Publishing Research Quarterly
https://doi.org/10.1007/s12109-024-09976-2

Detecting the Impact of Academics Self-Citations: Fi-Score

Luca Fiorillo[1,2,3]

Abstract

The academic landscape has witnessed significant transformations in recent years, primarily attributed to advancements in IT tools, which have advantages and drawbacks in the world of publications. The transition from traditional university library searches to the digital era, with access to various information sources such as Pubmed, Scopus, Web of Science and Google Scholar, has revolutionized research practices. Thanks to technology, researchers, academics and students now enjoy rapid and vast information access, facilitating quicker manuscript preparation and boosting bibliometric parameters. To identify authors "self-distorted" bibliometric parameters, different indices following the Hirsch index (h-index) (based on citations) have been proposed. The new "fi-score" evaluates the reliability of citation counts for individual authors and validates the accuracy of their h-index, comparing the number of citations to the h-index value to highlight value that is not within the norm and probably influenced or distorted by authors themselves. It examines how authors' citations impact their h-index, although they are not self-citing. The study calculated the fi-score on a sample of 194,983 researchers. It shows that the average value of the fi-score is 25.03 and that a maximum value admissible as good must not exceed 32. The fi-score complements existing indexes, shedding light on the actual scientific impact of researchers. In conclusion, bibliometric parameters have evolved significantly, offering valuable insights into researchers' contributions. The fi-score emerges as a promising new metric, providing a more comprehensive and unbiased evaluation of scholarly impact. By accounting for the influence of citations and self-citations, the fi-score addresses the limitations of traditional indices, empowering academic communities to recognize better and acknowledge individual contributions.

✉ Luca Fiorillo
luca.fiorillo2@studenti.unicampania.it; lfiorillo@unime.it

[1] Department of Biomedical and Dental Sciences, Morphological and Functional Images, University of Messina, Via Consolare Valeria, 1, 98100 Messina, Italy

[2] Multidisciplinary Department of Medical-Surgical and Odontostomatological Specialties, University of Campania "Luigi Vanvitelli ", 80121 Naples, Italy

[3] Department of Prosthodontics, Dr. DY Patil Dental College and Hospital, Dr. DY Patil Vidyapeeth, Pune 411018, Maharashtra, India

Published online: 08 April 2024

Keywords Bibliometrics · Fi-index · Publications · Citations · Score

Introduction

Bibliometric fields in academia primarily focus on the quantitative analysis of academic publications, authors and journals. They involve metrics like citation counts, h-indices and impact factors to assess research impact and productivity. The Hirsch index (h-index) gauges an author's impact based on their publications and how often those works are cited. The impact factor measures a journal's influence by averaging the citations its articles receive in a set time.

Non-bibliometric fields in academia encompass various academic disciplines where quantitative analysis of publications is not the primary focus. These fields include humanities, law, social sciences, natural sciences and engineering [16]. Instead of focusing on bibliometric parameters, they emphasize qualitative analysis, experimentation and theory development to advance knowledge [16], especially in areas where research's qualitative and theoretical aspects hold greater significance.

It has never been so simple to write a manuscript or produce a scientific paper, even with low-quality content, but to increase one's bibliographic history [18]. It would therefore be inherently wrong to assume that the publications of two people with comparable h-indexes had the same scientific impact in a given discipline. It is necessary to give importance to the position occupied in the list of authors. The h-index is a metric used to evaluate the impact and productivity of a researcher's or scientist's work. It is the highest number of "h" of a researcher's papers cited at least "h" times. In other words, an h-index of "h" means that a researcher has "h" papers, each of which has been cited at least "h" times. The h-index provides a quantitative measure of the quality and quantity of a researcher's contributions to their field, with higher values indicating more significant impact and productivity. In bibliometric fields of study, an author can have a high h-index by being part of a very productive research group that always gives a good but not particularly relevant or substantial contribution. In recent years, there has been an actual challenge towards achieving ever higher and more performing bibliometric parameters (citations numbers, h-index, etc.), so as to be suitable and strong in one's scientific disciplinary group: from the different university applications, to editorial roles in scientific journals, that require even more competitive parameters based on h-index [14]. Thanks to what has been said before, it is easy to understand how much it is possible to influence these parameters to give a boost to one's bibliometric parameters, either by cross-referencing with aggregated or even "unknown" authors (being careful not to create mutual publications) or again through the peer-review system where the request for references, by peer reviewers, has now become a must [7, 24, 25].

