The h-index outperforms other bibliometrics in the assessment of research performance in general surgery: A province-wide study

Bharat Sharma, MD,[a] Sylvain Boet, MD, MEd,[b,e] Teodor Grantcharov, MD, PhD,[a] Eunkyung Shin, BSc,[a] Nicholas J. Barrowman, PhD,[c] *and* **M. Dylan Bould, MB ChB, MEd,[d,e]**
Toronto and Ottawa, Ontario, Canada

Background. *The h-index is used as an objective measure of research impact. Its validity, however, is not known in the context of general surgery and comparisons with other bibliometric indices are lacking. We sought to evaluate the h-index as a reliable and valid measure of research performance in general surgery across 6 universities in the province of Ontario, Canada.*
Methods. *Bibliometric indices for 219 faculty members in general surgery were calculated using the Scopus and Web of Science online databases. We investigated agreement between the databases. A 2-way analysis of variance was used to compare the h-index of surgeons grouped by institutional affiliation and academic rank and to identify the relative impact of these factors on different bibliometric indices.*
Results. *The agreement on h-indices between the Scopus and Web of Science was problematic. The h-index was associated more strongly with academic rank (academic rank accounted for 33.3% of researcher's h-index) than of the number of publications (12.5%) or the number of citations per author (10.2%). The number of citations per paper was not associated with academic rank. The institutional affiliation affected bibliometric indices to a similar degree to academic rank.*
Conclusion. *Our data suggest better construct validity for the h-index than for other bibliometrics, although the agreement of h-index values between databases can be problematic for some researchers. The use of the h-index as a criterion-based assessment across different universities is problematic and that it should be used as a normative assessment tool, with comparisons with a specified population of interest. (Surgery 2013;153:493-501.)*

From the Department of Surgery,[a] St Michael's Hospital, University of Toronto, Toronto; the Department of Anesthesiology,[b] The Ottawa Hospital, University of Ottawa, Ottawa; the Departments of Pediatrics[c] and Anesthesiology,[d] The Children's Hospital of Eastern Ontario; and The Academy for Innovation in Medical Education,[e] The University of Ottawa, Ontario, Canada

THE H-INDEX was introduced in 2005 as a novel marker of research performance by Jorge E. Hirsch.[1] It is a citation-based metric that rewards sustained efforts of research productivity and impact. Metrics such as the h-index are used increasingly to guide decisions relating to academic promotion and research funding, which begs the questions: What is the evidence that research performance can be assessed appropriately with bibliometrics? And, is any 1 metric more appropriate than another to assess research impact?

A strength of the h-index is that it evaluates 2 surrogate measures of research quantity (evaluated by the number of publications) and quality (evaluated by the number of citations of publications) in a single number. A scientist has an h-index of *h* if *h* of his/her papers have at least *h* citations each and his/her other papers have less than or equal to *h* citations each. In other words, a scientist with a h-index of *h* has published at least *h* papers, each of which has been cited at least *h* times. For instance, 20 publications each cited once gives an h-index of 1, and 1 publication cited 20 times also gives an h-index of 1 (Table I). For instance, a researcher with 8 publications each cited 8 times

Funded by the Allan Waters Family Patient Simulation Centre, St. Michael's Hospital, University of Toronto, Toronto, Ontario.
Accepted for publication September 6, 2012.

Reprint requests: M. Dylan Bould, MB ChB, MEd, Department of Anesthesiology, The Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, Ontario, Canada. E-mail: dbould@cheo.on.ca.

0039-6060/$ - see front matter

http://dx.doi.org/10.1016/j.surg.2012.09.006

Table I. Examples of different bibliometric indices for 5 hypothetical scientists

Rank of publication by citations	*Citations per publication for 5 scientists*				
	A	*B*	*C*	*D*	*E*
1st	10	10	10	200	200
2nd	10	10	10	150	150
3rd	10	10	10	100	100
4th	10	10	10	10	50
5th	10	10	10	10	27
6th	10	10	10	10	20
7th	10	10	10	10	10
8th	10	10	—	10	6
9th	10	10	—	10	1
10th	10	10	—	10	0
11th	—	10	—	—	0
12th	—	10	—	—	0
N_p	10	12	7	10	12
N_c	100	120	70	520	564
N_c/N_p	10	10	10	52	47
h-index	10	10	7	10	7

