The American Journal of Surgery (2013) 206, 560-566

The American Journal of Surgery®

Clinical Science

The use of the Hirsch index in benchmarking hepatic surgery research

Alessandro Cucchetti, M.D.*, Federico Mazzotti, M.D., Sara Pellegrini, M.D., Matteo Cescon, M.D., Lorenzo Maroni, M.D., Giorgio Ercolani, M.D., Antonio Daniele Pinna, M.D.

Department of Medical and Surgical Sciences - DIMEC, Alma Mater Studiorum, University of Bologna, S.Orsola - Malpighi Hospital, Via Massarenti 9, Bologna 40138, Italy

KEYWORDS:
Hepatic surgery;
Surgical therapy;
H-index;
Academic research;
Bibliometric measures

Abstract

BACKGROUND: The Hirsch index (h-index) is recognized as an effective way to summarize an individual's scientific research output. However, a benchmark for evaluating surgeon scientists in the field of hepatic surgery is still not available.

METHODS: A total of 3,251 authors who published between 1949 and 2011 were identified using the Scopus identification number. The h-index, the total number of cited document, the total number of citations, and the scientific age were calculated for each author using both Scopus and Google Scholar.

RESULTS: The median h-index was 6 and the median scientific age, assessed with Google Scholar, was 19 years. The numbers of cited documents, numbers of citations, and h-indexes obtained from Scopus and Google Scholar showed good correlation with one another; however, the results from the 2 databases were modified in different ways by scientific age. By plotting scientific age against h-index percentiles an h-index growth chart for both Scopus database and Google Scholar was provided.

CONCLUSIONS: This analysis provides a first benchmark to assess surgeon scientists' productivity in the field of liver surgery.

Rapid progress in hepatic surgery over the past decades has been accompanied by vast increases in research publications. Determination of the quality and quantity of this published work is considered critical in the evaluation and comparison of scientists, both for employment purposes and for funding and resource allocation. The assessment of an individual's research output remains problematic, and there was a lack of consensus as to how to optimally score individual research performance.[1,2] New statistics have recently been proposed to provide more balanced methods of quantifying publication records. These include the h-index,[2,3] the g-index,[4] and the age-weighted citation rate (AWCR).[5] Among these, the citation metric that has been most actively debated, studied, and adopted is the h-index (Hirsch index). Hirsch's h-index is defined as the number of *h* publications cited at least *h* times in the literature.[3] It can quantify and predict research output and incorporates productivity and relevance of a body of work into a single statistic.[3,6] Even if criticized,[7] the h-index is now recognized as a simple and effective way to summarize an individual's scientific research output.[8] The popularity of this measure is demonstrated by the fact that Hirsch's original article has already received over 400 citations and over

The authors declare no conflict of interest.

* Corresponding author. Tel.: +39-051-6363721; fax: +39-051-304902.

E-mail address: aleqko@libero.it

Manuscript received October 17, 2012; revised manuscript January 11, 2013

0002-9610/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amjsurg.2013.01.037

110,000 downloads.[9] The method is increasingly used by academic, research, and federal institutions worldwide for research policymaking, monitoring of scientific developments, and comparisons between institutions, as well as individual scientists.[10,11]

As with any bibliometric indicator, however, a scientist's h-index must be viewed strictly in the context of the specialty area within which he or she is working and should not be extrapolated for comparison with the output of those operating in other scientific areas. This study aimed to establish the spread of h-indexes for those specifically working in the hepatic surgery field to set standards and provide a benchmark for evaluating individuals in this area.

Methods

Author identification

The first part of the study concerned the extraction of authorship data from hepatic surgery–related studies indexed in the MEDLINE and Embase databases from 1949 to 2011 inclusive. In order to focus on the surgeons involved, rather than other physician figures, the literature research was limited to the 4 top citation index general surgery journals based on the highest impact factors according to the *Journal Citation Report* of 2010.[12] MEDLINE and Embase were searched using the keywords "hepatic resection" or "hepatectomy" in the title or in the abstract text. The literature search was also limited to the English language and human species, resulting in a total of 1,015 unique articles checked for author selection. Author identification was performed up to the 14th position in the collaborators' list, resulting in a total of 3,377 authors. An iterative approach was adopted to identify homonyms and duplicate authors. Each single author was checked for Scopus identification number by comparing name, middle name, surname, affiliation, and article title obtained with the literature research. This approach lead to the exclusion of 126 duplicates and the final study sample consisted of 3,251 unique authors.

