



# Which h-index? An exploration within the Web of Science

Guangyuan Hu<sup>1</sup> · Lei Wang<sup>2</sup> · Rong Ni<sup>1</sup> · Weishu Liu<sup>3</sup>

Received: 31 December 2019 / Published online: 4 April 2020  
© Akadémiai Kiadó, Budapest, Hungary 2020

## Abstract

The h-index has attracted wide attention from both scientometricians and science policy makers since it was proposed in 2005. Advocates champion h-index for its simplicity embracing both quantity and quality, while also express concern about its abuse in research evaluation practices and database-dependence attribute. We argue that it is increasingly important to calculate and interpret the h-index precisely along with the rapid evolution of bibliographic databases. In memory of Dr. Judit Bar-Ilan, we join the h-index discussion in *Scientometrics* by further probing a similar “which h-index” question via comparing different versions of h-index within the Web of Science. In this article we put forward the reasons of different WoS h-indices from two perspectives, which are often neglected by bibliometric studies. We suggest that users should specify the details of data sources of h-index calculation for research promotion and evaluation practices.

**Keywords** H-index · Web of Science · Science Citation Index · Research evaluation · Bibliometric analysis

## Introduction

Born with glory, the h-index was first formally initiated by Jorge E. Hirsch in 2005 to measure a scholar’s output from the perspectives of both productivity and quality or, more precisely citation impact (Hirsch 2005). Since then, the h-index has attracted wide attention

---

✉ Weishu Liu  
wslu08@163.com

Guangyuan Hu  
hu.guangyuan@shufe.edu.cn

Lei Wang  
violetwl@163.com

Rong Ni  
nirong007@foxmail.com

<sup>1</sup> School of Public Economics and Administration, Shanghai University of Finance and Economics, Shanghai 200433, China

<sup>2</sup> Financial Research Department, Shanghai National Accounting Institute, Shanghai 200030, China

<sup>3</sup> School of Information Management and Engineering, Zhejiang University of Finance and Economics, Hangzhou 310018, Zhejiang, China

from academia and practitioners (Gingras 2016; Schubert and Schubert 2019). This indicator can be expanded to measure the output of any publication set of any research group or country, etc. (Bornmann and Daniel 2009). Given its limitations such as inter-field differences and insufficient weight to highly cited papers, criticisms sprang up and various h-index variants such as g-index and hg-index, to name just a few, were proposed successively (Alonso et al. 2009; Bornmann et al. 2011; Egghe 2006). Among them a fundamental question is that the h-index should be used without confusion.

In her highly-cited paper entitled “Which h-index? -A comparison of WoS, Scopus and Google Scholar” published in *Scientometrics*, Dr. Judit Bar-Ilan made a comparison of different calculations of h-index by using WoS, Scopus and Google Scholar bibliographic databases respectively (Bar-Ilan 2008).<sup>1</sup> In 2018, Judit Bar-Ilan also rejoined an open debate about the h-index through the *Scientometrics* journal which was triggered by an article entitled “Multiple versions of the h-index: Cautionary use for formal academic purposes” (Bar-Ilan 2018; Bornmann and Leydesdorf 2018; Costas and Franssen 2018; Teixeira da Silva and Dobránszki 2018a, b).

However, a more easily overlooked scenario is that even within Web of Science (WoS) different h-index values can also be generated. Jasco is an early pioneer discussing the names and time coverages of the WoS sub-datasets in computing the h-index (Jasco 2008). With the rapid evolution of WoS in recent years, a clear description of the data source in calculating the h-index is becoming increasingly important. This is particularly true when emerging countries such as China relies heavily on metrics to identify talents and make funding decisions (Tang and Hu 2018). In memory of Dr. Judit Bar-Ilan, we rejoin the discussion on this easily neglected issue in research evaluation practice by probing a similar “which h-index” question within the database of Web of Science.