Scientist A has 10 publications, each of which has been cited 10 times so their h-index is 10. Scientist B has 12 publications, each cited 10 times so his/her h-index is also 10, because the 11th publication is cited <11 times. For his/her h-index to increase to 11. He/she would need to have 11 publications each cited ≥11 times. Therefore, increases in the h-index require increases in both productivity and impact. Scientist C has an h-index of 7, even though all of his/her publications have also been cited 10 times, because the h-index cannot be higher than the total number of publications. Scientist D has published only 10 articles and has an h-index of 10 despite 3 very highly cited publications; therefore, the h-index is insensitive to a few "blockbuster" publications. Scientist E has an h-index of 7 because the top 7 articles have been cited ≥7 times and the 8th has been cited <7 times. The h-index is insensitive to a long "tail" of infrequently cited publications.

N_c, Total number of citations; N_p, total number of publications; N_c/N_p, citations per publication.

has an h-index of 8. If he/she publish a further 9th paper that has not yet been cited, then his/her h-index will remain at 8 until all 9 papers have been cited ≥9 times each. It would not matter if 1 of those 8 papers had been cited 100 times; to have an h-index of 100, the researcher must have 100 publications, all cited ≥100 times. As such, it is a fairly conservative metric. The h-index is, therefore, little affected by researchers who publish a high volume of low-impact papers or those who only have a few, high-impact publications.

First used in the basic sciences, the h-index has garnered interest recently in several health care specialities.[2-4] Previous research in urology and other medical disciplines has supported the construct validity of the h-index as an assessment of academic performance showing that it correlates with academic rank.[5-7] A limitation of this model of construct validity is that promotion may not only be based solely on research productivity and quality, especially in a more clinical faculty, but may also depend on administrative, educational, and clinical achievements. The existing literature has not quantified the association between the h-index and academic rank to clarify the usefulness of the model of validity,[5] and few studies have compared the h-index with other metrics.[6,7] The validity of the h-index as an assessment of individual research impact within the context of general surgery is also unknown.

Ball[8] noted that direct comparisons of h-index across disciplines as diverse as physics, chemistry, and computer science cannot be made owing to inherent differences in patterns of literature citations. Less is known about comparisons between different medical specialties; a recent study by Hedley-Whyte et al[9] demonstrated that chairpersons in pathology have greater h-indices than chairpersons in other specialties, which may either be owing to increased productivity or to different patterns of citation in different specialties. Hedley-Whyte et al[9] also found that h-indices varied by institution in the United States, and compared schools that were ranked by consolidating several previously published rating systems (Shanghai Jiao Tong University Academic Ranking of World Universities, *U.S. News and World Report*, and America's Best Graduate Schools, 2010 Edition) which themselves reflect research awards, articles published in *Nature* and *Science*, as well as citations. Schools ranked between the top 1% and 16% of US universities had a greater mean h-index value than schools identified as being ranked between the top 33% and 50%. The influence of institutional affiliation on bibliometric indices in the Canadian context, however, has not been explored previously.

Our study aimed to evaluate the reliability and construct validity of the h-index as an assessment of research performance of basic and clinical scientists in the departments of general surgery across 6 Ontario-based universities including analyses of h-index by academic rank and institutional affiliation. We aimed to compare the h-index with traditional measures of research impact: The number of publications per author, total citation count per author, and citations/paper. Our null hypotheses were (1) that the h-index would not differ significantly by academic rank, and (2) that there would be no difference in h-index according to institutional affiliation.

METHODS

Ethics. All the data that we present are in the public domain. The names of the faculty were taken from public University websites and citation information from Scopus and Web of Science databases, which are also in the public domain. We note that approval from the research ethics boards was not sought for any of the other h-index papers encountered in our review of the literature. Because it is widespread practice not to seek consent to report information that is in the public domain, we did not do so for this study, nor did we seek approval from our research ethics board.

Data collection. We elected to study universities in the province of Ontario, because it is the most populous province in Canada (2010 population 13,210,667[10]) and has 6 medically affiliated universities. We included all academic staff with appointments of lecturer, assistant professor, associate professor, and full professor in the University Departments of General Surgery in the province of Ontario, Canada: University of Ottawa, Queen's University, University of Toronto, Northern Ontario School of Medicine, University of Western Ontario, and McMaster University. The faculty and their associated academic ranks were identified using the departmental websites at each university. Academic staff with appointments of adjunct professors, professor emeritus, or cross-appointments from other departments were excluded ($n = 23$). We used 2 competing proprietary online bibliographic databases for data collection: Thomson Reuters' Web of Science, which has a more extensive historic coverage of publications and their citations, and Elsevier's SciVerse Scopus, which covers a greater number of journals but only since 1996.

