Evaluation of the productivity of hospital-based researchers: comparative study between the h-index and the h(fa)-index

Paul Sebo¹ · Sylvain de Lucia²

Received: 15 December 2020 / Accepted: 8 May 2021 / Published online: 23 June 2021
© The Author(s) 2021

Abstract

The h-index is often used to evaluate researchers and to measure their respective contributions to the scientific community. However, it does not take into account the relative contribution of each author to the articles. We aimed to assess the added value of the h(fa)-index, an index introducing a weighting factor for the first author. This cross-sectional study was conducted in March 2020 in Switzerland. We selected all hospital-based senior physicians practicing internal medicine or family medicine in the six university hospitals. Using *Web of Science*, we retrieved the h-index and computed the h(fa)-index of all physicians with a non-zero h-index. We used Spearman coefficients to examine the correlation between the two indices and Cohen's kappa to determine the inter-rater agreement. 244 physicians were included in the study (men: 63%). They had a median h-index of 3 (IQR 11) and a median h(fa) index of 4.5 (IQR 14). The correlation between the two indices was very high (ρ 0.98, p -value < 0.001). However, the inter-rater agreement was only moderate (Kappa 0.59 [95% CI 0.56–0.64]). The h(fa)-index was able to correct the categorization of the researchers in about 1/3 of the cases (agreement 63.4%). Although it is very challenging to evaluate researchers with a unique indicator, these data suggest that the h(fa)-index could be a better evaluation tool of researchers' productivity.

Keywords Academic research · Bibliometric study · H-index · H(fa)-index · Productivity · Switzerland

Introduction

One of the key aspects of scientific research is the publication of results, which not only allows new scientific knowledge to be disseminated, (Lafrenière et al., 2013; Mabe, 2010) but also increases the recognition of researchers (Bavdekar & Tullu, 2016; Post et al., 2012; Vale, 2015). Quantitative analysis of researchers' productivity (by documenting the number of articles published) and qualitative analysis of their publications (usually by looking

✉ Paul Sebo
paulsebo@hotmail.com

¹ Primary Care Unit, University of Geneva, Geneva, Switzerland

² Department of Community Health and Medicine, Geneva University Hospital, Geneva, Switzerland

at the number of citations and/or the impact factor of journals) has become essential for comparing researchers (Pfeiffer et al., 2016; Prpić, 1996). It also allows to dig some important topics related to research policy, such as the influence of gender and collaboration between institutions (Akers et al., 2016; Gasparyan et al., 2014, 2016; Lerchenmueller & Sorenson, 2016).

The h-index is a bibliometric variable that is increasingly used to evaluate researchers and to quantitatively measure their contributions to the scientific production. It takes into account both the productivity of researchers and the quality of their articles, since it combines the number of publications and citations (Hirsch, 2005). Indeed, the number of citations alone does not allow a reliable assessment of the overall quality of researchers' publications, as it can be affected by the publication of a few articles with many citations. The h-index has been launched by Jorge Hirsch in 2005 (Hirsch, 2005). First, it was used to determine the scientific quality of articles published by theoretical physicists. It is now used in many scientific fields, including medicine (Baldock, 2008). This index is defined as the number of articles h that have each been cited at least h times. It is easily determined using databases such as *Web of Science*, *Scopus* and *Google Scholar*.

Despite its advantages, the h-index is often criticized because it does not take into account the relative contribution of each author to the articles. To overcome this limitation, Butson and Yu proposed to introduce a weighting factor for the first author in the h-index (Butson & Yu, 2010). This modified h-index, called the $h(fa)$ -index, could provide a more appropriate impact index. It is calculated with the following formula: $h\text{-index} * (N_p + F_p) / N_p$, where N_p and F_p are respectively the total number of articles and the number of first author articles entering into the calculation of the h-index. This new indicator is relatively easy to calculate, as the h-index is immediately available through various databases and F_p is easy to obtain by manual verification.

The main objective of this study was to assess whether there were differences in the classification of researchers according to whether the h-index or the $h(fa)$ -index was used (inter-individual evaluation). For reaching this objective, we collected data allowing us to calculate these indices for senior hospital physicians practicing internal medicine or family medicine in Swiss university hospitals. We hypothesized significant differences in the classification of researchers between the two indicators. The use of the $h(fa)$ -index would then bring an added value to the h-index, allowing to correct the categorization of researchers and to characterize more precisely the contribution of each researcher to the scientific production.