## WoS, WoSCC and WoK

Web of Science is one of the most adopted data sources for bibliography searching and research evaluation for good or flawed science (Harzing and Alakangas 2016; Tang et al. 2020; Zhu and Liu 2020). But this term has been used interchangeably for different sub-datasets of WoS (Calver et al. 2017). To make sure we are in the same platform on terminology, let us first clarify some easily confusing notions: Web of Science, Web of Science Core Collection (WoSCC), and Web of Knowledge (WoK) and also their correlation with the famous Science Citation Index-Expanded (SCIE).

Currently, WoS is a platform providing access to Clarivate Analytics’ multidisciplinary bibliographic databases.<sup>2</sup> According to Thomson Reuters, the former owner of WoS,<sup>3</sup> the integrated WoS platform was previously known as WoK but renamed WoS in 2014

<sup>1</sup> According to Judit Bar-Ilan’s personal google scholar page, this paper is her most cited paper which has been cited 733 times. Data accessed on 22 November 2019 via [https://scholar.google.com/citations?user=mkb\\_14UAAA&hl=en&oi=sra](https://scholar.google.com/citations?user=mkb_14UAAA&hl=en&oi=sra).

<sup>2</sup> <https://apps.webofknowledge.com/>. Besides, we find a latest comprehensive article focusing on Web of Science during the proof process (Birkle et al. 2020).

<sup>3</sup> [http://wokinfo.com/media/pdf/WoK\\_5-13\\_ReleaseNotes.pdf](http://wokinfo.com/media/pdf/WoK_5-13_ReleaseNotes.pdf).

**Table 1** Web of Science platform

Citation indexes
Web of Science Core Collection
BIOSIS Citation Index
Chinese Science Citation Database <sup>SM</sup>
Data Citation Index
Russian Science Citation Index
SciELO Citation Index
Product Databases
Biological Abstract
BIOSIS Previews
Current Contents Connect
CABI: CAB Abstracts and Global Health®
FSTA®—the food science resource help
Inspec®
KJD-Korean Journal Database
MEDLINE®
Zoological Records
Derwent Innovations Index

(Torres-Salinas and Orduña-Malea 2014). The WoS platform contains citation indexes (including WoSCC), product databases, and Derwent Innovations Index as demonstrated in Table 1.<sup>4</sup>

WoSCC is a core database collection under the WoS platform. It was renamed from WoS in 2014.<sup>5</sup> Along with the expansion and integration of the WoSCC (Liu 2019; Jacso 2018; Rousseau et al. 2018), it now consists of two chemical indexes (i.e. Current Chemical Reactions and Index Chemicus) as well as the following eight citation indexes.

- Science Citation Index Expanded (SCIE).
- Social Sciences Citation Index (SSCI)
- Arts and Humanities Citation Index (A&HCI)
- Conference Proceedings Citation Index-Science (CPCI-S)
- Conference Proceedings Citation Index-Social Sciences and Humanities (CPCI-SSH)
- Book Citation Index-Science (BKCI-S)
- Book Citation Index-Social Sciences and Humanities (BKCI-SSH)
- Emerging Sources Citation Index (ESCI)

However, the phrase WoS is still often used to denote the WoSCC in practice. This phenomenon may be due to the confusion between these two concepts or just for simplification, however, both may introduce confusion.

<sup>4</sup> The contents under the Web of Science platform can be found in [http://images.webofknowledge.com/WOKRS513R8.1/help/WOK/hp\\_whatsnew\\_wok.html](http://images.webofknowledge.com/WOKRS513R8.1/help/WOK/hp_whatsnew_wok.html). More detailed information about the Web of Science platform can be found in <https://clarivate.com/webofsciencegroup/solutions/webofscience-platform/>.

<sup>5</sup> More information can be found in [http://wokinfo.com/media/pdf/WoK\\_5-13\\_ReleaseNotes.pdf](http://wokinfo.com/media/pdf/WoK_5-13_ReleaseNotes.pdf).