We used the *author finder* function of Web of Science to identify faculty members. Subject categories were limited to life sciences and biomedicine and to multidisciplinary science and technology. Relevant publications were identified, and the *create citation report* function was used to determine a Web of Science h-index, the total number of publications found, the total number of citations found, and the mean citations per publication.

We then used the *author search* function of Scopus and extended the documents linked to relevant author matches. Output was sorted by the number of citations per document, relevant publications were identified, and the *citation tracker* function was used to provide a Scopus h-index. Using the *overview options*, we also calculated a Scopus h-index without self-citation by any authors (this option is not available for Web of Science).

If neither of the search engines revealed a relevant author, then it was assumed that the h-index and other bibliometric indices were 0. Bibliometric indices for Web of Science and Scopus were calculated on the same day for each faculty member. Once all the data were collected, a second investigator checked a random selection of 25% of the data points using the same methodology. The difference between the Web of Science and Scopus h-index were calculated for each faculty member, and data points were also checked by a second investigator if the difference between Web of Science and Scopus h-indices was >1 standard deviation of the mean difference.

Descriptive statistics (median, interquartile range, range) were calculated for the Web of Science h-index, Scopus h-index, the Scopus h-index without self-citations by any authors, the total number of publications, the total number of citations, and the mean citations per publication for each academic rank.

Analysis of reliability. A Bland–Altman plot is considered to be an appropriate statistical technique to compare 2 measurements when neither of them is the gold standard. This analysis describes the "bias" or difference between 2 methods of measurement. In our study, neither Scopus nor Web of Science can be considered as the gold standard to measure citations of publications. Therefore, we assessed the agreement between the measurement of the h-index by Web of Science and Scopus with a Bland–Altman plot, which shows both the mean bias (ie, the average difference of the h-index between Scopus and Web of Science) and 1.96 SD limits of agreement (ie, a measurement of the range of difference in h-index between Web of Science and Scopus accounting for 95% of a normal distribution).[11] On the basis of previously published data from the specialty of anesthesiology, we decided a priori that 95% of researchers should have <3 h-index points difference when calculated by Web of Science compared with Scopus. We used this level of 3 points because this value represents about half of the difference between h-indices by academic rank (assistant professor versus associate professor, or associate professor versus full professor).[5,11,12]

Analysis of validity. According to current theories of assessment in education research, in the absence of a gold standard, construct validity can be investigated by considering a surrogate. In this case, we determined academic rank to be a surrogate measure of research performance. To examine the construct validity of different bibliometric indices, we compared each of the Web of Science–derived h-indices, total number of publications, total number of citations, and citations per paper by (1) the academic rank and (2) the institution using a 2-way analysis of variance for each metric. In cases where the combined effect of institution and academic rank (interaction) was greater than the sum of their individual effects (main effects), we examined graphically the consistency of the effects of each variable over levels of the other variable (simple effects). Inconsistency would cloud the effects of differences between ranks or institutions. Tests of consistency of the simple effects were also performed using a statistical correction to prevent error from multiple comparisons (Sidak's correction). In cases where consistent effects were found, we proceeded with analysis of the main effects of rank and institution. Statistical comparisons of the main effects of rank and institution were performed using a statistical correction to prevent error from multiple comparisons (Tukey's honest significance test). To examine which bibliometric index showed the strongest evidence for validity, we calculated how much of their variability was explained by academic rank and institution. We also made a planned subgroup analysis of University of Toronto data with recent historic data from anesthesia[5] also using a 2-way analysis of variance. Data were analyzed using SPSS 16.0 (Chicago, IL).

RESULTS

We analyzed the h-indices of 219 general surgical faculty members across 6 Ontario universities using Scopus and Web of Science.

Assessment of reliability. The agreement between the Scopus h-index values and the Web of Science h-index values was calculated using the Bland–Altman plot. The average agreement between the Scopus h-index and Web of Science h-index was excellent at 0.1 h-index points (95% confidence interval [CI], −0.2 to 0.4). The range in agreement of h-index as measured by the 2 databases, however, was considerable; 1.96 SD limits of agreement were −4.7 (95% CI, −5.3 to −4.2) for the lower limit to 4.9 (95% CI, 4.3–5.5) for the upper limit (Fig 1).