Measure of bibliometric indexes

In the second part of the study, the number of cited documents, the total number of citations, and the h-index were calculated for each of the selected authors using both Google Scholar and Scopus. The h-index is based on the set of the scientist's most cited papers and the number of citations the scientist has received in other publications. The h-index is the greatest number of articles an individual is the author or coauthor of that have each been cited h or more number of times. An h-index of 10, therefore, indicates that a scientific author has 10 publications, each with at least 10 citations. Google Scholar was searched using Harzing's software program Publish or Perish.[13] The program analyzes raw citations from Google Scholar and calculates the h-index and other statistics from the data. The AWCR index and publication age were also calculated using the Publish or Perish program: the AWCR is the average number of citations received by articles published by a single author, weighted by the age of the articles. In particular, it is the square root of the sum of all age-weighted citation counts over all articles that contribute to the h-index, and theoretically this allows younger and yet less cited papers to contribute to the AWCR, although they may not yet contribute to the h-index.[5] Publication age is the time between the first and the last publication (also called the scientific age).

The surname and first initial of each author was used for the search from both Scopus and Google Scholar. Scopus research was limited to the Health Sciences subject area and included the terms "hepatic resection" OR "hepatectomy" OR "liver transplantation" in the title or abstract. Google Scholar research was limited to the Medicine, Pharmacology, Veterinary Sciences fields and included the terms "hepatic resection" OR "hepatectomy" OR "liver transplantation" in the search field "All of the words." After the query results were reported, the affiliation and each reference were further reviewed to ensure that the results correctly included publications by the intended author by noting the journal in which it was published and, if necessary, linking to the article to review it. Any incorrect references were removed and the h-index recalculated.

Statistical analysis

All data were transferred to Microsoft Excel and statistical analysis was performed using the Statistical Package for Social Sciences version 13.0 (SPSS, Chicago, IL). Continuous data were explored for normality of their distribution using the Kolmogorov–Smirnov test: because no normal distribution was observed for any of the data analyzed, all the subsequent analyses were nonparametric. Data were thus reported in median and ranges, and relationships between each bibliographic measure were considered and outcome measures were explored with the Spearman correlation (rho). The Wilcoxon signed-rank test and the Kruskal-Wallis H test were also applied as appropriate. A P value of less than .05 was considered statistically significant in all the analyses.

Results

The distribution of the various bibliographic measures considered, among the 3,251 authors identified, is reported in Table 1. Briefly, the median number of cited documents per surgeon was 13, with both Scopus database and Google Scholar: the median number of citations per surgeon was 229 with the Scopus database and 313 with Google Scholar. The median h-index was 6 with both Scopus and Google Scholar and the median scientific age, assessed with Google Scholar, was 19 years (range, 0 to 57). Fig. 1

Table 1 Distributions of the bibliographic measures

Bibliographic measure	Authors considered (N = 3,251)
Scopus Cited Documents	
Median (range)	13 (1–938)
25th/75th/95th percentile	4/37/130
Scopus Number of Citations	
Median (range)	229 (0–18,004)
25th/75th/95th percentile	62/742/2,754
Scopus H-Index	
Median (range)	6 (0–81)
25th/75th/95th percentile	3/13/26
Google Scholar Cited Documents	
Median (range)	12 (1–1,000)
25th/75th/95th percentile	5/38/172
Google Scholar Number of Citations	
Median (range)	313 (0–26,377)
25th/75th/95th percentile	88/968/4,006
Google Scholar H-Index	
Median (range)	6 (0–117)
25th/75th/95th percentile	2/14/32
Google Scholar G-Index	
Median (range)	11 (0–182)
25th/75th/95th percentile	4/26/58
Google Scholar AWCR	
Median (range)	35.3 (0–3,016)
25th/75th/95th percentile	9.6/99.8/415.1
Scientific Age (years)	
Median (range)	19 (0–57)
25th/75th/95th percentile	12/28/46
0–10 years	677 (20.8%)
11–20 years	1137 (35.0%)
21–30 years	729 (22.4%)
>30 years	708 (21.8%)