## Calculation of the h-index

### Which WoS?

#### Different h-index values for different WoS

The first factor that influences the calculation of the h-index is the identification of all the publications belonging to an entity. Two different scenarios may happen. Firstly, the phrase WoS may denote the WoS platform. However, different institutions may choose to subscribe to different database packages according to their personalized demand. When scholars search with WoS platform's all databases search setting, different results may arise if accessed from different institutions. That is to say, different database package subscriptions under the WoS platform may generate different h-index values.

Secondly, the phrase WoS may also refer to WoSCC. A recent study has shown that different institutions may subscribe to different sub-datasets of WoSCC and also with varying years of coverage (Liu 2019). That is to say, the WoSCC-based h-index values may also be different when calculated in different institutions. Unfortunately, many scholars haven't specified the details of the sub-datasets when using the WoSCC as the data source (Dallas et al. 2018; Liu 2019).

#### Non-transparent calculation of the h-index: an empirical evidence

To have a better understanding of how prevalent this ambivalent situation is, we manually check a sample of SCIE/SSCI indexed publications over the period of 2017 and 2019 and examined their calculations of h-index by using WoS.<sup>6</sup> We use "Web of Science" and "h index" as the keywords to search in topic field and limit the citation indexes to SCIE and SSCI only.<sup>7</sup> 137 records published from 2017 to 2019 were retrieved. We further restricted the publishing language to English and 129 records were left. We ended up with 127 records with full text available for further analysis.<sup>8</sup>

Two authors read the full texts of these 127 records and tabulated how h-index was calculated if documented. Our examination showed that 99 out of the 127 records used the data from the WoS to calculate h-index. Table 2 summarizes their distribution by journal sources.

Yet over 40% of our sample (47 out of the 99 records) did not specify which sub-datasets of the WoS were adopted to calculate the h-index, including those professional publications in the category of Information Science and Library Science.

### Which citation count?

The second factor that influences the value of h-index is time cited counts of an entity's publications. One record's citation counts from Google Scholar, Scopus, and WoS are usually different (Martín-Martín et al. 2018), however, similar scenario also exists in WoS.

<sup>6</sup> The phrase WoS in this section denotes WoS platform or WoSCC, as appropriate.

<sup>7</sup> Since the data were accessed on Nov. 7 2019, not all the publications in 2019 were covered.

<sup>8</sup> We have tried various ways to find the full texts of all the 129 records including writing letters to the corresponding authors, and we ended up with 127 publications with full texts accessible.

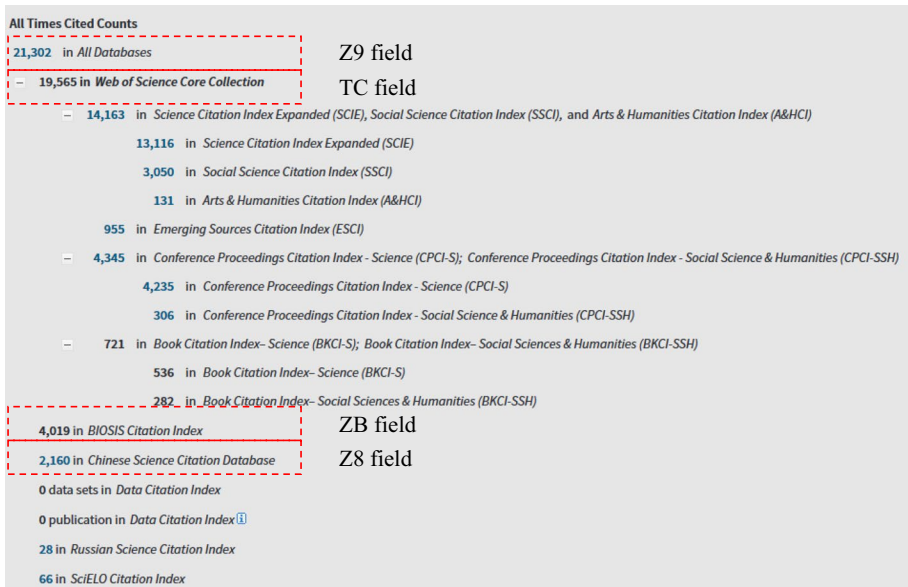
**Table 2** Journals haven't specified the details of the sub-datasets of WoS when calculating the h-index