The academic world has undergone substantial changes in recent years. This is undoubtedly linked to the new IT tools that help researchers, and have both advantages and a disadvantage for the academic world. In the past, the research was carried out in university libraries. If necessary, contacts were made between universities or between different states, perhaps to have known that a particular university library had a valuable text for the research. This involved an incredible lengthening

of time, not to mention the drafting of the manuscript, the revision and re-reading, before being "analogically" (ordinary mail, floppy disks, CD-ROMs) presented to the scientific journal for submission [1]. The computerization of research, with the different information sources (among the major: Pubmed, Scopus, Web of Science and Google Scholar), has allowed incredible and rapid access to information for academics, researchers, students and non-students. This is with the ultimate aim of also being able to draft the manuscripts more quickly and spread knowledge [2, 5].

Numerous computer tools were subsequently added to this, which made it possible to correct, summarize and cite references in an automated and rapid way to make the job even more accessible. Today, all of this is using chatbots and artificial intelligence. It is understood that these tools can currently be helpful for the creation of only some types of manuscripts and cannot carry out clinical studies of any kind, which, over time, will become the most respected types of manuscripts in the academic world [6].

This manuscript is intended to illustrate a new method for assessing the reliability of the citation count of a particular author while still evaluating the integrity of his h-index: the fi-score.

Materials and Methods

First of all, in this section, the indices and parameters needed to calculate the fi-score will be evaluated. The h-index is one of these indices, probably the most significant because it measures the scientific impact of researchers in bibliometrics fields through the citations that their manuscripts receive from other researchers. This value is used to weight a researcher's career. The citations received by each publication are considered to calculate this index. For example, having an h-index equal to 50 means having 50 publications cited at least 50 times each. This calculation resulted in a simple way to measure the specific scientific contribution of the researcher [17]. The total citation number of a researcher, instead, represents the other parameter that must be considered for calculating the fi-score.

To evaluate the average value of the fi-score, a sample of 194,983 worldwide researchers has been examined up to 2021. Researchers have been selected by an existing database annually published by Stanford University regarding the World's top 2% scientists ranking of 22 scientific fields and 174 subfields according to the standard Science-Metrix classification [19].

Results

The fi-score corresponds to a simple calculation, after what has been said above:

$$\text{Fi-score} = \frac{h - \text{index}^2}{\text{total citations}} \times 100$$

The aim is to identify how much an author has influenced their h-index with citations. This purpose is shared with the previously published fi-index [12]; it gives us objective information on how much the h-index of a particular author has been influenced by their self-citations. But, unlike the latter, the fi-score analyzes the citations and considers the data independently of the self-citations since the author could have a network of collateral researchers who can increase their citations in a targeted way. This index could be used to evaluate researchers during application procedures. The calculation of the fi-score is straightforward and intuitive. Successfully normalizing the value and identifying the universally acceptable value were the key challenges. Moreover, it was crucial to determine the threshold values that must not be exceeded under any circumstances.

To do this, an existing database of global researchers was used, published and with raw data available in the form of spreadsheets [19]. At this point, the data of interest was collected—in this case, h-index—and total citations recognized with the names "h21" and "nc9621," respectively (including self-citations) (Fig. 1). This data was selected individually to create other spreadsheets, and then the fi-score was calculated for 194,983 different researchers on the list. Some of the statistical values of interest are shown below in Table 1.