Materials and methods

Study site and study population

This was a secondary study based on data collected between 1 and 14 March 2020 for a bibliometric study aimed at assessing the productivity of senior hospital physicians in Switzerland (Sebo et al., 2020, 2021). We included all senior physicians (heads of division, staff physicians and senior registrars) practicing internal medicine or family medicine in one of the six Swiss university hospitals: three hospitals in predominantly French-speaking cantons (Western Switzerland: Geneva, Lausanne and Fribourg) and three hospitals in predominantly German-speaking cantons (Eastern Switzerland: Bern, Basel, Zurich).

Data collection

The list of physicians selected for the study was established by visiting hospital websites. Their main socio-demographic characteristics (gender, medical department [internal medicine, family medicine], title of professor [Y/N], hierarchical position [head of division, staff physician, senior resident] were then recorded. Finally, using *Web of Science*, their list of publications, the year of publication of their first article and their h-index were extracted, and their h(fa)-index was calculated. The procedure used to retrieve these bibliometric variables was described in detail in the main article (Sebo et al., 2020). Analyses were only conducted on data for physicians with a non-zero h-index.

Statistical analyses and sample size

Frequency tables were used to summarize the socio-demographic characteristics of physicians, and medians and interquartile ranges (IQR) were used to summarize their h-index and h(fa)-index. Spearman correlation coefficients were calculated to determine the correlation between the h-index and the h(fa) index. We considered correlations of 0.10 as "small", 0.30 as "medium" and 0.50 as "large". (Cohen, 1988).

We then sorted the physicians by their h-index and h(fa)-index, respectively, divided the two samples into ten quantiles and calculated the inter-rater agreement between the h-index and the h(fa)-index using the Cohen's kappa statistic. Cohen's kappa statistic is a measure originally designed to measure agreement between two raters. However, it was also used as a performance indicator in classification when comparing measures (Tian et al. 2018; Duro et al., 2012). We considered that an inter-rater agreement between 0.2 and 0.4 was "fair", between 0.4 and 0.6 was "moderate", between 0.6 and 0.8 was "substantial" and between 0.8 and 1.0 was "excellent". (Landis & Koch, 1977) To avoid including overly large subgroups in this analysis, we removed physician subgroups when they contained more than 10% of the total study sample (physician subgroups with an h-index of 1 and 2).

Finally, negative binomial regressions were used to assess the associations between physicians' socio-demographic characteristics and their h-index and h(fa)-index (=over-dispersion count data) (Negative Binomial Regression | Stata Annotated Output, 2019; Negative Binomial Regression | Stata Data Analysis Examples, 2019).

We calculated the required sample size for the primary study using Zhu and Lakkis' equations developed to compare two negative binomial rates (Zhu & Lakkis, 2014). We estimated that by comparing two separate groups, a sample size of 336 (168 in each group) would be sufficient to detect a difference of five publications with a type I and II error both set at 5%. We expected that the mean number of publications would be ten in one group and five in the other, the ratio of the number of participants would be one, and the dispersion parameter would be five. We performed all statistical analyses with STATA version 15.1 (College Station, USA).

Results

Of the 349 senior physicians working in one of Switzerland's six university hospitals, 244 had a non-zero h-index and were included in the analyses (men: 63%). 147 of these physicians (60%) were located in French-speaking Switzerland and 126 (52%) were affiliated to

a department of family medicine. In addition, the proportion of experienced physicians was relatively high: 102 (42%) were staff physicians or heads of department, 48 (20%) had the title of professor, and the mean number of years since the publication of the first article was 12 (SD 7).

The 244 physicians in the study had a median h-index of 3 (IQR 11, min–max 1–94) and a median h(fa)-index of 4.5 (IQR 14, min–max 1–130). The correlation between the h-index and the h(fa)-index was very high (Table 1). If we restricted the analysis to physicians with the highest h-index, the correlation decreased slightly (rho 0.98 for the total sample, vs. 0.89 for the group consisting of the 25% of physicians with the highest h-index). The decrease in correlation for high h-indices is clearly illustrated in Fig. 1 (scatterplot).

Although the correlation was very high, Fig. 2 shows that the inter-rater agreement between the h-index and the h(fa)-index was only moderate (agreement 63.4%, kappa 0.59 [95% CI 0.56–0.64], p -value < 0.001).