Journal	Records have used data from WoS to calculate the h-index	Records haven't specified the details of the sub-datasets of WoS
Scientometrics	8	5
PLoS One	5	4
Current Science	2	2
World Neurosurgery	2	2
Medicine	3	2
Journal of Informetrics	2	1
Academic Radiology	1	1
Journal of Intelligent and Fuzzy Systems	1	1
International Journal of Health Geographics	1	1
Journal of Vascular Surgery	1	1
Sustainability	2	1
American Journal of Dermatopathology	1	1
Journal of Hepato-Biliary-Pancreatic Sciences	1	1
Pediatric Surgery International	1	1
Technological Forecasting and Social Change	1	1
Orthopaedic Journal of Sports Medicine	1	1
Physica A-Statistical Mechanics and Its Applications	1	1
International Journal of Precision Engineering and Manufacturing-Green Technology	1	1
Journal of Medical Systems	1	1
Academic Psychiatry	1	1
Journal of Global Health	1	1
Spine Journal	1	1
Annals of Surgical Oncology	1	1
Croatian Medical Journal	1	1

**Table 2** (continued)

Journal	Records have used data from WoS to calculate the h-index	Records haven't specified the details of the sub-datasets of WoS
Plastic and Reconstructive Surgery	1	1
Medical Science Monitor	1	1
Transinformacao	2	1
Physica Status Solidi B-Basic Solid State Physics	1	1
Journal of Alzheimers Disease	1	1
Malaysian Journal of Library and Information Science	1	1
Journal of Neurosurgery	1	1
American Journal of Pharmaceutical Education	1	1
Journal of Plastic Surgery and Hand Surgery	1	1
Journal of the Royal Army Medical Corps	1	1
Molecules	1	1
International Journal of Mental Health Systems	1	1
Substance Abuse Treatment Prevention and Policy	1	1

The details of the used sub-datasets of WoS refer to the names of the used sub-datasets. Since many studies generally only report that they focus on publications of a specific period, we lessen the requirement on the reporting of sub-dataset's year of coverage. Only English SCIE/SSCI records are considered. Data accessed on Nov. 7 2019



**Fig. 1** Citation counts of one Nature paper. *Note:* Data accessed on Dec, 2 2019 from the WoS platform

Though the help file of WoS states “If you view Times Cited for a record from anywhere in the world, the value is always the same”,<sup>9</sup> different versions of one record’s citation count also exist in WoS.

According to the help file of WoS all databases, at least four citation count related field tags are provided: TC (times cited from WoSCC), Z8 (times cited from Chinese Science Citation Database), ZB (times cited from BIOSIS Citation Index), and Z9 (times cited from all citation indexes under the WoS platform).<sup>10</sup>

Figure 1 demonstrates that different versions of citation count of one Nature paper searched through the WoS platform. The article we chose is titled *Collective dynamics of ‘small-world’ networks*. As shown in the red brackets of Fig. 1, there exist large differences among different citation counts for the same paper. Even downloading the bibliographic data from the WoSCC, two citation count field tags are also provided (TC and Z9).<sup>11</sup> Therefore, different versions of citation count also influence the calculation of the h-index. Users generally use the TC field or the Z9 field to denote the citation count, however, many of them including the authors of this letter always haven’t specified which version of citation count is used. We also check the full texts of the abovementioned 99 records, but most of them haven’t mentioned this point.

<sup>9</sup> [https://images.webofknowledge.com/images/help/WOS/hp\\_times\\_cited\\_count.html](https://images.webofknowledge.com/images/help/WOS/hp_times_cited_count.html).

