Open Peer Review on Qeios

Measuring researchers' success more fairly: going beyond the H-index

Giulio Formoso[1]

1 AUSL di Reggio Emilia

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

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 warranted and can be achieved, although a perfect score without shortcomings 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.

Giulio Formoso

Local Health Authority / IRCCS of Reggio Emilia

Via Amendola 2; 42122 Reggio Emilia (Italy)

Email: giulio.formoso@ausl.re.it

Telephone: +39 0522335954

ORCID: https://orcid.org/0000-0002-1819-6887

Keywords: H-index, citation impact indicators, researchers' scientific output, authorship, scientific quality.

Citation impact indicators play a relevant role in the evaluation of researchers' scientific production. The H-index is an easily understandable system for assigning a score to the scientific output of researchers. It was proposed by the physicist Jorge Hirsch in 2005 and represents the number of articles with at least as many citations received from other scientific articles published in indexed journals (Hirsch, 2005). For example, a researcher with H-index =20 means he/she published 20 articles having at least 20 citations.

Being an indicator of researchers' professional success, it can also have a relevant impact on research funding and on future research outputs. As an example, in Italy an H-index of at least 10 (or at least 18, depending on the type of proposal) is required to principal investigators to compete for research grants by the Italian Ministry of Health (Ministero della Salute, 2021).

However, by taking only the number of article citations into account, this mechanism is deficient since:

- it does not consider the actual contribution of each author, frequently related to their position in the authors' list, so that researchers who may have contributed minimally count as those conceiving and writing the article (e.g. first ones in the authors' list count as penultimate ones);
- it does not consider the number of authors (e.g. being nineteenth out of 20 authors has the same impact as being the first of 3 authors), a circumstance also encouraging inflated numbers of authors in addition to inflating H-indexes of researchers not complying with actual criteria for authorship (e.g. in group authorship, but not only there) (International Committee of Medical Journal Editors; Resnik et al., 2016);
- in an editorial landscape where there are a multitude of journals, many of which showing limited scientific rigour, it assigns equal weight to citations by articles published in low-impact journals compared to citations by articles published in higher-impact journals (with more rigorous scientific standards);
- articles with number of citations lower than the H-index do not contribute to it, as well as citations exceeding the H-index. As for the latter: one author having few publications (e.g. 10) with lots of citations each (e.g. 100) will have the same H-index of a colleague having the same number of publications with much less citations each (e.g. 10);
- self-citations will bias its computation, especially in multi-authored papers where "coordinated efforts" among authors to collectively self-cite may happen.

Several authors have highlighted these limits (Bornmann et al., 2008; Ioannidis et al., 2016; Waltman, 2016; Koltun et al., 2021). For example, Ioannidis et al (2016) showed that many of the top 1,000 authors on total citations have had no first/last-authored cited papers, while several Nobel laureates and other extremely influential scientists rank low in such a list. Hirsch himself recently highlighted that his index could have "severe unintended negative consequences" and "fail spectacularly" also for favouring quantity over quality of scientific publications (Hirsch, 2020). Alternative scoring systems have been proposed, for example using either arithmetic or geometric distribution of each author's contribution to published articles (Waltman, 2016). Nonetheless, up to now none of these approaches has been implemented on a large scale. It is possible that they may appear complex and lack appeal compared to the easily understandable H-index. It is even more likely that addressing the problem of developing a fairer measure of scientific achievement would require a wide consensus among publishers, researchers and funders, that may have never been really sought. In any case, none of the alternative proposed scores has simultaneously addressed the problems listed above.

Measuring scientific achievement: how, ideally?

While the latter consensus will be essential for going beyond the H-index, there is space to favour this consensus through better explaining, with clear reasoning and easy math, the rationale and practicability of new approaches even considering some further enhancement. All the aforementioned elements could be put together, as in the proposal below trying to explain them as clearly as possible. Each author could be eventually assigned a score for each article – therefore considering his/her overall scientific production - based on his/her contribution to the article (an issue which has already been extensively addressed in the scientific literature) (Waltman, 2016), but also based on a weighted article relevance. This score would depend on: 1) number of authors of the article; 2) position in the list of authors (for those disciplines where authors are not listed alphabetically); 3) number of citations that the article receives (excluding self-citations); 4) the weight of these citations (based on the impact factor – IF, at the time of the citation - of the journals where the articles citing it are published).

The overall score for that author would come from the sum of scores obtained for each authored article, so that all citations can contribute to it.

An articulation of this type would make the algorithm apparently complex, but still quite practicable through automated indexing systems, and would define a suitable system for overcoming some of the limits of the current evaluation system.

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)+…+[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 (Waltman, 2016).

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.

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 which can distort the scientific profile of researchers. Information available in bibliographic databases can certainly allow the implementation of alternative scoring systems, using algorithms such as the one hypothesized above. The latter, which may have the advantage of simultaneously addressing different and relevant limits of the H-index, provides an example showing that limits of the H-index can be at least partially overcome, bearing in mind that it would be naïve to hypothesize a "perfect" score without any downside. Trying to weight authors' contribution quantitatively can never provide a complete representation of the relevance of their scientific output. Their position in the authors' list (when they are not listed alphabetically, as it happens in some scientific fields), 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 (International Committee of Medical Journal Editors; Resnik et al., 2016). In any case, when researchers' scientific achievements have to be assessed for reasons such as grant assignments, qualitative and subjective judgements are also needed, provided that they are presented with the highest level of transparency.

A proxy measure beyond the H-index would necessarily require a consensus to be reached mainly between publishers, ideally with the participation of representatives of researchers and research funders. Researchers should advocate for fairer measures of their scientific achievements, even if some of them may see their profile losing some appeal. Actually, there is a concrete risk that the current system, in spite of all its flaws, has begun quite familiar as well as ways to get scientific "profit" from it. This may be one of the reasons why attempts to challenge it have been half-hearted at best. The proposal in this paper provides a workable example of the main elements that may be considered in an updated, possibly improved scoring system, easily implementable in the era of artificial intelligence (which is for much higher things).

Acknowledgements

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

Statements and Declarations

Conflict of interest statement

The author has no competing interests to declare that are relevant to the content of this article (except for being first author in most of the papers he has authored).

Funding statement

No funds, grants, or other support was received.

References

- Bornmann, L., Mutz, R., Daniel, H.D. (2008). 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*, 59(5), 830–837. https://doi.org/10.1002/asi.20806
- Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences,* 102(46), 16569-16572. https://doi.org/10.1073/pnas.0507655102
- Hirsch, J.E. (2020). Superconductivity, What the H? The Emperor Has No Clothes. *Physics and Society,* 49, 4-9
- 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
- Ioannidis, J.P.A., Klavans, R., Boyack, K.W. (2016). Multiple Citation Indicators and Their Composite across Scientific Disciplines. *PLoS Biology* 14(7): e1002501 https://doi.org/10.1371/journal.pbio.1002501
- 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
- Ministero della Salute (2021). Bando della Ricerca Finalizzata 2021. https://www.salute.gov.it/imgs/C_17_bandi_283_0_file.pdf
- Resnik, D.B., Tyler, A.M., Black, J.R., Kissling, G. (2016). Authorship policies of scientific journals. *Journal of Medical Ethics*, 42(3), 199-202 doi:10.1136/medethics-2015-103171 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4769679/pdf/nihms754790.pdf
- Waltman L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics,* 10(2), 365-391. https://doi.org/10.1016/j.joi.2016.02.007