Finally, Table 2 shows the association between the two indices and physicians' socio-demographic characteristics. With the exception of hospital location, all physician characteristics were significantly associated with the h-index and the h(fa)-index, and the levels of association found were the same or nearly the same for the h-index and the h(fa)-index. The differences in incident rate ratios (IRRs) were at most 0.2 in absolute value. For example, in univariate analysis, these indices were identical and on average 30% higher for physicians in internal medicine than for those in family medicine (h-index: IRR 1.3 [95% CI 1.0–1.8], p -value 0.04; h(fa)-index: IRR 1.3 [95% CI 1.0–1.8], p -value 0.05). In multivariate analysis, the IRRs were identical as well (h-index: IRR 1.3 [95% CI 1.0–1.5], p -value 0.03; h(fa)-index: IRR 1.3 [95% CI 1.0–1.6], p -value 0.03).

Discussion

Main findings

In summary, we found that the senior physicians included in our study had a median h-index of 3 and a median h(fa)-index of 4.5. The correlation between the two indices was very high, although slightly lower for physicians with a higher h-index. Despite the high correlation, we found that the inter-rater agreement was only moderate. Finally, the levels of association with physicians' socio-demographic characteristics were almost the same for the h-index and the h(fa)-index.

Table 1 Correlation between the h-index and the h(fa)-index

Study sample	Number of physicians	Spearman, rho	p -value
Overall	244	0.98	< 0.001
50% of physicians with the highest h-index	118	0.97	< 0.001
25% of physicians with the highest h-index	56	0.89	< 0.001

Fig. 1 Scatterplot for the association between the h-index and the h(fa)-index ($N=244$)

10 quantiles

of h-index

10 quantiles of h(fa)-index

	1	2	3	4	5	6	7	8	9	10	Total
1	19	5	0	0	0	0	0	0	0	0	24
2	1	4	0	0	0	0	0	0	0	0	5
3	0	8	6	2	0	0	0	0	0	0	16
4	0	0	3	11	3	1	0	0	0	0	18
5	0	0	0	5	1	2	0	0	0	0	8
6	0	0	0	1	2	10	1	1	0	0	15
7	0	0	0	0	0	3	9	3	0	0	15
8	0	0	0	0	0	0	5	6	2	0	13
9	0	0	0	0	0	0	0	4	10	0	14
10	0	0	0	0	0	0	0	0	0	14	14
Total	20	17	9	19	6	16	15	14	12	14	142

Fig. 2 Interrater agreement between the h-index and the h(fa)-index (the two indices were categorized into ten quantiles after removing physicians with an h-index of 1 or 2, $N=142$)

Table 2 Unadjusted and adjusted associations between physicians' socio-demographic characteristics, and the h-index and the h(fa)-index (N = 244)

Characteristic	h-index, IRR ^a (95% CI)	<i>p</i> -value ^b	h-index, IRR ^a (95% CI)	Adj. <i>p</i> -value ^c	h(fa)-index, IRR ^a (95% CI)	h(fa)-index, IRR ^a (95% CI)	Adj. <i>p</i> -value ^c
German- vs. French-speaking Switzerland	1.2 (0.9–1.7)	0.14	1.3 (1.1–1.6)	0.02	1.2 (0.9–1.6)	1.3 (1.0–1.6)	0.04
Internal vs. Family medicine	1.3 (1.0–1.8)	0.04	1.3 (1.0–1.5)	0.03	1.3 (1.0–1.8)	1.3 (1.0–1.6)	0.03
Staff physicians and heads of division vs. Senior registrars	4.3 (3.4–5.5)	<0.001	1.9 (1.5–2.5)	<0.001	4.3 (3.3–5.4)	2.0 (1.5–2.5)	<0.001
Title of professor	3.8 (2.8–5.1)	<0.001	1.6 (1.2–2.1)	<0.001	3.9 (2.8–5.3)	1.7 (1.3–2.2)	0.001
Male vs. female physicians	2.2 (1.6–2.9)	<0.001	1.3 (1.0–1.6)	0.03	2.1 (1.6–2.8)	1.3 (1.0–1.6)	0.05
10–19 years vs. < 10 years since first publication	4.1 (3.1–5.2)	<0.001	2.6 (2.0–3.3)	<0.001	4.2 (3.2–5.4)	2.7 (2.1–3.5)	<0.001
≥ 20 years vs. < 10 years since first publication	9.1 (6.5–12.8)	<0.001	4.0 (2.8–5.8)	<0.001	8.9 (6.2–12.6)	3.8 (2.6–5.5)	<0.001

^aIRR = Incident rate ratio^bUnivariate negative binomial regression^cMultivariate negative binomial regression; model adjusted for all variables listed in the table

Comparison with existing literature

We showed that the $h(fa)$ -index was very useful in quantifying the productivity and impact of individual researchers. The correlation between the two indices was very high. However, in our sample of hospital-based physicians previously sorted into 10 categories according to the h -index, we found that the $h(fa)$ -index was able to correct the categorization of the researchers in about 1/3 of the cases (level of agreement: 63%). Of course, the results probably depend on the sample examined and the number of categories used, but they do suggest that the $h(fa)$ -index could improve the estimation made with the h -index. These data are consistent with those presented by Butson and Yu for a small sample of Australia's leading medical physics researchers (Butson & Yu, 2010).

