http://jhmi.sums.ac.ir
Health Man & Info Sci 2025, 12(1), 57–62

HMIS
Health Management and Information Science

Original Article

# Introduce a Novel Index for Scientific Output Ranking by Addressing Self-citation

**Ali Reza Soltanian[1*], Roya Ahmaddoost Razdari[2]**

[1]Department of Biostatistics, Modeling of Noncommunicable Diseases Research Center, Hamadan University of Medical Sciences, Hamadan, Iran
[2]Department of Biostatistics, Student Research Committee, Hamadan University of Medical Sciences, Hamadan, Iran

**Abstract**

**Introduction:** The Hirsch index is not comprehensive for scientific outputs ranking. Its structure does not include some issues, such as self-citation. Self-citation creates a scientometric complication. Using some scientometric indices, such as the h-index, is inadequate for ranking scientists alone, and they may turn to fake self-citations to increase their *h*-index over time. This study introduced a new index (*Ah*-index) based on self-citation to rank scientific outputs.

**Methods:** The information system integration division (ISID) of Hamadan University of Medical Sciences was used to calculate the Ah-index on August 5, 2023 (n=445). The bootstrapping technique was used to estimate uncertainty, standard error, and 95% confidence interval of Pearson's correlation coefficient between the new index and the h-index.

**Results:** This study shows that the correlation coefficient between h-index and Ah-index is 0.995 (C.I. 95%: 0.993-0.997). The authors' rank according to the h-index differed from the Ah-index, where scientists with high self-citations lost their position from the top of the list of authors according to our new indicator.

**Conclusion:** The findings of this study showed that by adjusting the self-citation of scientists, the author's position in the list of authors would be changed and become fairer.

**Keywords:** Bibliometrics, H-index, Information, Self-citation

**Article History:**
Received: October 9, 2024
Accepted: December 5 2024

**Please cite this paper as:**
Soltanian AR, Ahmaddoost Razdari R. Introduce a Novel Index for Scientific Output Ranking by Addressing Self-citation. Health Man & Info Sci. 2025; 12(1): 57-62. doi: 10.30476/jhmi.2025.102014.1212.

***Correspondence to:**
Ali Reza Soltanian,
Department of Biostatistics,
Hamadan University of Medical Sciences, Zip code 6517838687,
Hamadan, Iran
**Tel/Fax:** +98 81 38380026
**Email:** soltanian@umsha.ac.ir

## Introduction

Determining the quality of scientific research worldwide has always been an essential, fascinating, and controversial topic. Assessing research impact is critical for evaluating scientific contributions and allocating resources effectively. Various indicators have been developed to measure science and knowledge production (1). One of the most widely recognized metrics is the Hirsch index (h-index), introduced by physicist Jorge E. Hirsch in 2005 (2). The h-index assesses an individual or journal's scientific impact by quantifying the number of publications (quantity) and their citation count (quality). This dual-function approach differentiates prolific researchers and those who genuinely influence their fields. However, despite its widespread use, the h-index has notable limitations, leading to the development of alternative indices (3).

A key limitation of the h-index is its inability to account for a researcher's career length. The m-index was introduced to address this, which adjusts for a researcher's academic career, ensuring fairer comparisons between early-career and veteran researchers (4, 5). Furthermore, Waltman and Van Eck (2012) demonstrated inconsistencies in the h-index, arguing that it does not fully capture the overall scientific impact (2). Another shortcoming is that the h-index does not adequately consider highly cited articles, potentially underrepresenting groundbreaking research contributions. To resolve this issue, Egghe proposed the g-index in 2008, which gives greater weight to the most cited papers (4). However, like the h-index, the g-index does not address self-citation, a significant factor that can distort impact assessments (6, 7).

Numerous indices have been proposed to refine the evaluation of scientific impact, but a significant gap remains: the influence of self-

citation on researcher rankings. Self-citation can be a legitimate practice, linking current research to previous findings, but it can also be exploited to inflate impact metrics artificially. Some studies have shown excessive self-citation is often concentrated in lower-impact publications (8). Moreover, engineered self-citation strategies can undermine the reliability of scientometric evaluations, creating ethical concerns (9). Existing indices, such as the h-index, g-index, and m-index, do not explicitly address self-citation, leaving a loophole that may be exploited.