AWCR = age-weighted citation rate; h-index = Hirsch index.

reports the relationships observed between numbers of cited documents (Fig. 1A), numbers of citations (Fig. 1B), and h-indexes (Fig. 1C) obtained from Scopus and Google Scholar. All 3 bibliographic measures showed good correlations when results from Scopus and Google Scholar were compared; in particular, the number of cited documents had an rho of .821 (P = .001), the number of citations an rho of .815 (P = .001), and the h-index an rho of .838 (P = .001). Table 2 reports the correlations between each bibliographic measure considered and h-indexes calculated with both Scopus and Google Scholar. Of note, results from the same database demonstrated very good correlations, whereas scientific age showed a weak monotonic relationship with h-index (an rho of .345 vs Scopus h-index and .500 vs Google h-index).

The role of scientific age

Results were found to be modified in different ways by scientific age, as can be observed from Table 3. Regarding

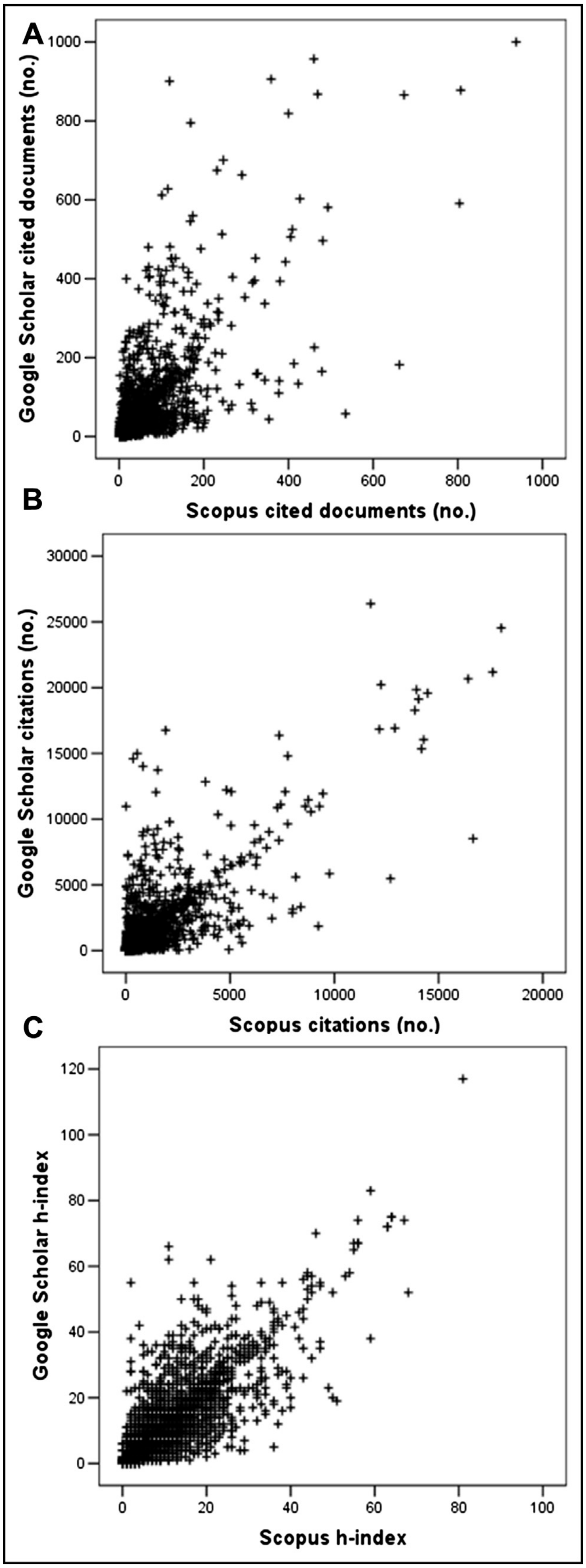

Figure 1 Relationships observed between Scopus and Google Scholar bibliometric measures: numbers of cited documents (A), numbers of citations (B), and Hirsch index (h-index) (C).