This index does not seem to add value when comparing groups of physicians rather than individual physicians. We found no difference in analyzing the association with various socio-demographic variables. The levels of association and p -values were very similar. For this type of subgroup analysis, the h -index, which is simpler and therefore quicker to obtain, seems to be sufficient.

Surprisingly, despite the relative simplicity of the $h(fa)$ -index, there are to our knowledge no published studies that have used this index in bibliometric studies or that have compared it to the h -index. Some authors have used an index similar to the $h(fa)$ -index, but even simpler to calculate, the h -index for articles limited to the first author only (McClelland et al., 2018; Vavryčuk, 2018). This index is probably too simplistic because it does not take into account the articles published in co-authorship, unlike the $h(fa)$ -index.

Perspectives

Although the h -index is increasingly used to evaluate researchers and to quantitatively measure their respective contributions to the scientific community, it has its critics (Bornmann & Marx, 2011; Schubert & Schubert, 2019; Waltman & Eck, 2012). The limitations that have been documented and discussed extensively in the literature have been summarized by Schubert & Schubert (Schubert & Schubert, 2019). The h -index depends on the duration of a researcher's career. There are differences in citation practices between fields, so the h -index should not be used to compare researchers from different disciplines. Since the h -index is easy to obtain, there is a risk of indiscriminate use (research performance is a complex, multifaceted endeavor that cannot be captured in a single indicator). It does not take into account the publication type. Finally, it does not take into account multiple co-authorship or self-citations. Baldock has shown, for example, that self-citations artificially improve the h -index (Baldock, 2008).

Whether we like it or not, it is likely that the h -index will not disappear soon and we will have to make do with it. There are however beyond any doubt ways to improve the evaluation of researchers. One solution would be to weight the h -index according to the position of the authors in the list of authors. The contribution of each author would unfortunately be difficult to estimate, as it would have to be done separately for each article. Another solution, much easier to implement, would be to introduce a weighting factor for the first author. This is the $h(fa)$ -index (Butson & Yu, 2010). This index could supplant the h -index in the future because it is fairly easy to obtain and probably evaluates the impact of researchers more reliably.

Thus, although it is certainly challenging to use a single quantitative indicator to evaluate researchers' publications, the $h(fa)$ -index would at least have the advantage of weighting

the score according to the number of first author articles entering in the calculation of the h-index. This could allow a more equitable evaluation of some researchers, as those working in small centers who are currently disadvantaged by the use of the h-index as an indicator of the impact of their research (Butson & Yu, 2010). However, although simple to calculate, the h(fa)-index is unfortunately not directly available using databases such as *Web of Science*, *Scopus* and *Google Scholar*. Therefore, in situations where promotion from junior to senior faculty or from senior faculty to department chair is not at the forefront, the h(fa)-index probably does not always add enough value to replace the h-index.

Limitations

First, we only included in the study physicians practicing general internal medicine and family medicine in Swiss university hospitals. The results can possibly not be generalized to other physicians or to other countries. Second, we cannot exclude certain errors in the lists of publications. Yet, even if this risk exists, it should be low. Indeed, we standardized search procedures, duplicated all searches and reviewed all publications to remove articles published by homonyms. In addition, we extracted the lists using the full first name of the physicians, not just the first letter of their first name.

Conclusion

In conclusion, in this sample of Swiss hospital-based senior physicians, we found that despite a high correlation between the h-index and the h(fa)-index, the inter-rater agreement was only moderate. The h(fa)-index was able to correct the categorization of the researchers in about 1/3 of the cases. Although it is certainly challenging to use a single quantitative indicator, the h(fa)-index could allow a more equitable evaluation of researchers.

Funding Open Access funding provided by Université de Genève.

Data availability The data underlying this article will be shared on reasonable request to the corresponding author.

Declarations

Conflict of interest The author declare that they don't have any conflict of interest.