1	nc9621	h21	fi-score
2	333.175	299	26,8330457
3	343.152	264	20,3105329
4	399.386	258	16,6665832
5	237.038	248	25,9468946
6	271.648	244	21,916598
7	209.040	243	28,2477038
8	267.236	241	21,733973
9	237.426	240	24,2601905
10	225.765	236	24,6699001
11	202.425	234	27,0500185
12	241.346	233	22,4942614
13	271.683	233	19,9824796
14	442.562	232	12,1619118
15	198.396	231	26,8962076

Fig. 1 Screenshot of Microsoft Excel® used spreadsheet. Calculation derived from the top-ranking researchers' published list. First column: author number (anonymized), Nc9621: total citations from 1996 to 2021; h21: h-index up to 2021; fi-score: fi-score calculation

Table 1 Main statistical data

Researchers' sample number	Average	Standard deviation
194,983	25,031416	6,84,622,183

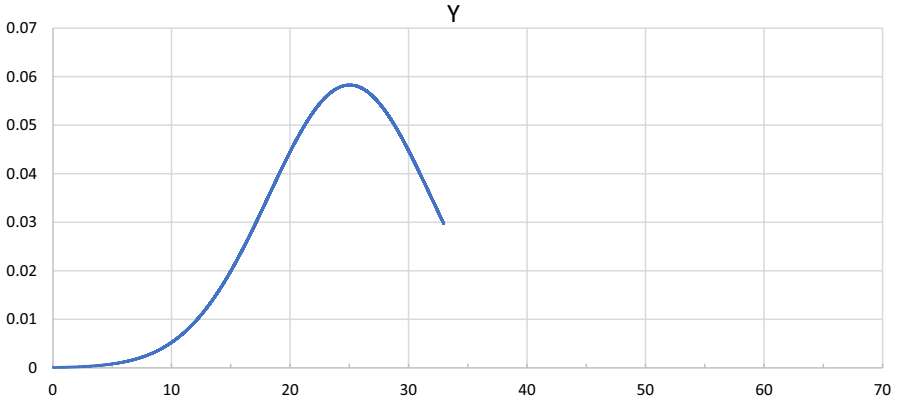

Fig. 2 Scatter plot with Gaussian curve about performed fi-score on the sample

Table 2 A particular case of nonsense fi-score

Total citation	h-index	Fi-score
1	1	100

Subsequently, using Microsoft Excel® software, the normal distribution of the values was calculated with a Gaussian curve, and a scatter plot of the values obtained was created (Fig. 2).

As can be seen from Fig. 2, it is intuitive that obtaining values that exceed the average fi-score value of 25.03 means that, in some way, the author was able to influence their number of citations. Considering the standard deviation obtained, a maximum admissible value is 31.87763783, rounded to 32. We do not worry about the lowest admissible value as this gives a favourable judgment to the researcher, always considering that having an h-index of at least 1, the fi-score cannot be 0. Below is a report of an absurd case justified by a lack of research history (Table 2). It reports an high fi-score despite lack of citations and manuscripts. This is the main limitation of this index, due to mathematical properties.

Discussion

Bibliometric indicators are used in the context of scientometric analysis to analyze the diffusion models of scientific publications and to evaluate their impact on scientific communities. They are indices that are developed by applying mathematical and statistical techniques. In addition to the analysis of scientific information distribution models, they are used in research evaluation processes. Bibliometric parameters,

including citation counts, h-index and impact factors, are essential tools for evaluating academic research. They help assess the impact of individual researchers, the quality of journals and the performance of academic institutions. Researchers use these metrics to gauge the significance of their work, while journal rankings influence manuscript submission choices. These parameters have become integral to the academic landscape, guiding decisions related to career progressions, publication selections and research fundings. Bibliometric parameters have been published to "normalize" the scientific production of a researcher as much as possible, and make it comparable; the most widely used being the h-index. To address this comparison, various indices have been proposed over the years:

- g-index by Egghe et al. [11]
- a-index by Jin et al. [20]
- h(2)-index by Kosmulski et al. [23]
- hg-index by Alonso et al. [4]
- q(2)-index by Cabrerizo et al. [10]
- r-index by Jin et al. [21]
- ar -index by Jin et al. [21]
- f-index by Franceschini et al. [15]
- f-index by Katsaros et al. [22]
- fi-index by Fiorillo [12, 13]
- ch-index by Ajiferuke et al. [3]
- m-index by Bornmann et al. [8]
- v-index by Riikonen et al. [26]
- b-index by Brown [9].