To fill this gap, we propose a novel scientometric index, the Ah-index, which integrates self-citation data into the traditional h-index framework. Unlike previous indices, the Ah-index accounts for self-citation frequency, offering a more balanced measure of scientific influence. This innovation ensures that researchers who rely excessively on self-referencing do not gain undue advantage in academic rankings. Additionally, by adjusting for self-citation, the Ah-index provides a more accurate reflection of an individual's impact within the scientific community.

Compared to existing variations of the h-index, such as the g-index and m-index, the Ah-index introduces a critical dimension by explicitly incorporating self-citation. The g-index corrects for highly cited papers but does not differentiate between genuine external citations and self-references. Similarly, the m-index adjusts for career length but overlooks the impact of self-citation. The Ah-index addresses both concerns by ensuring that citation counts are not artificially inflated through self-referencing, offering a more robust evaluation tool. Furthermore, while previous indices focus solely on citation count adjustments, the Ah-index provides insights into citation ethics and integrity within academic publishing.

Despite its advantages, the Ah-index is not without challenges. A key concern is that researchers might attempt to game the system by strategically adjusting self-citation patterns. Moreover, self-citation norms vary across disciplines; some fields naturally exhibit higher self-citation rates due to niche research topics and limited reference pools (10). Future research should explore field-specific thresholds for self-citation within the Ah-index framework to ensure fair evaluations. Additionally, comprehensive and reliable citation databases are necessary for accurate calculations, and discrepancies in data availability across institutions could pose implementation challenges.

Another important consideration is the potential unintended consequences of discouraging self-citation. While excessive self-citation can be problematic, legitimate self-referencing links past and current research. Penalizing self-citation too harshly could discourage authors from citing foundational work, leading to gaps in research continuity. Future studies should investigate strategies for distinguishing between necessary and excessive self-citation, refining the Ah-index accordingly.

Finally, adopting new scientometric indices often faces resistance from the academic community. The h-index is deeply entrenched in research evaluation systems, and shifting to a new metric may require widespread acceptance from funding agencies, universities, and researchers. Future research should focus on the practical implementation of the Ah-index across diverse disciplinary contexts to assess its effectiveness in real-world applications.

In conclusion, while existing scientometric indices provide valuable insights into research impact, they fail to account for the influence of self-citation. The Ah-index bridges this gap by integrating self-citation data into impact assessments, offering a more comprehensive and ethical ranking system. By addressing self-citation distortions, the Ah-index ensures a more accurate evaluation of scientific contributions, fostering greater integrity in academic research. Future research should further refine the index, addressing disciplinary differences and potential gaming strategies to enhance its robustness and applicability.

## Methods

As we know, "a scientist has index $h$ if $h$ of his or her $N_p$ papers have at least h citation each and the other ($N_p$-$h$) papers have $\leq h$ citations each" (1).

Let, in equation 1, "SC"," h", and "Ah" denote self-citation proportion, Hirsch index, and adjusted Hirsch index, respectively. Hence, the Hirsh index can be rewritten as follows for each scientist, journal, etc.:

$$Ah_{index} = h_{index} - (h_{index} \times SC) \quad \text{Eq. 1}$$

Consider a researcher with an $h$-index of 12 and a self-citation of 7%. In this case, the $Ah$-index will be equal to:

$$Ah_{index} = 12 - (12 \times 0.07) = 11.16$$

To illustrate and validate this proposal, the scientific data from the information system integration division (ISID) of Hamadan University of Medical Sciences was used on August 5, 2023 (https://isid.research.ac.ir/). Among the 488 scientists, those who had an h-index of at least one was considered at this time (n=445). Pearson's correlation coefficient was used to show a correlation between *the h*-index- and *the Ah*-index concerning equation 2.

$$\hat{\rho} = r = \frac{[\sum_{i=1}^{n} x_i y_i] - n\underline{x}\underline{y}}{\left[\left(\sum_{i=1}^{n} x_i - n\underline{x}^2\right)\left(\sum_{i=1}^{n} y_i - n\underline{y}^2\right)\right]^{\frac{1}{2}}} \quad \text{Eq. 2}$$

In addition, the bootstrap resampling method was used to infer the correlation coefficients. To estimate the standard error and 95% confidence interval by the bootstrap method, n=400 replacement samples were performed for B=1000 times from the ISID data set (Figure 1). The bootstrap package in R software 3.6.1 was used for these calculations.