Table 2 Relationships observed between each bibliographic measure considered and Google Scholar and Scopus Hirsch indexes

Bibliographic measure	Scopus h-index	Google Scholar h-index
Scopus total cited documents	.945	.827
Scopus total citation	.917	.762
Google Scholar total cited documents	.780	.958
Google Scholar total citation	.770	.914
Google Scholar ACWR	.768	.912
Scientific age	.345	.500

Values represent Spearman correlation (rho).
AWCR = age-weighted citation rate; h-index = Hirsch index.

Scopus results, the median number of cited documents ranged from 5, for authors with a scientific age less than 10 years, to 23 for authors with a scientific age greater than 30 years ($P = .001$), with the median number of citations from 59 to 602 ($P = .001$) and the median h-index from 3 to 9 ($P = .001$). A similar relationship was observed for Google Scholar results: the median number of cited documents ranged from 4 to 46.5 ($P = .001$), the median number of citations from 52 to 1,017 ($P = .001$), and the median h-index from 2 to 15 ($P = .001$), respectively. These 3 bibliographic measures change according to the database used for calculation and the scientific age. In particular, it should be noted that for authors with a scientific age less than 10 years, the number of cited documents ($P = .001$), number of citations ($P = .019$), and h-index ($P = .001$) are higher, even if only slightly, with Scopus when compared with Google Scholar results. These differences become less pronounced for authors with a scientific age between 10 and 20 years ($P = .049$, .001, and .020, respectively) and, from the 21st year onward, Google Scholar gives higher values than Scopus database results ($P = .001$ in all cases).

The Hirsch index growth curve

In Fig. 2, scientific age was plotted against h-indexes and values interpolated to obtain an h-index growth chart for both the Scopus database (Fig. 2A) and Google Scholar (Fig. 2B). These growth charts consist of a series of

Table 3 Relationships observed between each bibliographic measure considered and Google Scholar and Scopus databases by scientific age

Bibliographic measure	Scopus	Google Scholar	P*	Rho†
Scientific age <10 years (n = 677)				
Median cited documents (range)	5 (1–164)	4 (1–90)	.001	.817
25th/75th/95th percentile	2/12/38	2/9/26		
Median number of citations	59 (0–2,437)	52 (0–2,065)	.019	.810
25th/75th/95th percentile	20/175/547	14/158/535		
Median h-index	3 (0–29)	2 (0–20)	.001	.785
25th/75th/95th percentile	2/5/11	1/5/10		
Scientific age 11–20 years (n = 1,137)				
Median cited documents (range)	12 (1–423)	11 (1–296)	.049	.805
25th/75th/95th percentile	3/30/93	5/23/69		
Median number of citations	273 (0–8,875)	240 (0–14,802)	.001	.609
25th/75th/95th percentile	84/657/1,833	127/717/2,042		
Median h-index	6 (0–47)	5.5 (0–57)	.020	.829
25th/75th/95th percentile	3/12/21	2/11/21		
Scientific age 21–30 years (n = 729)				
Median cited documents (range)	22 (1–535)	23 (1–906)	.001	.831
25th/75th/95th percentile	8/60/169	9/57/185		
Median number of citations	437 (0–14,039)	617 (0–19,846)	.001	.641
25th/75th/95th percentile	144/1,181/4,045	214/1,409/4,493		
Median h-index	9 (0–64)	10 (0–75)	.001	.838
25th/75th/95th percentile	5/17/34	5/18/34		
Scientific age >30 years (n = 708)				
Median cited documents (range)	23 (1–938)	46.5 (1–1,000)	.001	.883
25th/75th/95th percentile	9/69/194	12/120/404		
Median number of citations	602 (0–18,004)	1,017 (0–26,377)	.001	.659
25th/75th/95th percentile	109/1,176/4,348	290/2,464/8,149		
Median h-index	9 (0–81)	15 (0–117)	.001	.786
25th/75th/95th percentile	4/17/33	6/24/44		

h-index = Hirsch index.
*P values are referred to the Wilcoxon signed-rank test.
†Rho is the results of Spearman correlation between Scopus and Google Scholar results.