Assessment of validity: Web of Science. As measured by Web of Science, the h-index showed a difference between the different academic ranks ($P < .001$; Table II), with academic rank accounting for 33.3% of the variation in h-indices. Mean (SD) h-indices were: lecturer 1.0 (1.8); assistant professor 2.9 (4.1); associate professor 7.3 (6.1); and full professor 23.1 (13.6).

H-indices were also different by institution ($P < .001$; Fig 2; Table II), with institutional affiliation accounting for 28.4% of the variation in h-indices. Mean (SD) h-indices by institution were: University of Toronto 12.4 (12.6); University of Ottawa 6.2 (7.2); University of Western Ontario 6.0 (6.2); Queen's University 5.9 (7.2); McMaster University 2.0 (3.4); and Northern Ontario School of Medicine 0.9 (1.4).

There was an interaction between academic rank and institution ($P < .001$); however, the effects of institution over different academic ranks and the effects of academic rank over different institutions were found to be consistent. For this reason, post hoc tests of the main effects of academic rank and institution were judged to be meaningful. Post hoc testing found differences in h-index between all academic ranks (all $P \leq .001$), except between lecturers and assistant professors ($P = .57$). There was also a difference between h-indices at the University of Toronto and all other institutions ($P \leq .01$) and between Northern Ontario School of Medicine and both the University of Ottawa and the University of Western Ontario ($P \leq .05$). No difference was found between the other departments of general surgery from the other universities.

Assessment of validity: Scopus. The Scopus derived h-index showed a difference between the different academic ranks ($P < .001$), accounting for 31.2% of the variation in the h-indices. Similar degrees of variance were explained using the Scopus derived h-indices whether or not self-citations were excluded (Table II). Mean (SD) h-indices calculated using the Scopus database were: lecturer 1.0 (2.0); assistant professor 2.8 (4.3); associate professor 8.0 (7.4); and full professor 23.0 (12.8). Again, there was an interaction between academic rank and institution ($P < .001$); however, the effects of institution over different academic ranks and the effects of academic rank over different institutions were again found to be consistent. For this reason, post hoc tests of the main effects of academic rank and institution were judged to be meaningful. Post-hoc testing found differences in h-index between all academic ranks (all $P \leq .001$), except between lecturers and assistant professors ($P = .64$). There was also a difference between h-indices at the University of Toronto and all other institutions ($P \leq .05$), between Northern Ontario School of Medicine and both the University of Ottawa, and the University of Western Ontario ($P \leq .05$), between University of Ottawa and McMaster ($P = .02$), and between the University of Western Ontario and McMaster ($P = .02$).

Comparison with other bibliometric indices. Both the total number of publications and the total number of citations as captured by Web of Science differed by both academic rank ($P < .001$)

