The H-index is an unfair measure of scientific achievements. A proposal to address its shortcomings

Giulio Formoso

Local Health Authority of Reggio Emilia

Via Amendola 2; 42122 Reggio Emilia (Italy)

E-mail: giulio.formoso@ausl.re.it; tel +39 0522335954

Conflict of interest statement: I declare that I do not have conflicts of interest Funding:

none

Abstract

Citation impact indicators play a relevant role in the evaluation of researchers' scientific production and can influence research funding and future research outputs. The H-index is widely used in this regard, in spite of several shortcomings such as not considering the actual contribution of each author, the number of authors, their overall scientific production and the scientific quality of citing articles. Several authors have highlighted some of these limits. Alternative systems have been proposed but have gained less fortune.

In order to show that fairer criteria to assess researchers' scientific impact can be achieved, a workable example is presented through a novel method, integrating the aforementioned elements by using information available in bibliographic databases.

A better, merit-based proxy measure is then warranted and can be achieved, bearing in mind that a perfect score without any shortcoming is a chimera. Any proposal on a new measure would require clear reasoning, easy math and a consensus between publishers, considering researchers' and research funders' point of view. In any case, the relevance of authors' scientific achievements cannot be adequately represented by a quantitative index only, and qualitative judgements are also necessary. But the time is ripe to make decisions on a fairer, although proxy, measure of scientific outputs.

Text

Introduction

Citation impact indicators play a relevant role in the evaluation of researchers' scientific production. In 2005 the physicist Jorge Hirsch proposed an index, the H-index, aimed at measuring the scientific output of researchers.[1] Technically, it measures the number of papers cited at least as many times from other articles published in indexed journals. As an example, a researcher with H-index =14 means he/she published 14 articles having at least 14 citations.

Since its proposal, the H-index has been more and more criticized, even by Hirsch himself.[2-4] For example, several Nobel laureates score relatively low on it and are not listed among the top 1,000 authors on total citations. Conversely, many others who do not have first/last-authored papers rank in a higher position.[5] Such kind of distorsions can happen because of several shortcomings of this index:

- all authors in a paper are weighted equally, independently of their actual contribution; each one gets the same credit for the paper being cited, either if she/he conceived and written it (generally the first author) or contributed less or even minimally;
- the number of authors is irrelevant (it does not make any difference being eight out of nine authors or the first of four authors), a situation possibly promoting the inclusion of some not having sufficiently contributed;
- citations by articles published in low-impact journals, possibly with limited scientific rigour, have the same weight as those in higher-impact journals;
- papers having lower number of citations than the H-index do not contribute to it, as well as citations over and above the H-index;
- self citations will bias its computation.

Alternative systems have been proposed, for example using either arithmetic or geometric distribution of each author's contribution to published articles.[6] As a matter of fact, these scores have gained less fortune: they may be considered less practical than the easily understandable Hindex.

Research funding and future research outputs can be influenced by such scoring systems. For example, principal investigators submitting research proposals to the Italian Ministry of Health

should have an H-index of at least 10 (or at least 18, depending on the type of proposal).[7]

Measuring scientific achievement: how, ideally?

A fairer measure of scientific achievement would require a wide consensus among publishers, researchers and funders. It may be possible to favour this consensus by explaining, with clear reasoning and easy math, the feasibility of new methods. All the above-mentioned elements can be put together, as in the proposal below trying to explain them as clearly as possible. The goal would be to make each article contribute to each author's score – therefore considering his/her overall scientific production, but also his/her contribution to the article. Since not all citing articles have the same scientific weight, a measure for this weight should be incorporated too. Ideally, such a score would depend on: number of authors in the article, position in the list of authors (for those disciplines where authors are not listed alphabetically), number of citations that the article receives, weight of these citations (considering for example the impact factor – IF - of the citing journals). In this way, each citation of each authored article would contribute to the overall score for that author. Automated indexing systems have all the necessary information to develop a quite simple algorithm such as the one proposed below, as an example.

Methods: some (simple) math

Let's make the example of an article with five authors. The article score would depend on the number of citations that article receives (possibly weighted by the relevance of the citing articles). This score should be divided among the five authors - therefore the more the authors, the fewer points each author gets - and benefit those higher in the list. Below is some simple math with an equation for such an example:

$y+(y-x)+(y-2x)+(y-3x)+(y-4x)=z$ where:

y= first author's score x= linear score reduction for each subsequent author z= number of times the article is cited.

If z is weighted by the relevance of the citing articles:

$z=IF_1+IF_2+\ldots+IF_n$, where $IF_{1,2\ldots n}$ are the impact factors of the journals (at the time of citation) where the citing articles 1,2…n are published. Journal IF from past publications (e.g. before 1975, when the IF system started to be implemented) may not be attainable, but such weighting could be considered at least for research purposes.

Box 1 provides a general representation of this equation (n= number of authors).

