

Measuring Academic Performance for Healthcare Researchers with the H Index: Which Search Tool Should Be Used?

Vanash M. Patel^a Hutan Ashrafian^a Alex Almoudaris^a Jonathan Makanjuola^a
Chiara Bucciarelli-Ducci^b Ara Darzi^a Thanos Athanasiou^a

^aDepartment of Surgery and Cancer, Imperial College London, St. Mary's Hospital, London, and ^bNIHR Cardiovascular Biomedical Research Unit, Bristol Heart Institute, Bristol, UK

Key Words

H index · Benchmarking · Health services research · Bibliographic databases

Abstract

Objectives: To compare H index scores for healthcare researchers returned by Google Scholar, Web of Science and Scopus databases, and to assess whether a researcher's age, country of institutional affiliation and physician status influences calculations. **Subjects and Methods:** One hundred and ninety-five Nobel laureates in Physiology and Medicine from 1901 to 2009 were considered. Year of first and last publications, total publications and citation counts, and the H index for each laureate were calculated from each database. Cronbach's alpha statistics was used to measure the reliability of H index scores between the databases. Laureate characteristic influence on the H index was analysed using linear regression. **Results:** There was no concordance between the databases when considering the number of publications and citations count per laureate. The H index was the most reliably calculated bibliometric across the three databases

(Cronbach's alpha = 0.900). All databases returned significantly higher H index scores for younger laureates ($p < 0.0001$). Google Scholar and Web of Science returned significantly higher H index for physician laureates ($p = 0.025$ and $p = 0.029$, respectively). Country of institutional affiliation did not influence the H index in any database. **Conclusion:** The H index appeared to be the most consistently calculated bibliometric between the databases for Nobel laureates in Physiology and Medicine. Researcher-specific characteristics constituted an important component of objective research assessment. The findings of this study call to question the choice of current and future academic performance databases.

Copyright © 2012 S. Karger AG, Basel

Introduction

Research performance has traditionally been evaluated through bibliometrics that included the total number of publications and the total number of citations [1]. Total paper counts do not reflect research quality and ci-

Table 1. Characteristics and differences between the citation indexing databases [6–8, 10, 11, 17]

	Web of Science	Scopus	Google Scholar
Date of inauguration	since early 1960s, but accessible via Internet in 2004	11/2004	11/2004
Number of journals	10,969	16,500 (>1,200 open access journals)	not revealed (theoretically all electronic resources)
Language	English (plus 45 other languages)	English (plus more than 30 other languages)	English (plus any language)
Subject coverage	science, social science and arts and humanities	science and social science	not revealed
Period covered	1900 to present	1966 to present	not revealed
Updating	weekly	1–2/week	monthly
Developer	Thompson Scientific (US)	Elsevier (The Netherlands)	Google Inc. (US)
Fee-based	yes	yes	no
H index calculation	yes	yes	only using Harzing's Publish or Perish software

tation numbers do not provide an accurate account of research breadth. As a result, there is an accepted need to utilize improved markers of research performance to quantify research excellence with increased precision and objectivity [2]. In 2005, Hirsch [3] proposed the 'H index', which measures the importance, significance, and broad impact of a scientist's cumulative research contributions. A scientist with an index of H has published a number of papers (H), each of which has been cited by others at least H times [3]. The use of the H index to measure research performance is rapidly increasing, although the use of the H index has not been previously validated for medical researchers. The H index has been criticized for the technical shortcomings associated with self-citations, field dependency and multiple authorship (these are common to the majority of current bibliometrics) [4, 5]. Specifically, an H index value taken as an isolated measure has been criticized for yielding an unreliable interpretation of the research performance of an individual and may not be fully reflective of those publications with the greatest impact [4, 5].

The strongest indication of its acceptability is that its calculation has been integrated into the citation databases of Web of Science, Scopus and Google Scholar [6–10]. There are differences in the scope of these databases (table 1) [11], which use disparate systems to count citations [1]. The databases have been shown to produce quantitatively and qualitatively different citation counts for general medical research [1].

A number of differences in H index calculation between databases have been reported when evaluating re-

searchers' performance in the fields of computing science and mathematics [12, 13]. It is unclear whether the same H index calculation discrepancy exists when assessing medical researchers. Differences in H index when calculated across different databases may have implications if used in assessing a medical researcher's performance or for decisions on academic promotion.