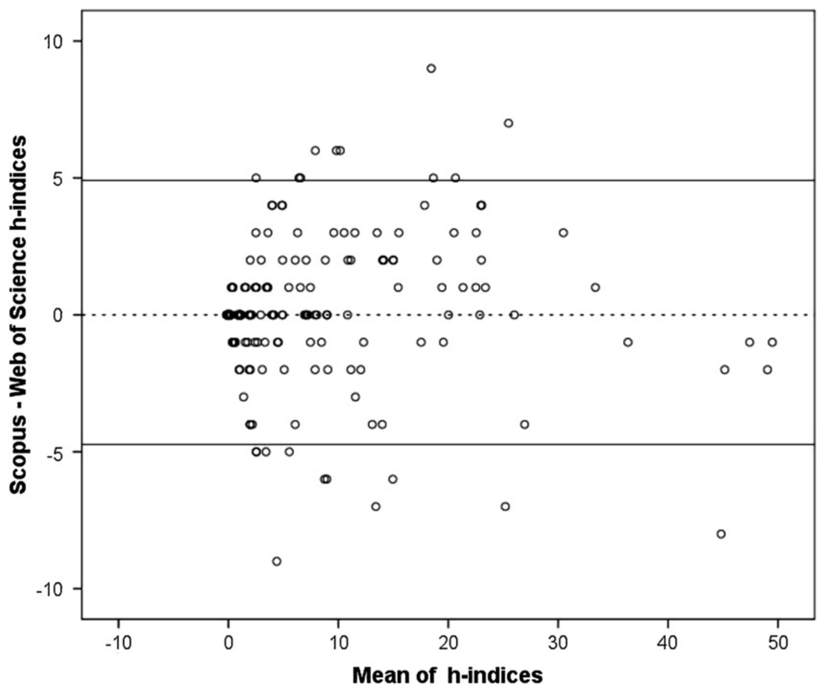

Fig 1. The Bland–Altman plot for Scopus– and Web of Science–derived h-indices. The *dashed line* indicates the mean bias (the mean of the differences between the Scopus– and Web of Science–derived h-indices) and the *solid lines* are +1.96 SD of the mean bias. Because of discreteness, many of the data points were in identical positions. So that all of the data points could be displayed, a small normal random perturbation with mean 0 and SD 0.1 was added to each horizontal coordinate (the mean of the h-indices).

Table II. Degree of variance explained by different bibliometric indices

	Variance owing to	
	Academic rank (%)	*Institution (%)*
h-index (Web of Science)	33.3	28.4
h-index (Scopus)	31.2	26.1
h-index (Scopus, self-citations excluded)	31.6	25.3
Total number of papers	12.5	12.4
Total number of citations	10.2	12.3
Citations per paper	NS	NS

NS, No significant difference in bibliometric index by variable.

and by institution ($P < .001$); however, academic rank explained considerably less variation in these metrics than for the h-index (Table II).The number of citations per paper did not differ significantly by either academic rank ($P = .07$) or institution ($P = .74$; Table II).

Comparison between surgery and anesthesiology. In the pre-planned comparison between the subgroup of the Department of Surgery in Toronto ($n = 84$) and recent historic data from the Department of Anesthesia in Toronto ($n = 284$, data collected May–June 2010),[5] a difference was found between the h-index of anesthesiology and surgery ($P < .001$), as well as there being a difference by academic rank ($P < .001$). Academic rank had much more of an effect on the h-index than specialty accounting for 58.7% of the variation in h-index with the specialty accounting for only 11.1% of the variation (Fig 3).

DISCUSSION

Similar to previous findings by our group in the specialty of anesthesiology,[5] we found that there can be a significant variation in h-indices depending on the type of online search database used. The Web of Science database includes publications from 1990 to present, whereas the Scopus database includes only publications from 1996. The 2 databases also have different coverage of journals, and studies have found that different databases produce quantitatively and qualitatively different citation counts.[13,14] It is probably unlikely that competing commercial databases will become sufficiently aligned to eliminate this problem. A further threat to reliability is how different databases identify individual researchers based on their name and institutional affiliations, which may be problematic if researchers have a common name,