The fi-score in fact aims to develop a corrective factor as a new index. It should be assumed that the length of a researcher's career significantly influences the fi-score as the number of citations tends to increase over time. Moreover, the presence of self-citations can alter the h-index.

Let us briefly examine some of the alternative indices proposed. The g-index measures the quality of a researcher by considering the performance of their best articles. The a-index includes only manuscripts within a researcher's h-index, averaging their citations. The h(2)-index gives more weight to highly cited works. For example, an h(2)-index of 10 means the author has at least ten works with 100 citations each. The hg-index combines "h" and "g" indices to penalize authors with low h-indices, reducing the influence of a few highly cited works compared to the rest of their production. The q(2)-index is based on two indices, offering a comprehensive view of a researcher's production. The r-index is a modification of the a-index, proposing to take the square root of the sum of citations used for calculating the h-index, rather than dividing by the number of h-indices. The ar-index "completes the h-index" by considering citation intensity and the age of publications used to calculate the h-index [21]. Moreover, the "m" quotient proposes accounting for time by dividing the h-index by the number of years since the first publication. The first f-index relates the h-index to the publication age. Other indices include the m-index, calculating the average citations received from articles considered for the

h-index; the v-index, representing the percentage of articles useful for the h-index calculation; and the e-index, considering excess citations not accounted for in the h-index. Additionally, the f-index considers coterminal citations, extending co-citations to distinguish individuals in papers citing other papers. The ch-index evaluates the number of citers rather than citations. Lastly, the b-index assesses self-citations but not the actual number; instead, it considers the self-citation rate constant in an author's publications. Returning to the fi-score, it is obvious to think how to obtain a high h-index value with the minimum possible number of citations is a "good" and easy thing for a researcher. A researcher with a high h-index and a low number of citations should cast some doubts even if they are not self-citations.

As already mentioned, there are different ways to be able to have one's works cited, whether they are valid or not. If the citations number are noticeably low, it should raise an immediate concern, especially in case of a high h-index, and of short careers [3, 4, 9–11, 15, 17, 20–23, 26]. A researcher who is part of an important research team can make their indices increase quickly, mainly if they deal with a debated topic, which is why everything should be evaluated sectorally. However, the study considered many researchers worldwide who deal with different sectors, so the value should be mostly reliable. Having calculated the score for a sample of around 200,000 researchers, and showing an intuitive potential, it would be interesting to use this score in competition between peers in the academic field.

Limitations

While the fi-score presents a promising new approach to assess the reliability of citations and evaluate the influence on the h-index, it is essential to acknowledge certain limitations in this study:

- Sample size: The study might have been conducted on a sample of researchers, which may only partially represent part of the academic community. A more extensive and diverse dataset would be necessary to validate the findings and make broader conclusions.
- Field-specific variations: The fi-score might perform differently across various research fields. Different disciplines have unique citation patterns, and the influence of citations on the h-index could vary significantly. Additional studies should explore the applicability of the fi-score in other academic domains.
- Inclusion of self-citations: The study excluded self-citations from the fi-score calculation. While this addresses potential biases related to self-promotion, it may overlook the legitimate impact of self-cited work. Future studies could investigate the optimal approach to handle self-citations in bibliometric analyses.
- Temporal analysis: The current study did not consider the temporal evolution of citations and their impact on the h-index over time. Understanding how citations change over a researcher's career may provide deeper insights into the dynamics of academic impact.
- Potential citation manipulation: The fi-score may not account for cases where researchers artificially manipulate citations or engage in unethical practices to

increase their h-index. Robust measures to detect and prevent such manipulations should be developed.