Box 1. General equation for the proposed score $y+(y-x)+(y-2x)+\ldots+[yn-(n-1)x]=z$

or

$$ny-\sum_{i=1}^{n-1} ix = z$$

That is: $ny–n/2(n-1)x=z$

If - as in the current system - all authors had the same score, that would be z/n (i.e. the total score of the article divided by the number of authors). In a system trying to differentiate authors in this regard, it is necessary to establish a "bonus" that could be assigned to the first author, versus a situation of equal subdivision of the article score. In general terms: y=bz/n where b= bonus coefficient for the first author, which must be >1 to be rewarding. According to this, box 2 shows the math to get to each author's scores, according to a linear reduction.

Box 2. Obtaining the coefficient for the linear reduction of each author's score

Setting up a system of equations:

$ny–n/2(n-1)x=z$ (see box 1)

$y=bz/n$

Solving it: $x=2z(b-1)/n(n-1)$

The final equation in box 2 cannot be applied only when there is a single author: in that case the author's score would be equal to the number of citations or weighted citations.

To have scores >0 for all authors, b must also be <2. Box 3 shows the math.

Box 3. Explaining the upper limit of the "bonus" coefficient

y-(n-1)x>0 (score of the last author >0) therefore,

considering that y=bz/n and x=2z(b-1)/n(n-1):

bz/n–[2z(b-1)(n-1)]/[n(n-1)]>0

Solving we get the condition 2-b>0.

This proposal is similar in principle to other ones using weights based on arithmetic counting.[3]

By applying the equations above to the previous example (5 authors) and setting the bonus for the first author = 1.5, for an article with a score of 10 (i.e. with 10 citations), this score would be distributed among the authors as follows:

y=1.5*10/5=3 x=2*10*(1.5-1)/5*(5-1)=10/20=0.5

Therefore, the highest score would be =3 points, the second one =2.5, the third one =2, the fourth one =1.5, the fifth one =1 (total =10).

If the number of citations (z) were 6 and the authors were 9: y=1.5*6/9=1; x=1/12. The highest score would be =1 and then scaling by 1/12 down to 4/12, which would be the lowest score.

Moving b towards 2 would naturally lead to a greater score differentiation between the authors (more rewarding for the first one and for those in a high position). Which value of b may represent the most "balanced" one is a matter that could be discussed.

It would be appropriate to assign the second best score to the last author, who is often the research group leader, going on with the subsequent scores starting from the second author onwards. As for disciplines where authors are listed alphabetically, a score differentiation may be considered for the first (and, in case, for the last) author only.

Discussion: what next?

The time is ripe to make decisions on a fairer and more meritocratic measure of scientific outputs, considering the relevant limits of the H-index. Even comparison is also seen in twitter indicators versus traditional citation based indicators (Sakib, 2021). Information available in bibliographic

databases can certainly allow the implementation of alternative scoring systems, using algorithms such as the one hypothesized above. The latter is just an example that limits of the H-index can be at least partially overcome, bearing in mind that a perfect score without any shortcoming is a chimera. The relevance of authors' scientific output cannot be adequately represented by a quantitative index only. Their position in the authors' list (when they are not listed alphabetically), the number of authors in a paper, the overall and weighted number of citations can just help provide a fairer proxy measure than the current one. Possible drawbacks on willingness of researchers to collaborate cannot be feared, as far as the collaboration is productive and mutually useful rather than aimed at simply inflating H-indexes of researchers not complying with criteria for authorship.[8,9] In any case, when researchers' scientific achievements need to be assessed, for example when deciding grant assignments, qualitative and subjective judgements are also needed. A score can provide some sort of proxy and the proposal above provides a workable example of the main elements that may be considered in an updated scoring system.

Acknowledgements

The author wants to thank Prof. Emilio Cirillo (University La Sapienza, Rome) for his constructive feedback

References

Hirsch JE. An index to quantify an individual's scientific research output. *PNAS* 2005;**102**:16569-72

Bornmann L, Mutz R, Daniel HD. Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. Journal of the American Society for Information Science and Technology. 2008;59(5):830–837.

Hirsch JE. Superconductivity, What the H? The Emperor Has No Clothes. *Physics and Society* 2020;**49**:4-9

Koltun V, Hafner D (2021) The h-index is no longer an effective correlate of scientific reputation. PLoS ONE 16(6): e0253397. https://doi.org/10.1371/journal.pone.0253397

Ioannidis JPA, Klavans R, Boyack KW (2016). Multiple Citation Indicators and Their Composite across Scientific Disciplines. *PLoS Biol* **14**(7): e1002501

Waltman L. A review of the literature on citation impact indicators. *Journal of Informetrics* 2016;**10**:365-91

Ministero della Salute. Bando della Ricerca Finalizzata 2021. https://www.salute.gov.it/imgs/C_17_bandi_283_0_file.pdf (last access: Mar 09, 2022)

International Committee of Medical Journal Editors. Defining the Role of Authors and Contributors http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html (last access: Mar 09, 2022)

Resnik DB, Tyler AM, Black JR, et al. Authorship policies of scientific journals. *J Med Ethics* 2016;42:199-202

Sakib, SM Nazmuz. "Comprehensive Examination of Novel Altmetrics Data in Relation to Citation Based Impact: A Case Study in the Field of Information Sciences."