Use of the $h$ index in neurosurgery

Clinical article

*Janet Lee, M.D., Kristin L. Kraus, M.Sc., and William T. Couldwell, M.D., Ph.D.

Department of Neurosurgery, University of Utah, Salt Lake City, Utah

Object. Assessing academic productivity through simple quantification may overlook key information, and the use of statistical enumeration of academic output is growing. The $h$ index, which incorporates both the total number of publications and the citations of those publications, has been recently proposed as an objective measure of academic productivity. The authors used several tools to calculate the $h$ index for academic neurosurgeons to provide a basis for evaluating publishing by physicians.

Methods. The $h$ index of randomly selected academic neurosurgeons from a sample of one-third of the academic programs in the US was calculated using data from Google Scholar and from the Scopus database. The mean $h$ index for each academic rank was determined. The $h$ indices were also correlated with various other factors (such as time spent practicing neurosurgery, authorship position) to identify how these factors influenced the $h$ index. The $h$ indices were then compared with other citation statistics to evaluate the robustness of this metric. Finally, $h$ indices were also calculated for a sampling of physicians in other medical specialties for comparison.

Results. As expected, the $h$ index increased with academic rank and there was a statistically significant difference between each rank. A weighting based on position of authorship did not affect $h$ indices. The $h$ index was positively correlated with time since American Board of Neurological Surgery certification, and it was also correlated with other citation metrics. A comparison among medical specialties supports the assertion that $h$ index values may not be comparable between fields, even closely related specialties.

Conclusions. The $h$ index appears to be a robust statistic for comparing academic output of neurosurgeons. Within the field of academic neurosurgery, clear differences of $h$ indices between academic ranks exist. On average, an increase of the $h$ index by 5 appears to correspond to the next highest academic rank, with the exception of chairperson. The $h$ index can be used as a tool, along with other evaluations, to evaluate an individual's productivity in the academic advancement process within the field of neurosurgery but should not be used for comparisons across medical specialties. (DOI: 10.3171/2008.10.JNS08978)

Key Words • bibliometrics • citations • $h$ index • impact

Abbreviations used in this paper: AWCR = age-weighted citation rate; df = degrees of freedom.

* Dr. Lee and Ms. Kraus contributed equally to this study.
vide more balanced methods of quantifying the publication record. These include the $h$ index, 10 the $g$ index, 3 and the AWCR. 14 Among these, the citation metric that has been most actively debated, studied, and adopted is the $h$ index. The $h$ index was introduced in 2005 to objectively quantify an individual’s scientific research output. According to Hirsch, 15 the $h$ index combines the number of publications and the number of citations in a single number. The value $h$ is the number of papers published by an author that have received at least $h$ citations. Thus, an author who has published 10 papers that have each been cited at least 10 times will have an $h$ index of 10. Hirsch maintains that the $h$ index provides an unbiased assessment 10 and is also the best metric for predicting future output, 9 although others have questioned its value. 15, 16, 20, 21 The identification of potential limitations in the $h$ index, however, has not limited the broad application of this metric across disciplines.

The $h$ index has only begun to be reported for authors in the medical profession, 22, 23 but can be readily calculated using such tools as the Thomson Institute of Scientific Information Web of Science database, Scopus, and Google Scholar, and thus may come into more frequent use. We proposed to calculate an approximation of $h$ index values for faculty in neurosurgery at various points in their career to provide some general guidelines from which to consider the use of this statistic in publishing by physicians, and also for purposes of providing relative comparison of individuals at differing academic ranks for use in academic retention and promotion. We also calculated comparative $h$ index values for selected physicians in other specialties.

Methods

Sample Identification

A list of neurosurgery training programs was generated from the SFmatch.org website in March 2008. Thirty programs were selected using a random number generator. Within each program, faculty members were categorized by academic tenure—assistant professor, associate professor, professor, or department or division director (chairperson)—according to information listed on each program’s website or affiliated medical school website in March and April 2008. In cases in which the faculty rank could not be determined from the website, the department chairperson was contacted individually for confirmation of faculty rank. Research faculty members and non–tenure-track clinical faculty were excluded. In addition, individuals who had not previously completed a neurosurgery residency or did not possess an American Board of Neurological Surgery certificate (or were not Board eligible, that is, neurointensivists and neurointerventionalists) were also excluded. The remaining faculty members were then listed alphabetically within their tenure rank.

For each of the 30 randomly selected programs, one member was identified from each tenure group using a random number generator. In cases in which an institution lacked a faculty member of a particular rank, this field was omitted from the calculation. When a chairperson at a program was also the only professor or associate professor in the program, they were included in both rank categories; otherwise, the chairperson was excluded from the professor or associate professor list during the calculations.

For comparison with other medical specialties, the editorial boards of two leading journals in each specialty were reviewed. The $h$ index values for the first 5 physicians on the editorial advisory board (alphabetically by surname) of each journal were calculated using the methods described below. Editors-in-chief and associate editors were not used. From these data, the mean $h$ index for each journal and the specialty as a whole was calculated.

Calculation of $h$ Index Values Using Google Scholar

Because of its more widespread availability and inclusivity (longer time period), Google Scholar was selected as the primary database. Google Scholar was searched using the software program Harzing’s Publish or Perish (http://www.harzing.com/pop.htm). The program analyzes raw citations from Google Scholar and calculates the $h$ index and other statistics from the data. The surname and first initial of each author was used for the search. The searches were limited to the “Biology, Life Sciences, Environmental Sciences” and “Medicine, Pharmacology, Veterinary Sciences” fields. After the query results were reported, each reference was reviewed to ensure that it was correctly included with 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 was automatically recalculated.