<sup>10</sup> [https://images.webofknowledge.com/images/help/WOK/hs\\_alldb\\_fieldtags.html](https://images.webofknowledge.com/images/help/WOK/hs_alldb_fieldtags.html).

<sup>11</sup> [https://images.webofknowledge.com/images/help/WOS/hs\\_wos\\_fieldtags.html](https://images.webofknowledge.com/images/help/WOS/hs_wos_fieldtags.html).

## Conclusion

With the rapid update of the WoS in recent years, the simple statement of the use of WoS as the data source without clarifications will bring about confusion and inconsistency in metrics. This study tries to distinguish the concepts of WoK, WoS and WoSCC and further reveals the fact that different WoS may generate different h-index values. We argue that h-index, despite its deficiencies, if used, should be at least used consistently by detailing how they are calculated. We hope to remind both bibliometricians and research evaluators of the need to pay attention to the possible various h-index values even within WoS database.

Twelve years after Dr. Judit Bar-Ilan's high-cited paper on "which h-index", this article expands the discussion on "which h-index" in research evaluation, but within one widely utilized bibliographic database, WoS. Similarly, for other database-dependent metrics, this phenomenon also exists. Given the increasing use of various bibliographic databases, the features and also limitations of each database should be expressed explicitly (Falagas et al. 2008; Liu 2017; Liu et al. 2018; Tang et al. 2017; Zhu et al. 2019). We write this paper, partly in memory of Dr. Judit Bar-Ilan; while at the same time, we suggest that researchers and evaluation practitioners should pay attention to the details of data sources especially when using the WoS (Dallas et al. 2018; Liu 2019).

**Acknowledgements** This research is financially supported by the National Natural Science Foundation of China (#71801189 and #71904168) and Zhejiang Provincial Natural Science Foundation of China (#LQ18G030010 and #LQ18G010005). All of the views expressed here are those of the authors who also take full responsibility for any errors.

## Compliance with ethical standards

**Conflict of interest** The authors declare that there is no conflict of interest.

## References

- Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). H-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, 3(4), 273–289. <https://doi.org/10.1016/j.joi.2009.04.001>.
- Bar-Ilan, J. (2008). Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257–271. <https://doi.org/10.1007/s11192-008-0216-y>.
- Bar-Ilan, J. (2018). Comments on the Letter to the Editor on "Multiple versions of the h-index: Cautionary use for formal academic purposes" by Jaime A. Teixeira da Silva and Judit Dobránszki. *Scientometrics*, 115(2), 1115–1117. <https://doi.org/10.1007/s11192-018-2681-2>.
- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), 363–376. [https://doi.org/10.1162/qss\\_a\\_00018](https://doi.org/10.1162/qss_a_00018).
- Bornmann, L., & Daniel, H. D. (2009). The state of h index research. *EMBO Reports*, 10(1), 2–6. <https://doi.org/10.1038/embor.2008.233>.
- Bornmann, L., & Leydesdorff, L. (2018). Count highly-cited papers instead of papers with h citations: use normalized citation counts and compare "like with like"! *Scientometrics*, 115(2), 1119–1123. <https://doi.org/10.1007/s11192-018-2682-1>.
- Bornmann, L., Mutz, R., Hug, S. E., & Daniel, H. D. (2011). A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics*, 5(3), 346–359. <https://doi.org/10.1016/j.joi.2011.01.006>.