## Results

In the study, the *h*-index and self-citation of 445 scientists at Hamadan University of Medical Sciences were considered, and Table 1 is based on the first 15 scientists of the university. The h-index and the *Ah*-index ranked the scientists. As shown in Table 1, the ranking system of the scientists is different according to *the h*- and *Ah*- index. For example, consider the third scientist with an *h*-index of 31 and a self-citation of 27% (Table 1). He/she receives the eighth rank by the *Ah*-index, while he/she receives the second rank concerning the *h*-indexing system. Therefore, the results show that after adjusting self-citation, the order of the scientists changed. In other words, scientists 1 and 2 in Table 1 have the same *h*-index, but the self-citation of the first and second scientists is 5% and 2%, respectively. A comparison of their *h*-index shows that both are in the first place, while the *Ah*-index, with adjusting self-citation, declares a more accurate ranking; i.e., the second scientist gets the first rank, and the first scientist gets the second rank, which is a fairer ranking.

This study shows that the Pearson's correlation coefficient between *h*-index and *Ah*-index (r=0.995, SE=0.0010; C.I. 95%: 0.993-0.997) is severe. In addition, the graphical representation between *the h*- and *Ah*- indices point to a strong correlation between them (Figure 2).

## Discussion

For the first time, this study intends to introduce another index (i.e., *the Ah*-index) similar to the

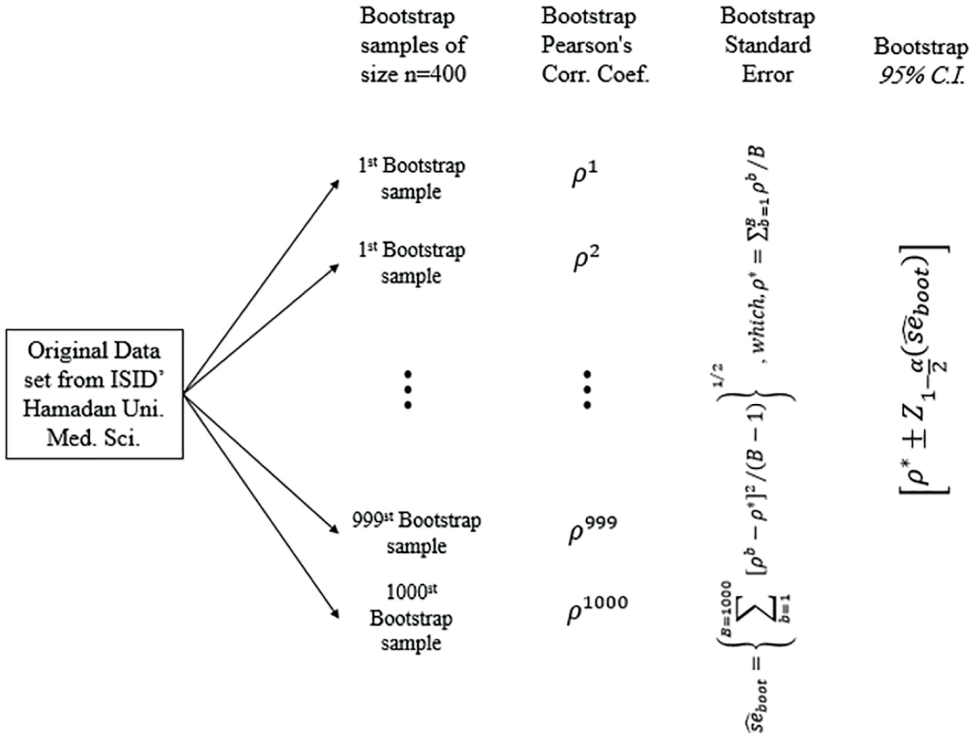


**Figure 1:** The bootstrap algorithm for estimating Pearson's correlation coefficient, the standard error, and the 95% confidence interval of Pearson's correlation coefficient.