The aim of this study was to compare the results of the most popular bibliometric databases for a universally accepted cohort of medical scientists in terms of their H index scores.

Subjects and Methods

The cohort of Nobel laureates in Physiology and Medicine (who represent healthcare research excellence and quality) from 1901 to 2009 were chosen to investigate whether or not the H index of medical scientists varies between Web of Science, Scopus and Google Scholar. The same cohort was used to determine whether a scientist's age, country of institutional affiliation at the time of the award and whether the laureate was a physician or not (physician status) had an influence on the calculation of the H index by each database.

A list of all the Nobel laureates in Physiology and Medicine (1901–2009) was obtained by searching the official website of the Nobel Prize (www.nobelprize.org). These individuals were selected because not only are they high scientific achievers in medical research, but also they are representative of global medical science in terms of age, country of institutional affiliation at the time of the award, and physician status. The full names, date of birth, date of bereavement, country of institutional affiliation at the time of the award, and physician status for each laureate were extracted by reading their biography from the official website. These spe-

cific parameters were chosen to evaluate any influence such demographics may have upon the determination of bibliometric indices. For example, it is known that a disproportionately high number of laureates originate from the United States and Europe; however, it is not known whether this factor predicts higher bibliometric scores. The same rationale was applied to the other demographics.

Two authors (V.P., H.A.) calculated the year of first publication, year of last publication, total number of publications, total number of citations and the H index for each laureate from Google Scholar, Web of Science, and Scopus. Publications after 31st December 2009 were excluded, and all data were collected within 7 days from all three citation databases. The data were cross-checked independently by another author (J.M.) to confirm accuracy of data collection. For laureates that had deceased before 31st December 2009, publications after the date of bereavement were excluded. The specific methodology for bibliometric extraction from each database is outlined below.

Google Scholar

Harzing's [8] Publish or Perish software, which analyses raw citations from Google Scholar, was used to calculate a series of bibliometrics. The quoted initial and surname were inputted in the Author Impact Analysis field to generate a list of publications authored by a specific laureate. The search was not restricted to any specialty. Publications from the list that were not authored by the specific laureate were deselected. The bibliometrics were extracted from the results field.

Web of Science

The surname and initial were inputted to search for publications authored by a specific laureate. The Distinct Author Set feature uses citation data to create sets of articles likely written by the same person. This feature was used as a tool to focus the search to compile a list of all the publications for each laureate. The bibliometrics were extracted by creating a citation report from this list.

Scopus

The surname and initials or first name were inputted to search for publications authored by a specific laureate. The Scopus Author Identifier uses an algorithm that matches author names based on their affiliation, address, subject area and source title, dates of publication, citations and co-authors. This identifier was used as a tool to focus the search to compile a list of all the publications for each laureate. The bibliometrics were extracted by viewing the citation overview of this list.

Statistical Analysis

One-way analysis of variance (ANOVA) was used to assess differences of bibliometrics between the databases. Cronbach's alpha statistic was used to measure the reliability of bibliometrics between the databases. Multivariate linear regression analysis was used to explore whether the H index from each database was influenced by a scientist's age, country of institutional affiliation at the time of the award, and physician status. Data were analysed with the use of SPSS for Windows (Rel. 18.0.0. 2009, Chicago, Ill., USA, SPSS Inc.).

Results

There were 101 Nobel Prizes in Physiology and Medicine awarded to 195 laureates between 1901 and 2009, with a median age of 56 years (32–87) at the time of the award; 10 laureates were female and 125 laureates were deceased. Of the 101 Nobel Prizes, 37 (36.6%) were given to one laureate only, 31 (30.7%) were shared by two laureates and 32 (31.7%) were shared between three laureates. At the time of the award 92 (47.2%) laureates were affiliated to North American countries, 92 (47.2%) laureates were affiliated to European countries, and 11 (5.6%) laureates were affiliated to other countries (table 2). Ninety-nine (50.8%) laureates were non-physicians and 96 (49.2%) were physicians. Three laureates (prize winning years 1948, 1986 and 2007) were excluded from the analysis because accurate data could not be retrieved due to commonality of names. Data were not available for 68 laureates in Web of Science (prize winning years ranging from 1901 to 1976) and 29 laureates in Scopus (prize winning years ranging from 1901 to 1985). Table 3 gives examples to show the variation of bibliometric data extracted from the three databases.