- Calver, M. C., Goldman, B., Hutchings, P. A., & Kingsford, R. T. (2017). Why discrepancies in searching the conservation biology literature matter. *Biological Conservation*, 213, 19–26. <https://doi.org/10.1016/j.biocon.2017.06.028>.
- Costas, R., & Franssen, T. (2018). Reflections around 'the cautionary use' of the h-index: Response to Teixeira da Silva and Dobránszki. *Scientometrics*, 115(2), 1125–1130. <https://doi.org/10.1007/s11192-018-2683-0>.
- Dallas, T., Gehman, A. L., & Farrell, M. J. (2018). Variable bibliographic database access could limit reproducibility. *BioScience*, 68(8), 552–553. <https://doi.org/10.1093/biosci/biy074>.
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152. <https://doi.org/10.1007/s11192-006-0144-7>.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal*, 22(2), 338–342. <https://doi.org/10.1096/fj.07-9492LSF>.
- Gingras, Y. (2016). *Bibliometrics and research evaluation: Uses and abuses*. Cambridge: MIT Press.
- Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. <https://doi.org/10.1007/s11192-015-1798-9>.
- 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>.
- Jasco, P. (2008). The pros and cons of computing the h-index using Web of Science. *Online Information Review*, 32(5), 673–688. <https://doi.org/10.1108/14684520810914043>.
- Jasco, P. (2018). The scientometric portrait of Eugene Garfield through the free ResearcherID service from the Web of Science Core Collection of 67 million master records and 1.3 billion references. *Scientometrics*, 114(2), 545–555. <https://doi.org/10.1007/s11192-017-2624-3>.
- Liu, W. (2017). The changing role of non-English papers in scholarly communication: Evidence from Web of Science's three journal citation indexes. *Learned Publishing*, 30(2), 115–123. <https://doi.org/10.1002/leap.1089>.
- Liu, W. (2019). The data source of this study is Web of Science Core Collection? Not enough. *Scientometrics*, 121(3), 1815–1824. <https://doi.org/10.1007/s11192-019-03238-1>.
- Liu, W., Hu, G., & Tang, L. (2018). Missing author address information in Web of Science—An explorative study. *Journal of Informetrics*, 12(3), 985–997. <https://doi.org/10.1016/j.joi.2018.07.008>.
- Martín-Martín, A., Orduña-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, 12(4), 1160–1177. <https://doi.org/10.1016/j.joi.2018.09.002>.
- Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming metric-wise: A bibliometric guide for researchers*. Cambridge, MA: Chandos Publishing. <https://doi.org/10.1016/C2017-0-01828-1>.
- 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 handbook of science and technology indicators*. Springer handbooks. Cham: Springer. [https://doi.org/10.1007/978-3-030-02511-3\\_12](https://doi.org/10.1007/978-3-030-02511-3_12).
- Tang, L., & Hu, G. (2018). Evaluation woes: Metrics beat bias. *Nature*, 559(7714), 331. <https://doi.org/10.1038/d41586-018-05751-4>.
- Tang, L., Hu, G., & Liu, W. (2017). Funding acknowledgment analysis: Queries and caveats. *Journal of the Association for Information Science and Technology*, 68(3), 790–794. <https://doi.org/10.1002/asi.23713>.
- Tang, L., Hu, G., Sui, Y., Yang, Y., & Cao, C. (2020). Retraction: The other face of collaboration? *Science and Engineering Ethics*. <https://doi.org/10.1007/s11948-020-00209-1>.
- Teixeira da Silva, J. A., & Dobránszki, J. (2018a). Multiple versions of the h-index: Cautionary use for formal academic purposes. *Scientometrics*, 115(2), 1107–1113. <https://doi.org/10.1007/s11192-018-2680-3>.
- Teixeira da Silva, J. A., & Dobránszki, J. (2018b). Rejoinder to “Multiple versions of the h-index: Cautionary use for formal academic purposes”. *Scientometrics*, 115(2), 1131–1137. <https://doi.org/10.1007/s11192-018-2684-z>.
- Torres-Salinas, D., & Orduña-Malea, E. (2014). Ruta dorada del open access en Web of science. *Anuario ThinkEPI*, 8, 211–214.
- Zhu, J., Hu, G., & Liu, W. (2019). DOI errors and possible solutions for Web of Science. *Scientometrics*, 118(2), 709–718. <https://doi.org/10.1007/s11192-018-2980-7>.
- Zhu, J., & Liu, W. S. (2020). A tale of two databases: The use of Web of Science and Scopus in academic papers. *Scientometrics*. <https://doi.org/10.1007/s11192-020-03387-8>.