percentile curves that illustrate the distribution of the h-index, obtained with Scopus or Google Scholar, in relationship with scientific age, in order to track how the h-index can change over time. Fig. 2C reports the difference in the median curve (50th percentile) between the 2 databases. Considering an author with a scientific age of 5 years and an h-index of 5, his productivity can be placed above the 75th percentile of the hepatic surgery scientific community with both Scopus and Google Scholar methods, meaning that this author had a very good scientific production. Considering an author with a scientific age of 10 years and an h-index of 18, his productivity can be placed above the 95th percentile of the hepatic surgery community, meaning that this author had an excellent scientific production. Of particular note is that percentile curves were different between Scopus's and Google Scholar's h-indexes. In fact, as can be observed from Fig. 2C, the difference between the 50th percentile h-indexes of Scopus and Google Scholar was minimal within the first 10 to 15 years, after which Google Scholar provided higher h-indexes than the Scopus database.

The age-weighted citation rate

The possibility of younger and yet less cited articles being included in a different bibliographic measure was explored by analyzing the age-weighted citation rate. The scatter-plot of Fig. 3A effectively shows that the AWCR can handle scientific age in a single measure (rho .358). However, as can be observed from Fig. 3B,C, the AWCR was found to be strictly related to both the Google Scholar number of citations (rho .960) and the h-index (rho .912).

Comments

The assessment of academic research represents a critical issue of investigation, because the determination of the quality and quantity of published work is of paramount importance both for employment purposes and for funding allocation.[1–3] In the field of hepatic surgery, a metric of comparison is still not available; nonetheless, such a benchmark seems to be important in the presence of the remarkable development of liver surgery observed in recent decades.[14] Results from this study could suggest standards for evaluating individuals in this specific surgical area. The literature research and the author selection for the h-index calculation of the present study identified surgeon scientists on the basis of their h-index and their scientific age in a growth chart, useful for the assessment of scientific research in the field of hepatic surgery worldwide.

There are some significant results from this study that deserve particular attention. Attention must be first focused on the difference observed in h-indexes obtained from Scopus and Google Scholar databases in relationship with scientific age. A weak monotonic relationship was found between scientific age and h-index; in fact, after a scientific age of about 15 to 20 years, Google Scholar gives higher h-indexes than the Scopus database. This is a result of the literature availability being from only 1996 for the h-index calculation of Scopus. This means that comparisons can be reliably performed between the 2 databases only for younger

