The h-index is an index that attempts to measure both the productivity and impact of the published work of a scientist or scholar. 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 group of scientists, such as a department or university or country, as well as a scholarly journal. The index was suggested by Jorge E. Hirsch, a physicist at UCSD, as a tool for determining theoretical physicists' relative quality and is sometimes called the Hirsch index or Hirsch number.

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 Np papers have at least h citations each, and the other (Np − 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. 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.
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 relatively 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.

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.\[3\] Little systematic investigation has been made on how academic recognition correlates with $h$-index over different institutions, nations and fields of study.

Among the 22 scientific disciplines listed in the Thomson Reuters Essential Science Indicators Citation Thresholds, physics has the second most citations after space science.\[4\] 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.\[4\] 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.\[4\] 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).\[4\]

## Calculation

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 a $h$-index and an i10-index.\[5\] Each database is likely to produce a different $h$ for the same scholar, because of different coverage: Google Scholar has more citations than Scopus and Web of Science but the smaller citation collections tend to be more accurate. In addition, specific databases, such as the INSPIRE-HEP database can automatically calculate $h$-index for researchers working in High Energy Physics.

The topic has been studied in detail by Lokman I. Meho and Kiduk Yang.\[6\][7] Web of Knowledge was found to have 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.\[7\] The exclusion of
conference preprints is a problem for scholars in computer science, where conference preprints are considered an important part of the literature,[8] but reflects common practice in most scientific fields where conference preprints are unrefereed and are accorded less weight in evaluating academic productivity.

[citation needed] With electronic pre-publication and very long printing lags for some journals, these 'in press' citations can be considerable. 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. [9] 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, that one should assume false negatives in the databases are more problematic than false positives and take the maximum $h$ measured for an academic. [10]

**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:[11] (However, most of these are not exclusive to the $h$-index.)

- The $h$-index does not account for the number of authors of a paper. In the original paper, Hirsch suggested partitioning citations among co-authors. Even in the absence of explicit gaming, the $h$-index and similar indexes tend to favor fields with larger groups, e.g. experimental over theoretical.
- The $h$-index does not account for the typical number of citations in different fields. Different fields, or journals, traditionally use different numbers of citations.
- The $h$-index discards the information contained in author placement in the authors' list, which in some scientific fields is significant.[12][13]
- The $h$-index is bounded by the total number of publications. This means that scientists with a short career are at an inherent disadvantage, regardless of the importance of their discoveries. For example, Évariste Galois' $h$-index is 2, and will remain so forever. Had Albert Einstein died after publishing his four groundbreaking Annus Mirabilis papers in 1905, his $h$-index would be stuck at 4 or 5. This is also a problem for any measure that relies on the number of publications. However, as Hirsch indicated in the original paper, the index is intended as a tool to evaluate researchers in the same stage of their careers. It is not meant as a tool for historical comparisons.
- The $h$-index does not consider the context of citations. For example, citations in a paper are often made simply to flesh out an introduction, otherwise...
having no other significance to the work. $h$ also does not resolve other contextual instances: citations made in a negative context and citations made to fraudulent or retracted work. This is also a problem for regular citation counts.

- The $h$-index gives books the same count as articles making it difficult to compare scholars in fields that are more book-oriented such as the humanities.
- The $h$-index does not account for confounding factors such as "gratuitous authorship", the so-called Matthew effect, and the favorable citation bias associated with review articles. Again, this is a problem for all other metrics using publications or citations.
- The $h$-index has been found to have slightly less predictive accuracy and precision than the simpler measure of mean citations per paper.\[14\] However, this finding was contradicted by another study.\[15\]
- The $h$-index is a natural number which reduces its discriminatory power. Ruane and Tol therefore propose a rational $h$-index that interpolates between $h$ and $h + 1$.\[16\]
- The $h$-index can be manipulated through self-citations,\[17\] and if based on Google Scholar output, then even computer-generated documents can be used for that purpose, e.g. using SCIgen.\[18\]

### Alternatives and modifications

Various proposals to modify the $h$-index in order to emphasize different features have been made.:\[19\][20][21][22][23][24] As the variants have proliferated, comparative studies have become possible and they demonstrate that most proposals do not differ significantly from the original $h$-index as they remain highly correlated with it.\[25\]

- An individual $h$-index normalized by the average number of co-authors in the $h$-core has been introduced by Batista et al.\[19\] They also 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 $h$s 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 $hn$, where $n$ is the number of years since the first published paper of the scientist,\[11\] also called $m$-quotient.\[26\][27]
- 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.) was proposed by Gałolewski and Grzegorzezki.\[28\]
- Successive Hirsch-type-index introduced independently by Kosmulska\[29\] and Prathap.\[30\] 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 $i$.
- Bornmann, Mutz, and Daniel recently proposed three additional metrics, $h^2_{\text{lower}}$, $h^2_{\text{center}}$, and $h^2_{\text{upper}}$, 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, $h^2_{\text{lower}}$, the area captured by the $h$-index, $h^2_{\text{center}}$, and the area from publications with the highest visibility, $h^2_{\text{upper}}$. Scientists with high $h^2_{\text{upper}}$ percentages are perfectionists, whereas scientists with high $h^2_{\text{lower}}$ percentages are mass producers. As these metrics are percentages, they are intended to give a qualitative description to supplement the quantitative $h$-index.\[31\]
The g-index proposed by Egghe\textsuperscript{[32]} can be seen as the \(h\)-index for an averaged citations count.

K. Dixit and colleagues argue 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.\textsuperscript{[33]}

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 \(n\).\textsuperscript{[34]}

An \(s\)-index, accounting for the non-entropic distribution of citations, has been proposed and it has been shown to be in a very good correlation with \(h\)\textsuperscript{[35]}

The \(e\)-index, 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 (isohindex group).\textsuperscript{[36][37]}

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 (Nature article (http://www.nature.com/nature/journal/v489/n7415/full/489201a.html) ). It is possible to try the predictions using an online tool (http://klab.smpp.northwestern.edu/h-index.html)

See also

- Bibliometrics
- Eddington number (cycling) An earlier metric of the same form.
- Durfee square, a quantity defined in the same way for integer partitions
- Comparison of Research Networking Tools and Research Profiling Systems

References


5. \(^{a}\) Google Scholar Citations Help (http://scholar.google.com/intl/en/scholar)


Further reading


doi:10.1038/436900a.


doi:10.1038/4441003a.


doi:10.1371/journal.pone.0000127.


External links

Lists of h-indices

- The H-index for economists
- The H-index for computer science