The median year of first publication was significantly different between the three databases with the lowest year in Google Scholar (1933; 1843–1989) compared to Scopus (1957; 1880–1996) and Web of Science (1970; 1968–1991) ($p < 0.0001$ for all comparisons). The median year of last publication was not significantly different in Google Scholar (1997; 1903–2009) and Scopus (2000; 1880–2009) ($p = 0.408$), but was significantly higher in Web of Science (2008; 1970–2009) ($p < 0.0001$ for both comparisons). The median total number of publications was significantly different between the three databases with the highest in Google Scholar (236; 3–1,000) compared to Web of Science (109; 1–1,255) and Scopus (67; 1–992) ($p < 0.0001$ for all comparisons). The median total number of citations was significantly different between the three databases with the highest in Web of Science (10,579; 2–95,460) compared to Google Scholar (6,521; 4–97,988) and Scopus (991; 0–41,470) ($p = 0.029$ Web of Science vs. Google Scholar, $p < 0.0001$ for other comparisons). The median H index was not significantly different between Google Scholar (35; 1–166) and Web of Science (43; 1–161) ($p = 0.066$), but was significantly lower in Scopus (13; 0–111) ($p < 0.0001$ for both comparisons).

The H index was the most reliably calculated bibliometric across the three databases (Cronbach's alpha = 0.90) (table 4). This reliability was greater between Web

Table 2. Country of affiliation at time of award

	Laureates (n = 195)
North America	
USA	90
Canada	2
Europe	
UK	31
Germany	16
France	10
Sweden	8
Switzerland	6
Austria	5
Denmark	5
Belgium	4
Italy	2
The Netherlands	2
Hungary	1
Portugal	1
Spain	1
Others	
Australia	6
Russia	2
Argentina	1
Japan	1
South Africa	1

Table 3. Examples of bibliometric outcomes of Nobel laureates in Physiology and Medicine

	James Watson	Francis Crick	Maurice Wilkins
Nobel Prize year	1962	1962	1962
Birth year	1928	1916	1916
Country of affiliation at time of award	USA	UK	UK
Physician status	non-physician	non-physician	non-physician
Google Scholar			
Year of first publication	1948	1945	1934
Year of last publication	2009	2004	2003
Total number of publications	984	180	91
Total number of citations	23,551	10,180	2,247
H index	57	32	23
Web of Science			
Year of first publication	1972	1970	1970
Year of last publication	2008	2007	1995
Total number of publications	161	25	12
Total number of citations	5,567	3,336	1,091
H index	36	15	8
Scopus			
Year of first publication	1969	1950	1948
Year of last publication	2008	2007	1995
Total number of publications	21	87	38
Total number of citations	994	6,562	244
H index	6	33	8

Table 4. Reliability of bibliometrics from Google Scholar, Web of Science and Scopus

	Google Scholar	Web of Science	Scopus	Cronbach's alpha
Total Nobel laureates (n = 195)				
Data available	192	127	166	
Data not retrievable/available	3	68	29	
Bibliometrics, median (range)				
Year of first publication	1933 (1843–1989)	1970 (1968–1991)	1957 (1880–1996)	0.670
Year of last publication	1997 (1903–2009)	2008 (1970–2009)	2000 (1880–2009)	0.832
Total number of publications	236 (3–1,260)	109 (1–1,255)	67 (1–992)	0.759
Total number of citations	6,521 (4–97,988)	10,579 (2–95,460)	991 (0–41,470)	0.789
H index	35 (1–161)	43 (1–166)	13 (0–111)	0.900

of Science and Scopus (Cronbach's alpha = 0.91), than Google Scholar and Scopus (Cronbach's alpha = 0.85) or Google Scholar and Web of Science (Cronbach's alpha = 0.82).

Univariate regression analysis demonstrated that younger laureates were more likely to have significantly

higher H indexes returned by each of the three databases ($p < 0.0001$) (table 5). The H index from Google Scholar was higher in laureates from North America in comparison to laureates from European and other countries, respectively ($p < 0.0001$). Physicians had a lower H index than non-physicians in Scopus ($p = 0.018$).