**Table 1:** The *h*- and *Ah*-index of 15 academic scientists from different disciplines at Hamadan University of Medical Sciences.

| Author | Self-citation (%) | *h*-index | *Ah*-index | Author's rank by: | |
|---|---|---|---|---|---|
| | | | | *h*-index | *Ah*-index |
| 1 | 5 | 32 | 30.40 | 1 | 2 |
| 2 | 2 | 32 | 31.36 | 1 | 1 |
| 3 | 27 | 31 | 22.63 | 2 | 8 |
| 4 | 8 | 30 | 27.6 | 3 | 3 |
| 5 | 7 | 29 | 26.97 | 4 | 4 |
| 6 | 7 | 28 | 26.04 | 5 | 5 |
| 7 | 23 | 28 | 21.56 | 5 | 13 |
| 8 | 8 | 27 | 24.84 | 6 | 6 |
| 9 | 10 | 26 | 23.4 | 7 | 7 |
| 10 | 15 | 25 | 21.25 | 8 | 14 |
| 11 | 6 | 24 | 22.56 | 9 | 9 |
| 12 | 10 | 24 | 21.6 | 9 | 12 |
| 13 | 7 | 24 | 22.32 | 9 | 10 |
| 14 | 6 | 24 | 22.56 | 9 | 9 |
| 15 | 5 | 23 | 21.85 | 10 | 11 |

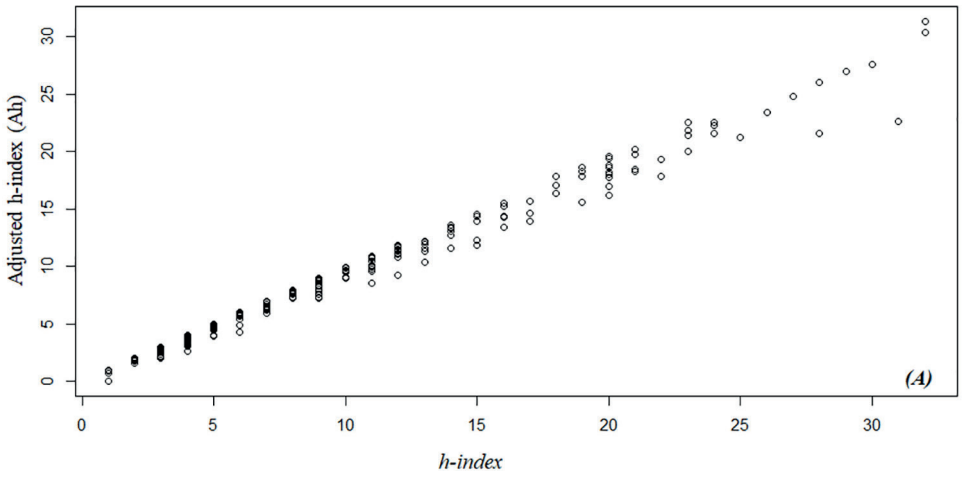


**Figure 2:** A scatter plot for pairwise correlation between the *h*- and *Ah*- index of 446 scientists from various disciplines under ISID of Hamadan University of Medical Sciences was used on August 5, 2023, with *h*-index≥1 (https://isid.research.ac.ir/).

*h*- h-index for ranking scientists, journals, research institutions, etc. This study used Pearson's correlation coefficient and bootstrap resampling technique to prove the correlation between the h- and Ah- indices.

The findings of this study showed that the correlation between the two indices of *the h*- and *Ah*- index is positive and very strong. This shows that *the Ah*-index can be used the same as *the h*-index for ranking and sorting scientists, journals, etc., in Scopus, Web of Science, PubMed, and other databases. Self-citation is inherently appropriate(10), but its involvement in ranking systems prevents doubt in the number of citations (11). In the ranking system, self-citation cannot be ignored, nor can it be eliminated. Therefore, it is better to introduce an intermediate index.

The increase in self-citation may be one of the reasons for the salience of articles (11). Therefore, it is necessary to introduce a scientific indicator that can partially prevent such a problem. Our new indicator, the *Ah*-index, can reduce the effect of self-citation in research ranking.

Corresponding to the bootstrap findings, the *h*- and *Ah*- index behave similarly. Notice that the *Ah*-index, in addition to having the characteristics of the h-index, is also adjusted according to self-citation. In addition, our findings showed that adjusting self-citation changes the author's rank and becomes more real. After adjusting it, authors with many self-citations will lose their rank from the top of the author list. One of the problems that the *Ah*- index can cause is the opposition of some scientists, who believe (11)

that self-citation is unimpeded. In response to such ambiguities, the *Ah*- index does not intend to eliminate self-reference; we want to make the index more rational by controlling and adjusting its value to rank.

In a simulation study, Bartneck C and colleagues showed that authors can significantly increase the h index by citing their works. Finally, they provided an index called q to identify this behavior. Researchers believe the strategy of unfair self-citation is helpful for authors cited less by others and less productive (12).