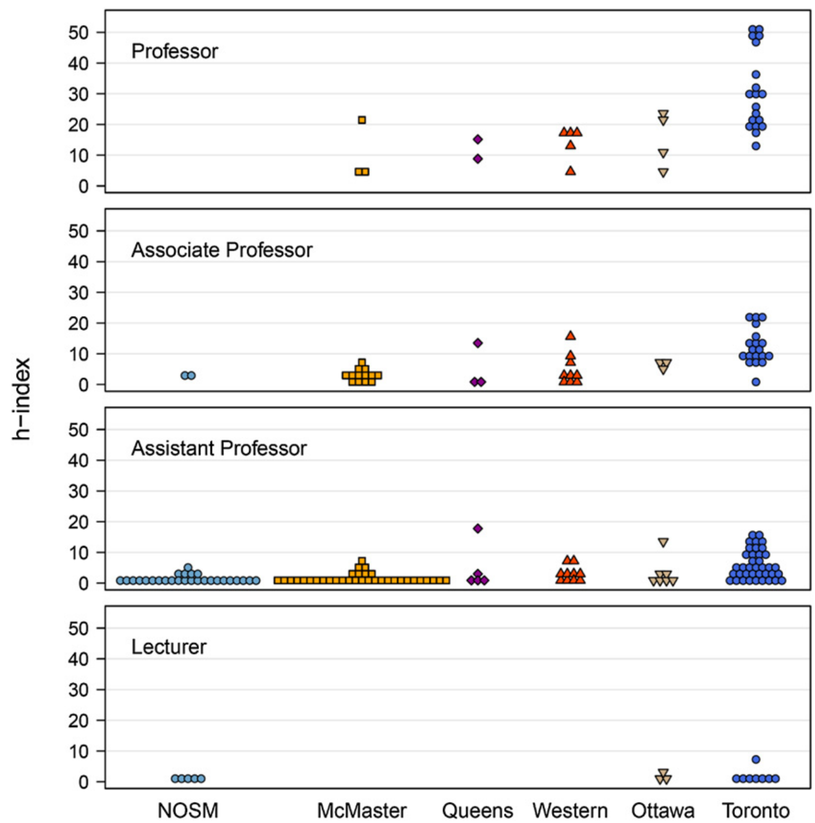

Fig 2. h-Indices by academic rank in departments of surgery in 6 Ontario universities. *Dot plot* of Web of Science h-index for each academic rank. Each dot represents a faculty member and each institution is marked by a different color. (Color version of figure is available online.)

have changed their name, use different initials in different publications, or have changed institutions frequently.[5] A solution to eliminate this threat to reliability would be to have a universally recognized database in which each author validates (ie, confirm his/her authorship) for each publication. This solution is already used in the Système d'Interrogation, de Gestion et d'Analyse des Publications Scientifiques (SIGAPS) system, which relies on PubMed and is mandatory for all physicians in France; this system is currently available only in French.[15] For now, we can just recommend that the readership be aware of these issues, so that they can decide whether they consider h-indices from several databases or to always use one that they feel most comfortable with, and advocate that the search database used to evaluate h-indices should be standardized and transparent.

The h-index differs significantly by both academic rank and institution in academic departments of general surgery across all Universities in the province of Ontario, Canada. Bould et al,[5] Benway et al,[6] and Choi et al[7] have shown previously that the h-index correlates with academic rank in anesthesiology, neurosurgery, and radiation oncology, respectively. These findings have implications for departmental chairs and promotions committees who may find the h-index a useful metric when assessing researchers for recruitment, promotion, or allocation of academic time and funding. Our study builds on this evidence and demonstrates that the h-index also correlates with academic rank in general surgery. What this present study adds is a quantification of the association of the h-index with academic rank and a comparison with other bibliometrics. The academic rank accounts for >33% of the variation in h-index compared with just 12.5% of variation in the number of publications, 10.2% of variation in the total number of citations, and no effect explained in the number of citations per paper. The relatively high degree of variation in h-index explained by academic rank supports its construct validity for the assessment of research productivity in general surgery compared with the other metrics: Publications, citations, and citation per paper. However, the large amount of unexplained variation demonstrates the limitation of this model of construct validity. We acknowledge that there is no "gold standard" for comparison and that

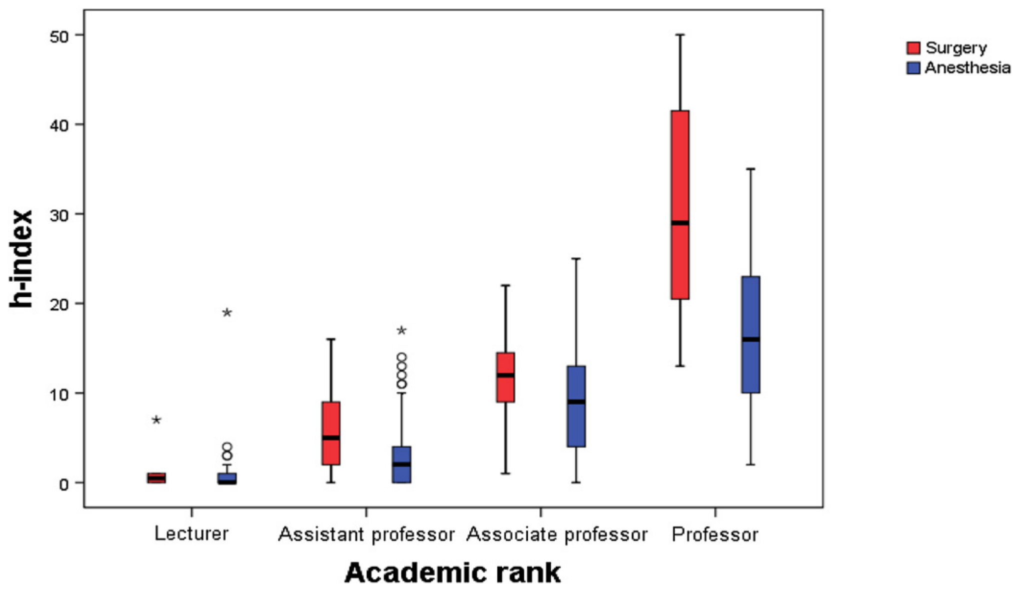

Fig 3. *Box plot* of Web of Science h-index for each academic rank across 2 departments at an Ontario University. The height of the box is the interquartile range (IQR). The *dark horizontal line* is the median. The *lower whisker* extends to the lowest value within 1.5 IQR of the *lower quartile*, and the *upper whisker* extends to the highest value within 1.5 IQR of the *upper quartile*. More extreme values are represented by *circles* and *asterisks*. (Color version of figure is available online.)