Ethical approval Not required (under Swiss law, ethical approval is not required when collecting non personal health data).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akers, K. G., Sarkozy, A., Wu, W., & Slyman, A. (2016). ORCID Author identifiers: A primer for librarians. *Medical Reference Services Quarterly*, 35, 135–144.
- Baldock, C. (2008). The h-index and medical physics. *Australasian Physical and Engineering Sciences in Medicine*, 31, xi–xii.
- Bavdekar, S. B., & Tullu, M. S. (2016). Research publications for academic career advancement: An idea whose time has come. But is this the right way? *Journal of Postgraduate Medicine*, 62, 1–3.
- Bornmann, L., & Marx, W. (2011). The h index as a research performance indicator. *European Science Editing*, 37(3), 77–80.
- Butson, M. J., & Yu, P. K. N. (2010). The first author h-index (h(fa)-index): levelling the field for small and large institute medical and science scholars. *Australasian Physical and Engineering Sciences in Medicine*, 33, 299–300.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates.
- Duro, D. C., Franklin, S. E., & Dubé, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, 118, 259–272.
- Gasparyan, A. Y., Akazhanov, N. A., Voronov, A. A., & Kitas, G. D. (2014). Systematic and open identification of researchers and authors: focus on open researcher and contributor ID. *Journal of Korean Medical Science*, 29, 1453–1456.
- Gasparyan, A. Y., Yessirkepov, M., Gerasimov, A. N., Kostyukova, E. I., & Kitas, G. D. (2016). Scientific author names: Errors, corrections, and identity profiles. *Biochemia Medica*, 26, 169–173.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102, 16569–16572.
- Lafrenière, D., Menuz, V., Hurlimann, T., & Godard, B. (2013). Knowledge dissemination interventions: A literature review. *SAGE Open*. <https://doi.org/10.1177/2158244013498242>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lerchenmueller, M. J., & Sorenson, O. (2016). Author disambiguation in PubMed: Evidence on the Precision and recall of author-ity among NIH-funded scientists. *PLoS ONE*, 11(7), e0158731.
- Mabe, M. A. (2010). Scholarly communication: A long view. *New Review of Academic Librarianship*, 16, 132–144.
- McClelland, S., Mitin, T., Jagsi, R., Thomas, C. R., & Jaboin, J. J. (2018). Importance of first and second authorship in assessing citation-based scholarly activity of US radiation oncology residents and subsequent choice of academic versus private practice career. *Journal of the American College of Radiology*, 15, 1322–1325.
- Negative Binomial Regression | Stata Annotated Output [Internet]. (2019). Available from: <https://stats.idre.ucla.edu/stata/output/negative-binomial-regression/>.
- Negative Binomial Regression | Stata Data Analysis Examples [Internet]. (2019). Available from: <https://stats.idre.ucla.edu/stata/dae/negative-binomial-regression/>.
- Pfeiffer, M., Fischer, M. R., & Bauer, D. (2016). Publication activities of German junior researchers in academic medicine: Which factors impact impact factors? *BMC Medical Education*, 16, 190.
- Post, R. E., Weese, T. J., Mainous, A. G., & Weiss, B. D. (2012). Publication productivity by family medicine faculty: 1999 to 2009. *Family Medicine*, 44, 312–317.
- Prpić, K. (1996). Characteristics and determinants of eminent scientists' productivity. *Scientometrics*, 36, 185–206.
- Schubert, A., & Schubert, G. (2019). All Along the H-Index-Related Literature: A Guided Tour. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handb Sci Technol Indic* (pp. 301–334). Springer Verlag.
- Sebo, P., de Lucia, S., & Vernaz, N. (2020). Scientific publications in internal medicine and family medicine: A comparative cross-sectional study in Swiss university hospitals. *Family Practice*. <https://doi.org/10.1093/fampra/cmaa124>.
- Sebo, P., de Lucia, S., & Vernaz, N. (2021). Gender gap in medical research: a bibliometric study in Swiss university hospitals. *Scientometrics*, 126, 741–755.
- Tian, Y., Zhang, H., Pang, Y., & Lin, J. (2018). Classification for single-trial N170 during responding to facial picture with emotion. *Frontiers in Computational Neuroscience*, 12, 68.
- Vale, R. D. (2015). Accelerating scientific publication in biology. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 13439–13446.

- Vavryčuk, V. (2018). Fair ranking of researchers and research teams. PLoS ONE. <https://doi.org/10.1371/journal.pone.0195509>.
- Waltman, L., & van Eck, N. J. (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, 63, 406–415.
- Zhu, H., & Lakkis, H. (2014). Sample size calculation for comparing two negative binomial rates. *Statistics in Medicine*, 33, 376–387.