[43]

- It has been argued that "For an individual researcher, a measure such as Erdős number captures the structural properties of network whereas the \( h \)-index captures the citation impact of the publications. One can be easily convinced that ranking in coauthorship networks should take into account both measures to generate a realistic and acceptable ranking." Several author ranking systems such as eigenfactor (based on eigenvector centrality) have been proposed already, for instance the Phys Author Rank Algorithm.[44]

- The \( c \)-index accounts not only for the citations but for the quality of the citations in terms of the collaboration distance between citing and cited authors. A scientist has \( c \)-index \( n \) if \( n \) of [his/her] \( N \) citations are from authors which are at collaboration distance at least \( n \), and the other \((N-n)\) citations are from authors which are at collaboration distance at most \( n \).[45]

- An \( s \)-index, accounting for the non-entropic distribution of citations, has been proposed and it has been shown to have a very good correlation with \( h \).[46]

- The \( e \)-index, the square root of surplus citations for the \( h \)-set beyond \( h^2 \), complements the \( h \)-index for ignored citations, and therefore is especially useful for highly cited scientists and for comparing those with the same \( h \)-index (iso-\( h \)-index group).[47][48]

- Because the \( h \)-index was never meant to measure future publication success, recently, a group of researchers has investigated the features that are most predictive of future \( h \)-index. It is possible to try the predictions using an online tool.[49] However, later work has shown that since \( h \)-index is a cumulative measure, it contain intrinsic auto-correlation that led to significant overestimation of its predictability. Thus, the true predictability of future \( h \)-index is much lower compared to what has been claimed before.[50]

- The \( h \)-index has been applied to Internet Media, such as YouTube channels. The \( h \)-index is defined as the number of videos with \( \geq h \times 10^5 \) views. When compared with a video creator's total view count, the \( h \)-index and \( g \)-index better capture both productivity and impact in a single metric.[51]

- The \( i^{10} \)-index indicates the number of academic publications an author has written that have at least ten citations from others. It was introduced in July 2011 by Google as part of their work on Google Scholar.[52]

- The \( h \)-index has been shown to have a strong discipline bias. However, a simple normalization \( \frac{h}{\langle h \rangle_d} \) by the average \( h \) of scholars in a discipline \( d \) is an effective way to mitigate this bias, obtaining a universal impact metric that allows to compare scholars across different disciplines.[53] Of course this method does not deal with academic age bias.

- The \( h \)-index can be timed to analyze its evolution during one's career, employing different time windows.[54]

See also
- Bibliometrics
- Comparison of Research Networking Tools and Research Profiling Systems
- Eddington number (for cycling), an earlier metric of the same form

References


Further reading


**External links**

- H-index for economists (http://ideas.repec.org/top/top.person.hindex.html)
- The H-index for computer science (http://www.cs.ucla.edu/~palsberg/h-number.html)


Categories: Citation metrics | Academic publishing | Index numbers

- This page was last modified on 7 September 2015, at 11:43.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.