academic rank is only a surrogate for research achievements and may also be influenced by alternatives to the research career track, including excellence in administration, education, and even clinical work. Further research investigating that unexplained variation may provide answers to whether we can improve further the assessment of research success beyond current subjective means and existing metrics. Future research is warranted to look at bibliometrics in distinct career paths in surgery.

Our study also explored the impact of institution on bibliometric indices. Hedley-Whyte et al[9] have demonstrated previously that American universities apparently ranked in the top 1% and 16% of universities by several group rankings (Shanghai Jiao Tong University Academic Ranking of World Universities, *U.S. News and World Report*, and America's Best Graduate Schools) have a greater mean h-index than institutions ranked between the top 33% and 50%. Similarly, Ponce and Lozana[4] showed that h-indices vary across institutions and correlate with university funding, faculty size, and publication counts. Our study is the first we know of that compares the relative effect of academic rank and institutional affiliation. Our data suggest that institutional affiliation is as important as academic rank in explaining the variation in h-index values. In the context of Ontario, the institution accounted for almost 30% of the variation in h-indices across general surgery faculty.

Measures of departmental research productivity are important because they may attract future faculty/researchers and impact national/international rankings, all of which have implications for institutional success. In our study, faculty members at a research institution traditionally considered a "research powerhouse" and consistently ranked as a top tier research university in Canada[16] had a mean h-index that was significantly greater than those of other institutions in the province of Ontario, even when accounting for the academic rank. For instance, at Queen's University, the mean h-index for full professors is 12.0; this is comparable to a mean h-index of 12.4 for associate professors at the University of Toronto. Furthermore, as demonstrated previously in anesthesia, much variation also exists within each academic rank level in general surgery. For example, professors at the University of Toronto can have a range of h-index values from 4 to as high as 49. Again, such a high range of h-index value may be explained partly by the alternatives to the research career track for academic promotion.

Our findings suggest that the h-index performs better than the publication count or citation count of a researcher; however, the unexplained variation in h-index values within an academic rank suggest that decisions regarding academic promotion at an institution go beyond this single numerical value, and that factors such as mentorship, teaching ability, and leadership roles and abilities also play a central role.[3] Direct comparisons between h-indices of researchers across institutions should be viewed cautiously; the context in which the h-index is viewed is paramount. When assessing research productivity on the basis of a particular h-index, it is vital to consider what population one wishes to use for comparison.

Using data from a recent study by our group in anesthesiology,[5] we compared the h-indices across the general surgery department and the anesthesia department at a single institution (the University of Toronto) to explore the impact of specialty on the h-index. Conclusions on the effect of specialty on h-index must be extremely guarded from this comparison and should only be used for hypothesis generation, because we only compared 2 of many potential specialties. We found that the effect of specialty on the h-index was much less important than the academic rank, in contrast to our analysis of the effect of institution. Previous studies outside health care suggest that h-indices cannot be compared across disciplines owing to different patterns of publication and citation. For example, some fields such as physics have peak h-indices of around 110, whereas others such as computer sciences have h-index values as high as 70. Radicchi et al[17] have stated previously that the likelihood of any 1 paper being cited varies enormously among scientific fields, and that this likelihood was proportional to the average number of citations within that field. Because the number of citations is a function of the total number of investigators in the field, comparisons of the h-index (which is a direct function of citations) across scientific fields is not valid owing to the widely disparate number of investigators in each field. Differences in journals, citation patterns, and the scientific community result in variations that may make cross-discipline comparisons invalid. Our subanalysis between surgery and anesthesia supports the hypothesis that the h-index is context specific with respect to specialties within medicine, and future research is required to identify how important this factor is across a wide range of specialties. Given the variability of h-index across institutions and disciplines, and problematic reliability depending on which database is used for calculation of the h-index, the use of h-index for academic promotion should be cautious. Recommendations could be made only within institutions, with reference to whether the academic has a research, educational, administrative, or clinical focus, and with a clear and transparent choice on which database is chosen to calculate the h-index.