- Correlation with research quality: The study focused on quantifying citations and their influence on the h-index, but it did not directly measure the quality or impact of the research itself. Future research should investigate the relationship between citation metrics and the actual scientific value of the publications.
- Comparison with existing indices: The fi-score was introduced as a new approach, but it has yet to be compared comprehensively with other established bibliometric indices. A comparative analysis with existing metrics would help assess the advantages and limitations of the fi-score.
- Subjectivity in author contributions: Bibliometric indices, including the h-index, often fail to capture the nuances of author contributions within a publication. These metrics may not accurately represent collaborative research and shared authorships.
- Causal inference: The study focused on the correlation between citations and the h-index but did not establish causal relationships. Determining whether a high h-index leads to more citations or vice versa remains challenging. The sample has been obtained by 2% of top-ranked researchers worldwide, so the average value of the fi-score could be influenced.

Conclusion

The proposed fi-score opens the door to a more comprehensive evaluation of researchers and their contributions. It provides a more equitable basis for academic evaluations, funding decisions and editorial roles in scientific journals. Moreover, future perspectives may involve further advancements in artificial intelligence and machine learning algorithms, allowing for a more sophisticated and nuanced analysis of research contributions. This could lead to the development of personalized metrics that consider the uniqueness of different research fields and the diversity of academic career trajectories. However, it is essential to be cautious in pursuing bibliometric parameters. The focus should remain on fostering genuine and impactful research rather than engaging in practices solely aimed at artificially boosting these metrics. Encouraging collaboration, cross-disciplinary research, and quality over quantity will be crucial in shaping the future of academic evaluation and recognition.

Funding Open access funding provided by Università degli Studi della Campania Luigi Vanvitelli within the CRUI-CARE Agreement.

References

1. Adam, David, and Jonathan Knight. 2002. "Publish, and be damned." *Nature* 419 (6909): 772–776. https://doi.org/10.1038/419772a.
2. Aga, S.S., and S. Nissar. 2022. "Essential guide to manuscript writing for academic dummies: An editor's perspective." *Biochemistry Research International* 2022: 1492058. https://doi.org/10.1155/2022/1492058.
3. Ajiferuke, Isola, and Dietmar Wolfram. 2010. "Citer analysis as a measure of research impact: Library and information science as a case study." *Scientometrics* 83 (3): 623–638. https://doi.org/10.1007/s11192-009-0127-6.
4. Alonso, S., F.J. Cabrerizo, E. Herrera-Viedma, and F. Herrera. 2010. "hg-index: A new index to characterize the scientific output of researchers based on the h- and g-indices." *Scientometrics* 82 (2): 391–400. https://doi.org/10.1007/s11192-009-0047-5.
5. Bartholomew, R.E. 2014. "Science for sale: The rise of predatory journals." *Journal of the Royal Society of Medicine* 107 (10): 384–385. https://doi.org/10.1177/0141076814548526.
6. Beall, Jeffrey. 2012. "Predatory publishers are corrupting open access." *Nature* 489 (7415): 179–179. https://doi.org/10.1038/489179a.
7. Bi, Henry H. 2023. "Four problems of the h-index for assessing the research productivity and impact of individual authors." *Scientometrics* 128 (5): 2677–2691. https://doi.org/10.1007/s11192-022-04323-8.
8. Bornmann, Lutz, Rüdiger Mutz, and Hans-Dieter Daniel. 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.
9. Brown, Richard J.C. 2009. "A simple method for excluding self-citation from the *h*-index: the *b*-index." *Online Information Review* 33 (6): 1129–1136. https://doi.org/10.1108/14684520911011043.
10. Cabrerizo, F.J., S. Alonso, E. Herrera-Viedma, and F. Herrera. 2010. "q2-index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core." *Journal of Informetrics* 4 (1): 23–28. https://doi.org/10.1016/j.joi.2009.06.005.
11. Egghe, Leo. 2006. "Theory and practise of the g-index." *Scientometrics* 69 (1): 131–152.
12. Fiorillo, Luca. 2022. "Fi-index: A new method to evaluate authors hirsch-index reliability." *Publishing Research Quarterly* 38 (3): 465–474. https://doi.org/10.1007/s12109-022-09892-3.
13. Fiorillo, Luca, and Marco Cicciù. 2022. "The use of Fi-index tool to assess per-manuscript self-citations." *Publishing Research Quarterly* 38 (4): 684–692. https://doi.org/10.1007/s12109-022-09920-2.
14. Fiorillo, Luca, and Vini Mehta. 2023. "Research, publishing or a challenge?" *International Journal of Surgery Open* 61: 100713. https://doi.org/10.1016/j.ijso.2023.100713.
15. Franceschini, Fiorenzo, and Domenico A. Maisano. 2010. "Analysis of the Hirsch index's operational properties." *European Journal of Operational Research* 203 (2): 494–504. https://doi.org/10.1016/j.ejor.2009.08.001.
16. Hammarfelt, Björn. 2016. "Beyond coverage: Toward a bibliometrics for the humanities." In *Research assessment in the humanities: Towards criteria and procedures*, edited by Michael Ochsner, Sven E. Hug, and Hans-Dieter Daniel, 115–131. Cham: Springer International Publishing.
17. Hirsch, J.E. 2005. "An index to quantify an individual's scientific research output." *Proceedings of the National Academy of Sciences of the United States of America* 102 (46): 16569–16572. https://doi.org/10.1073/pnas.0507655102.
18. Inouye, Kelsey, and David Mills. 2021. "Fear of the academic fake? Journal editorials and the amplification of the 'predatory publishing' discourse." *Learned Publishing* 34 (3): 396–406. https://doi.org/10.1002/leap.1377.