Table 5. Regression analyses of laureate characteristics and H index values from Google Scholar, Web of Science and Scopus

Characteristic	Univariate analysis ^a	p value	Multivariate analysis ^{a, b}	p value
Birth year				
Google Scholar	0.616 (0.500–0.731)	<0.0001	0.561 (0.449–0.672)	<0.0001
Web of Science	0.273 (0.209–0.336)	<0.0001	0.280 (0.218–0.343)	<0.0001
Scopus	0.643 (0.542–0.744)	<0.0001	0.594 (0.497–0.690)	<0.0001
Country of affiliation at time of award				
Google Scholar	–0.006 (–0.009–0.003)	<0.0001	–0.004 (–0.007–0.000)	0.067
Web of Science	–0.001 (–0.004–0.001)	0.352	–0.002 (–0.06–0.001)	0.159
Scopus	–0.003 (–0.007–0.000)	0.053	–0.002 (–0.007–0.003)	0.467
Physician status				
Google Scholar	–0.002 (–0.004–0.001)	0.223	0.003 (0.000–0.007)	0.025
Web of Science	0.001 (–0.001–0.003)	0.521	0.003 (0.000–0.006)	0.029
Scopus	–0.003 (–0.006–0.001)	0.018	0.003 (0.000–0.007)	0.079

^a Values are expressed as unstandardized B coefficients (95% confidence intervals).

^b Variables included birth year, country of affiliation at time of award and physician status.

Multivariate regression model (table 5) showed that the impact of birth year on the H index remained significant. There was no significant influence of country of institutional affiliation at the time of award on the H index in any database. The non-physician bias of Scopus became non-significant, but physicians had a higher H index in Google Scholar and Web of Science ($p = 0.025$ and $p = 0.029$, respectively).

Discussion

The H index was more consistently calculated between the databases than other bibliometrics. However, in general, it is important to consider which database to use for computing the H index for a variety of reasons. As demonstrated by our study, H index scores are influenced by age and physician status. Resultantly, individual scores may be discordant between databases as each covers a different time span and list of journals. The scope of each database must be transparent including algorithms used such that researchers are able to make informed decisions on which source to base their performance metrics on. This study showed that physician's status could constitute an important component of objective research assessment in healthcare.

There appeared to be a paucity of laureates with H index scores greater than 60 before circa 1900 as demonstrated in figure 1. This could be due to several factors: firstly, there could have been fewer publications in circu-

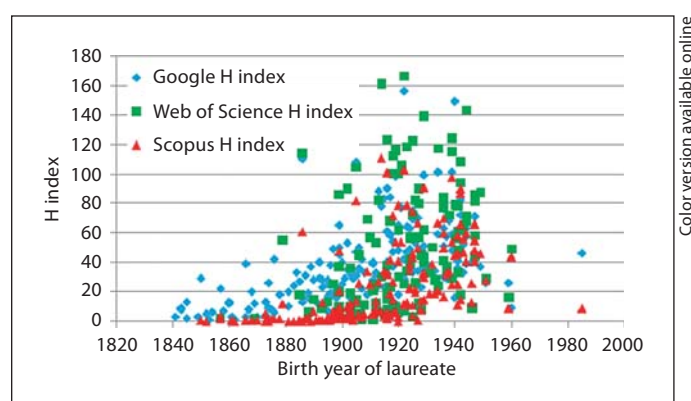


Fig. 1. Scatter plot of birth year of laureates versus their H index.

lation during this period, hence probability of citation. Secondly, with the advent of electronic publishing of manuscripts in the contemporary era, manuscripts reach wider audiences and thus are more likely to be cited. Hence finally, what was represented is the degree to which journals indexed their historic manuscripts.

The use of the H index in Nobel laureates reflects the application of this bibliometric tool as a valuable measure of research excellence and quality. The future use of the H index could include its application in assessment, promotion and support of healthcare scientists.