This novel indicator may not be a comprehensive index for scientometric ranking. However, it does not have the shortcomings of the *h*- index in a dynamic ranking system, e.g., Scopus and Web of Science databases. On the other hand, scientists or journals are confident that self-citation contributes to their ranking and will be cautious.

Note that the suggested index (*Ah*-index) is a float, not an integer. This feature is both a strength and a weakness. The float characteristics of the *Ah*-index allow scientific institutions to implement the ranking system by two decimal places. In addition, this feature makes the ranking fairer. The *Ah*-index cannot be interpreted simply as *h*-index because it is a combination of two components (equation 1), and it is a decimal value and not an integer. These are the weakness of this index. Therefore, we recommend using this index along with Hirsch's index in the ranking system of scientists and other scientific outputs. However, if researchers can develop an index in the future that can be interpreted like the Hirsch index, it will be a great success in scientometrics.

As mentioned earlier, our index is difficult to interpret, but it efficiently ranks scientists and prevents undue self-citation. For this reason, more work should be done on the Ah index to find a simple and understandable interpretation for people.

Although our index is imperfect, it is a step in developing Hirsch's index and can prevent self-citation without reason. This index does not allow authors to unnecessarily self-cite to increase the Hirsch's index. Our index can be the primary basis for the development of scientometric indices in such a way as to eliminate arbitrary and unreasonable self-citation. It is suggested that this index be worked on and a complete index with an integer value be introduced. Its interpretation is simple, like Hirsch's index.

As a limitation, *Ah*-index is challenging to interpret, but it is very efficient in ranking scientists and prevents undue self-citation. For this reason, more work should be done on the Ah index to find a simple and understandable interpretation for people.

## Conclusion

We introduced a new index to adjust the effect of self-citation in the ranking of scientometric indices such as *the h*-index. The simulation by the bootstrap method showed that the new *Ah*- index is as valid as the *h*- index. Therefore, the *Ah*-index, a combination of *h*-index and self-citation, can be a relatively good index for ranking scientific outputs.

## Acknowledgments

The Vice-Chancellor for Research and Technology of Hamadan University of Medical Sciences, Hamadan, Iran (IR.UMSHA.REC.1402.188) supported this study.

## Authors' Contribution

Soltanian AR designed this work. Soltanian AR and Ahmaddoost-radar R wrote this manuscript. All authors contributed to the article and approved the submitted version.

## Conflict of Interest

There are no conflicts of interest.

## References

1. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A*. 2005;102(46):16569-72. doi: 10.1073/pnas.0507655102.
2. Waltman L, Van Eck NJ. The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*. 2012;63(2):406-15. doi: 10.1002/asi.21678.
3. Egghe L. The Hirsch index and related impact measures. *Annual review of information science and technology*. 2010;44(1):65-114. doi: 10.1002/aris.2010.1440440109.
4. Bornmann L, Daniel HD. What do we know about the h index? *Journal of the American Society for Information Science and technology*. 2007;58(9):1381-5. doi: 10.1002/asi.20609.
5. Egghe L. How to improve the h-index. *The scientist*. 2006;20(3):15-6.

6. Schreiber M. A case study of the Hirsch index for 26 non-prominent physicists. *Annalen der Physik*. 2007;519(9):640-52. doi: 10.1002/andp.20075190904.
7. Egghe L, Ravichandra Rao I. The influence of the broadness of a query of a topic on its h-index: Models and examples of the h-index of n-grams. *Journal of the American Society for Information Science and Technology*. 2008;59(10):1688-93. doi: 10.1002/asi.20843.
8. Hyland K. Self-citation and self-reference: Credibility and promotion in academic publication. *Journal of the American Society for Information Science and technology*. 2003;54(3):251-9. doi: 10.1002/asi.10204.
9. Moed HF. Citation analysis in research evaluation. Berlin: Springer Science & Business Media; 2006. doi: 10.3233/ISU-2006-26217.
10. Fowler J, Aksnes D. Does self-citation pay? *Scientometrics*. 2007;72(3):427-37. doi 10.1007/s11192-007-1777-2.
11. Ioannidis JP. A generalized view of self-citation: direct, co-author, collaborative, and coercive induced self-citation. *J Psychosom Res*. 2015;78(1):7-11. doi: 10.1016/j.jpsychores.2014.11.008.
12. Khaji A, Saadat S. Is self-citation ethical? *Iranian Journal of Microbiology*. 2014;6(1).:35-44.