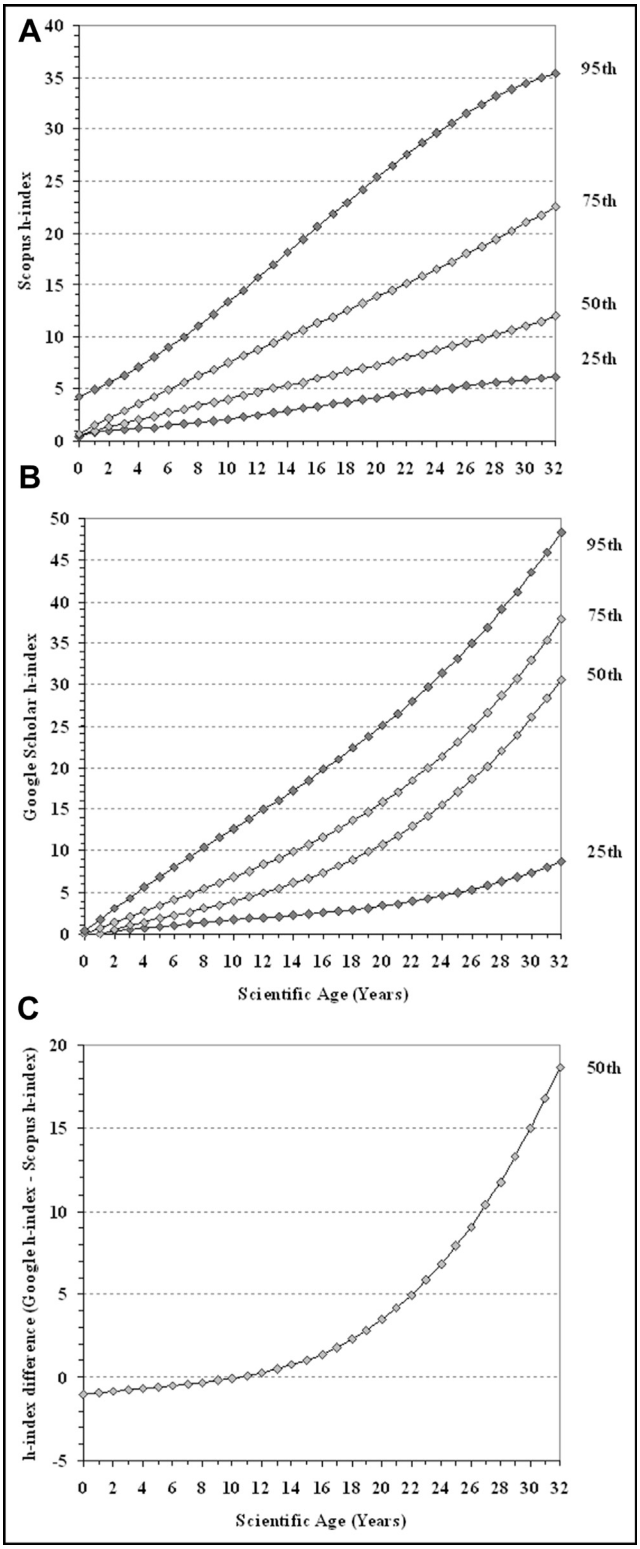

Figure 2 Hirsch-index (h-index) growth chart for Scopus database (A) and Google Scholar (B). (C) Reports the difference in the median curve (50th percentile) between the 2 databases.