The strengths of this study included using a large cohort of researchers who had been selected by peer review

for receipt of the Nobel Prize for their achievements in medical research. The broad range of the laureates' birth year allowed the impact of the time period covered by each database on the H index to be investigated. Most of the laureates were equally affiliated at the time of the award to either North American or European countries, which strengthened the evaluation of geographical bias of the databases in relation to the H index. There were similar numbers of physician and non-physician laureates, so the physician bias of the databases in relation to the H index could be assessed. The methodology of data collection was robust given the agreement of two independent authors with the retrieval of the results and furthermore the short information retrieval period occurring within 1 week.

Potential limitations of this study include difficulty in distinguishing between articles belonging to authors with similar names. Web of Science and Scopus use identifiers to group an author's publications together, and Scopus claims to have achieved 99% certainty for 95% of its records [6, 10, 14]. These individual author sets were used to focus our search for a laureate's publications to ensure that data collection was comprehensive. Harzing's Publish or Perish software does not have an author identifier feature, which made data extraction from Google Scholar more difficult and time-consuming. The reliability of Google Scholar is unknown because the coverage and methods of the database are not transparent [15]. This study did not take into account the age at which the Nobel Prize was awarded, and the follow-up

citations after the Nobel Prize award may also have affected final citation and H index results. When considering the significant influence of the laureate's birth year on the H index, it was not possible to determine whether publications of older laureates had a longer time to get cited or whether publications of younger laureates were more likely to get cited because a greater proportion of their work was more accessible. The language bias of native English speakers in relation to the H index was not tested. The results of this study cannot be translated to other scientific disciplines because the value of the H index is discipline-dependent [4]. The overlap of publications and citations between the databases was not considered, so it was not possible to determine the degree to which the H index increases by combining the unique publications and citations from each database [12, 16, 17].

Conclusions

Google Scholar and Web of Science returned significantly higher H index scores for laureates in comparison to Scopus. However, the H index was the most reliably calculated bibliometric across the three databases. Researcher-specific characteristics should be considered when calculating the H index from citation databases. Ultimately, this will lead to more objective, fair and transparent assessment of healthcare research performance for individuals, institutions, and countries.

References

- 1 Kulkarni AV, Aziz B, Shams I, Busse JW: Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *JAMA* 2009;302:1092–1096.
- 2 Ashrafian H, Rao C, Darzi A, Athanasiou T: Benchmarking in surgical research. *Lancet* 2009;374:1045–1047.
- 3 Hirsch JE: An index to quantify an individual's scientific research output. *Proc Natl Acad Sci USA* 2005;102:16569–16572.
- 4 Bornmann L, Daniel H: What do we know about the H index? *J Am Soc Information Sci Technol* 2007;58:1381–1385.
- 5 Bornmann L, Mutz R, Daniel H: The H index research output measurement: two approaches to enhance its accuracy. *J Informetrics* 2010;4:407–414.
- 6 Elsevier: Scopus website, 2004.
- 7 Google: Google Scholar Beta website, 2004.
- 8 Harzing AW: Publish or Perish, 2007, available from <http://www.harzing.com/pop.htm>.
- 9 Kulasegarah J, Fenton JE: Comparison of the H index with standard bibliometric indicators to rank influential otolaryngologists in Europe and North America. *Eur Arch Otorhinolaryngol* 2010;267:455–458.
- 10 Thomson-Reuters: ISI Web of Knowledge website, 2004.
- 11 Patel VM, Ashrafian H, Ahmed K, Arora S, Jiwan S, Nicholson J, Darzi A, Athanasiou T: How has healthcare research performance been assessed? – A systematic review. *J R Soc Med* 2011;104:251–261.
- 12 Bar-Ilan J: Which h-index? – A comparison of WoS, Scopus and Google Scholar. *Scientometrics* 2008;74:257–271.
- 13 Meho LI, Rogers Y: Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science. *J Am Soc Information Sci Technol* 2008;59:1711–1726.
- 14 Qiu J: Scientific publishing: identity crisis. *Nature* 2008;451:766–767.
- 15 Mingers J, Lipitakis E: Counting the citations: a comparison of web of science and Google Scholar in the field of business and management. *Scientometrics* 2010;85:613–625.
- 16 Bakalbas N, Bauer K, Glover J, Wang L: Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomed Digit Libr* 2006;3:7.
- 17 Falagas ME, Pitsouni EI, Malietzis GA, Pappas G: Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* 2008;22:338–342.