Some of our study limitations should be noted. Our data are reliant on available website databases and departmental websites to search of faculty members. As such, it is difficult to guarantee that all relevant publications for an author have been included. Errors may occur if an author has a common name, has worked in numerous geographically disparate locations, or has used a combination of initials. We employed safeguards, such as including random sampling checks and reassessment of extreme data points, to avoid such errors and suggest proving unique identification numbers for authors so that they may be identified more accurately across databases. Furthermore, the h-index does not account for the authorship position or the type of publication, that is, a review versus a primary research paper,[18] and this limitation will likely skew comparisons between faculty members. Also, the h-index calculation does not account for when papers are cited, and indeed some publications can be "sleeping beauties" that are perhaps ahead of the field but will later be cited many times. Our study sheds some light on the value of the h-index as we move toward greater research productivity and accountability. Although research productivity may be better captured by the h-index than some other metrics, any single numeric value still overlooks other important academic pursuits, such as teaching, mentorship, leadership, and patient care, and much work remains to be done on how best to assess these other vital academic roles.

Finally, current assessment theory suggests that a process of "validation" can never be fully complete but that as much evidence as possible must be sought to support the use of any form of assessment in a particular context.[19]

REFERENCES

1. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005;102: 16569-72.
2. Lee J, Kraus KL, Couldwell WT. Use of the h index in neurosurgery. Clinical article. J Neurosurg 2009;111:387-92.
3. Fuller CD, Choi M, Thomas CR. Biliometric analysis of radiation oncology departmental scholarly publication productivity at domestic residency training institutions. Am Coll Radiol 2009;6:112-8.
4. Ponce FA, Lozano AM. Academic impact and rankings of American and Canadian neurosurgical departments as assessed using the h-index. J Neurosurg 2010;113:447-57.
5. Bould MD, Boet S, Sharma B, Shin E, Barrowman NJ, Grantcharov T. h-indices in a university department of anaesthesia: an evaluation of their feasibility, reliability, and validity as an assessment of academic performance. Br J Anaesth 2011;106:325-30.
6. Benway BM, Kalidas P, Cabello JM, Bhayani SB. Does citation analysis reveal association between h-index and academic rank in urology? Urology 2009;74:30-3.

7. Choi M, Fuller CD, Thomas CR Jr. Estimation of citation-based scholarly activity among radiation oncology faculty at domestic residency-training institutions: 1996-2007. Int J Radiat Oncol Biol Phys 2009;74:172-8.
8. Ball P. Achievement index climbs the ranks. Nature 2007; 448:737.
9. Hedley-Whyte J, Milamed DR, Hoaglin DC. Chairpersons of pathology in the United States: limited benchmarks for publications. Am J Clin Pathol 2010;134:185-92.
10. Statistics Canada. Quarterly demographic estimates [updated 2011 Jul 5]. Available from http://www.statcan.gc.ca/daily-quotidien/100929/t100929b2-eng.htm.
11. Mantha S, Roizen M, Fleisher L, Thisted R, Foss J. Comparing methods of clinical measurement: reporting standards for Bland and Altman analysis. Anesth Analg 2000;90:593.
12. Bland MJ, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;327:307-10.
13. Falagas M, Pitsouni E, Malietzis G, Pappas G. Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. FASEB J 2008;22:338.
14. Kulkarni A, Aziz B, Shams I, Busse J. Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. JAMA 2009; 302:1092.
15. Devos P, Lefranc H, Dufresne E, Beuscart R. From bibliometric analysis to research policy: the use of SIGAPS in Lille University Hospital. Stud Health Technol Inform 2006;124: 543-8.
16. RE$EARCH Infosource Inc. Top 50 Canadian Research Universities List [updated 2010; cited 2011 Apr 20]. Available from: http://www.researchinfosource.com/media/2010Top50List.pdf.
17. Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: toward an objective measure of scientific impact. Proc Natl Acad Sci U S A 2008;105:17268-72.
18. Callaham M, Wears R, Weber E. Journal prestige, publication bias, and other characteristics associated with citation of published studies in peer-reviewed journals. JAMA 2002;287:2847-50.
19. Kane MT. Current concerns in validity theory. Journal of Educational Measurement 2001;38:319-42.