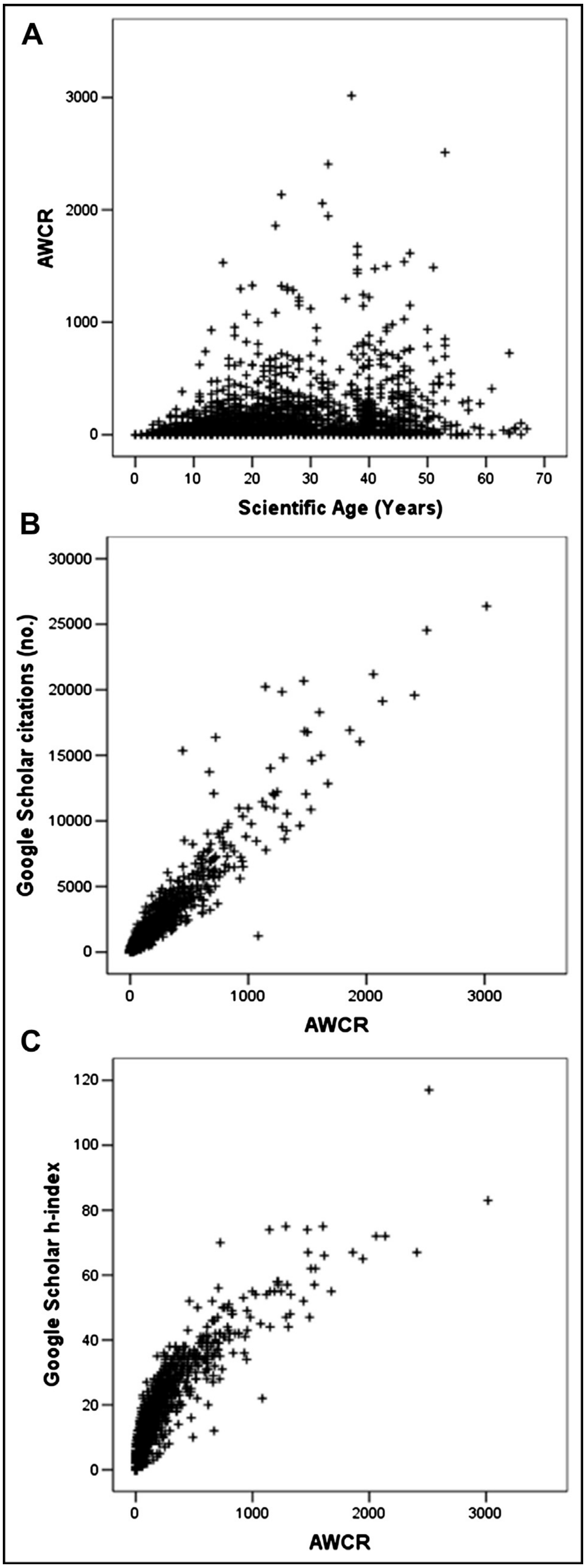

Figure 3 Relationship observed between the age-weighted citation rate (AWCR) and scientific age (A; rho .358), number of citations (B; rho .960), and the Hirsch index (h-index) (C; rho .912).

surgeon scientists; on the contrary, this approach will give an incorrect comparison in older authors. These observations support the suggestion of using the same database for academic research comparison, especially for older authors. Looking at the growth h-index curves provided, the finding that there are some younger authors (around the 95th percentile) who have an h-index that grossly corresponds to the 25th percentile of older ones is of interest. This finding indicates that there are relatively young surgeons who have already produced a relevant amount of scientific works, comparable, or even higher, to that of older surgeons. Thus, the chronological age of a scientist should not represent an obstacle for employment in the academic field, funding allocation, or career progression. This observation supports the need to assess academic activity on the number of high-quality research reports rather than on academic age. Application of a global surgical benchmark would allow the constructive appraisal of academic surgery to achieve its future goals, especially among young scientists.[14]

The second result that should attract attention is the relationship observed between the various bibliometric measures considered and the databases used. In fact, good correlations were found between the number of cited documents, the number of citations, and the h-index in both Scopus and Google Scholar databases. It should also be noted that the age-weighted citation was found to be strictly related to both h-index and the number of citations. These findings mean that all bibliometric measures that are commonly used to assess relevance of a body of work have similar accuracy; however, in a short period of time, the h-index has become a widely recognized measure of quantifying an individual's research output and to compare academic and public institutions with one another.[4–11] Given its widespread use, the proposed h-index benchmark provides a useful metric in the comparison of research in the present surgical field, with important practical aspects in the evaluation of the academic career of surgeons. In comparing the h-index of candidates by means of the present growth curves, institutions would be able to identify the candidate who is best qualified for career advancement: for example, in Italy, the National Agency for the Evaluation of Universities and Research Institutes established in 2012 a national scientific evaluation for career advancement toward associate professor or full professor based on the h-index.[15] Similarly, centers of academic surgical excellence can be recognized and targeted as sites of significant academic recruitment and research output.[14]

Some comments should be reserved for the limits of the h-index.[16] First, this metric may not fully capture the quality of a scientific research: a published article could be cited not because of its quality but rather because of its poor value that raised harsh criticism. Unfortunately, this represents an unavoidable limit of this citation metric. Second, the typical interfield differences in the h-index values, caused by differences among fields in productivity and citation practices, can also be found when the field was selected, as was found in this analysis. In fact, in the area of hepatic surgery, intrafield differences could probably be expected among surgeons who perform only general hepatic surgery and surgeons involved in liver transplantation, and this is an aspect that probably requires further dedicated investigations. Third, the h-index can be affected by journal impact factor or citation index. In fact, it can be expected that the higher the rank of a scientific journal, the higher the probability would be of a specific article, published in such a journal, being cited by other authors because such a journal would be considered more authoritative in the field. An indirect measure of such a relationship can probably be obtained from calculating the g-index, which aims to improve the h-index by giving more weight to highly cited articles, which articles published in more authoritative scientific journals can be. This represents another aspect that deserves further analysis.[4] All these observations highlight the need for caution in the use of the h-index, avoiding indiscriminate use, because research performance has to be considered a complex multifaceted endeavor that should not be underestimated.[17]