19. Ioannidis, John P.A., Kevin W. Boyack, and Jeroen Baas. 2020. "Updated science-wide author databases of standardized citation indicators." *PLoS Biology* 18 (10): e3000918. https://doi.org/10.1371/journal.pbio.3000918.
20. Jin, Bihui. 2007. "The AR-index: complementing the H-index." https://sci2s.ugr.es/sites/default/files/files/TematicWebSites/hindex/Jin2007.pdf.
21. Jin, BiHui, LiMing Liang, Ronald Rousseau, and Leo Egghe. 2007. "The R- and AR-indices: Complementing the h-index." *Chinese Science Bulletin* 52 (6): 855–863. https://doi.org/10.1007/s11434-007-0145-9.
22. Katsaros, Dimitrios, Leonidas Akritidis, and Panayiotis Bozanis. 2009. "The f index: Quantifying the impact of coterminal citations on scientists' ranking." *Journal of the American Society for Information Science and Technology* 60 (5): 1051–1056. https://doi.org/10.1002/asi.21040.
23. Kosmulski, Marek. 2006. "A new Hirsch-type index save time and works equally well as the original h-index." https://sci2s.ugr.es/sites/default/files/files/TematicWebSites/hindex/kosmulski2006.pdf.
24. Magadán-Díaz, Marta, and Jesús I. Rivas-García. 2022. "Publishing industry: A bibliometric analysis of the scientific production indexed in scopus." *Publishing Research Quarterly* 38 (4): 665–683. https://doi.org/10.1007/s12109-022-09911-3.
25. Negahdary, Masoud, Mahnaz Jafarzadeh, Ghasem Rahimi, Mahdia Naziri, and Aliasghar Negahdary. 2018. "The modified h-index of scopus: A New way in fair scientometrics." *Publishing Research Quarterly* 34 (3): 430–455. https://doi.org/10.1007/s12109-018-9587-y.
26. Riikonen, Pentti, and Mauno Vihinen. 2008. "National research contributions: A case study on Finnish biomedical research." *Scientometrics* 77 (2): 207. https://doi.org/10.1007/s11192-007-1962-y.