The literature research and the author selection used in the present study can have potential shortcomings. It is possible that some authors, who have significantly contributed to the growth of liver surgery, were left out from the search criteria. It is possible, in fact, that some surgeons, who have contributed to liver transplantation development, may not have been included because they published in journals other than those included, for example, transplantation targeted journals. The exclusion of transplantation journals in the inclusion criteria was determined a priori to avoid the inclusion of other physician figures, such as anesthesiologists or hepatologists.[11] It should be noted, however, that is fairly unlikely that a surgeon who has really contributed to the development of liver surgery has never published a scientific article in 1 of the 4 most important surgical journals. This does not mean that these figures have not played a decisive role in the growth of liver surgery, but it contrasts with the aim of this study, which was to identify surgeons who have shaped liver surgery as we now know it. Attention should also be given to the other databases for h-index calculation such as ISI's Web of Knowledge database[12] or Firefox's add-in for Google Scholar.[18] Although it was not the aim of this study to compare h-index calculation to other databases, this feature probably represents an issue that deserves further investigation. Another item to note would be the potential bias represented by homonyms, which in this study were kept to a minimum because the Scopus identification number was applied for author identification, and all references were accurately checked and any incorrect references were removed.

Conclusion

This study provides an original first benchmark useful to assess surgeon scientists' productivity in the field of liver surgery. The h-index can be considered an accurate tool in the determination of the quality and quantity of published work. Because the h-index calculation is influenced by the database used, the report of a scientist's h-index must include the database used for computation.

Acknowledgment

The authors thank Susan West for her writing assistance. The h-index of the corresponding Author, calculated with Google Scholar, can be found here: http://scholar.google.it/citations?user=6V_7iUEAAAAJ&hl=it

References

1. Ball P. Index aims for fair ranking of scientists. Nature 2005;436:900.
2. Hirsch JE. Does the H index have predictive power? Proc Natl Acad Sci USA 2007;104:19193–8.
3. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA 2005;102:16569–72.
4. Egghe L. Theory and practice of the g-index. Scientometrics 2006;69:131–52.
5. Jin B. The AR-index: complementing the h-index. ISSI Newsletter 1997;3:6.
6. Lee J, Kraus KL, Couldwell WT. Use of the h index in neurosurgery. J Neurosurg 2009;111:387–92.
7. Horne R, Petrie KH, Wessely S. H-index pathology: implications for medical researchers and practitioners. BMJ 2009;339:b5356.
8. Ball P. Achievement index climbs the ranks. Nature 2007;448:737.
9. Fersht A. The most influential journals: impact factor and Eigenfactor. Proc Natl Acad Sci USA 2009;106:6883–4.
10. Podlubny I, Kassayova K. Law of the constant ratio. Towards a better list of citation superstars: compiling a multidisciplinary list of highly cited researchers. Res Eval 2006;15:154–62.
11. Poynard T, Thabut D, Jabre P, et al. Ranking hepatologists: which Hirsch's h-index to prevent the "e-crise de foi-e"? Clin Res Hepatol Gastroenterol 2011;35:375–86.
12. Thomson Reuters Community Forum. Available at: http://community.thomsonreuters.com/. Accessed January 2, 2012.
13. Harzing, AW. Publish or Perish 2007. Available at: http://www.harzing.com/pop.htm. Accessed June 5, 2013.
14. Ashrafian H, Rao C, Darzi A. Benchmarking in surgical research. Lancet 2009;374:1045–7.
15. Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca. Available at: http://www.anvur.org/?q=en. Accessed June 5, 2013.
16. Costas R, Bordons M. The h-index: advantages, limitations and its relation with other bibliometric indicators at the micro level. J Informetr 2007;1:193–203.
17. Martin BR. The use of multiple indicators in the assessment of basic research. Scientometrics 1996;36:343–62.
18. Mozilla. Available at: https://www.mat.unical.it/ianni/wiki/ScholarHIndexCalculator. Accessed January 10, 2011.