Calculation of $h$ Index Values Using Scopus

Because we recognize that different databases may provide different pools of references and citations, we selected another database to calculate the $h$ index for comparison. The Scopus abstract and citation database is a subscription database available from Elsevier (www.scopus.com). The database cross-references citations and references from 15,000 peer-reviewed journals, various conference proceedings, and book series for citations after 1995. The Scopus tool also provides citation statistics for articles and authors listed in the database.

Similar to the search of Google Scholar, the surname and first initial of each author were used to search the Scopus database. In some cases, variations on the search (such as surname and affiliation) were used to identify the correct individual. Once the correct individual was found, the “Details” page was viewed. From the “Details” page, the “Identify Unmatched Authors” button was used to identify other indexed variations of the author’s name. The title of the latest published paper of any approximate matches was viewed to determine whether the approximate match should be joined with the rest of the author’s publications. These were added using the “Group With Author” button. Once all variations of the author’s name listing were grouped, the $h$ index of the author was reported.
Use of the $h$ index in neurosurgery

Statistical Analysis

Power calculations were based on the assumption that this study should be able to detect a 5-point difference in the $h$ index between each academic rank using Google Scholar. The study aimed to achieve an 80% power at a 0.05% level of significance.

The correlation coefficient between the $h$ index calculated from Google Scholar and Scopus was determined. The $h$ scores appeared relatively normally distributed, although they were bounded by a lower limit of zero. Therefore, the mean $h$ indices with 95% CIs were identified for each academic rank. An analysis of variance was performed to identify differences among the groups.

To evaluate several factors that may affect the $h$ index of an individual, additional calculations were completed on related data extracted for a subset of individuals. A weighted calculation of the $h$ index based on relative contribution of authorship was calculated for a random subset of 20 individuals. Five individuals from each academic rank were selected, and their $h$ index was recalculated using a weighting system with full credit for first and last authorship, half credit for second authorship, and a quarter credit for any other contributing authorship.

A $g$ index and AWCR were also calculated for this subset of individuals using the Publish or Perish program. The $g$ index is the largest number such that the top $g$ articles published by an author (ranked in decreasing order by the number of citations of each) received together at least $g^2$ citations. The AWCR is the average number of citations received by papers published by a single author, weighted by the age of the papers. A Pearson correlation coefficient between each of these measures and an $h$ index was calculated.

To evaluate the effect of time spent working within the field on the $h$ index, the time from American Board of Neurological Surgery certification to the present for each chairperson included in the study was calculated. A Pearson correlation coefficient between the calculated time and $h$ index was then calculated.

Results

As expected, the $h$ index calculated from the Google Scholar data correlated with academic rank. The mean $h$ indices within the field of neurosurgery were 5.1 (95% CI 2.1–8.0) for assistant professors, 10.7 (95% CI 7.7–13.7) for associate professors, 16.0 (95% CI 13.0–19.0) for professors, and 24.7 (95% CI 21.8–27.6) for chairpersons (Fig. 1). There was a statistically significant difference between the mean $h$ indices of the different academic ranks ($F = 32.1$, $p < 0.0001$).

For comparison, the $h$ index calculated from the Scopus database also correlated with academic rank, but the differences between the ranks were more condensed. The mean $h$ indices within the field of neurosurgery were 4.9 (95% CI 2.5–87.3) for assistant professors, 8.3 (95% CI 5.9–10.7) for associate professors, 10.1 (95% CI 7.7–12.5) for professors, and 14.8 (95% CI 12.5–17.2) for chairpersons. Using the Scopus data, there was a statistically significant difference between the mean $h$ indices of the different academic ranks ($F = 11.96$, $p < 0.0001$).

The $h$ indices calculated from Google Scholar and Scopus data demonstrated a high correlation ($r = 0.77$, $df = 113; p < 0.0001$). Because of the limitations of citation date built into the Scopus database as well as the more general availability of the Google Scholar search engine, further statistical analysis was only performed using the statistics calculated from Google Scholar.

In an effort to determine the impact of relative contributing authorship on $h$ score, weighted $h$ indices were calculated for a subset of individuals. The results with weighted and nonweighted $h$ indices showed that, overall, the weighted $h$ index did not appear to differ from the nonweighted $h$ index ($r = 0.99$, $df = 18$; $p < 0.0011$; Fig. 2).

Similarly, to determine whether highly cited papers would impact the $h$ index, the correlation between the $h$ index and the $g$ index was then calculated within the same subset of individuals, demonstrating a positive correlation ($r = 0.98$, $df = 18$; $p < 0.0001$; Fig. 3). Furthermore, an AWCR was also calculated to evaluate whether the average number of citations adjusted for the age of the contributing papers impacted the $h$ index. The AWCR also demonstrated a positive correlation with the $h$ index ($r = 0.91$, $df = 18$; $p < 0.0001$).
The correlation coefficient between time from American Board of Neurological Surgery certification and $h$ index was 0.68 ($df = 27; p < 0.0001$), demonstrating a positive correlation between the length of time within neurosurgery and the $h$ index.

The mean $h$ index for the medical specialties varied substantially (Fig. 4). Physicians practicing general surgery had the highest mean $h$ index at 32.8, followed by those in urology (32.6), oncology (28.5), cardiology (28.3), neurosurgery (27.2), and orthopedic surgery (15.8).

**Discussion**

In a short period of time, the $h$ index has become a widely recognized measure of quantifying an individual’s research output.\(^1\) In the past, the total number of citations, total number of publications, and citations per publication have also been used to assess productivity and relevance of a body of work. However, the $h$ index incorporates these factors into a single statistic. Because of the versatility of the $h$ index, it has even been used to compare academic and public institutions with one another.\(^17\) Furthermore, it has been shown to have predictive properties of future achievement for the authors.\(^9\) We have completed a sampling of the $h$ index for neurosurgery faculty at 30 academic neurosurgery programs around the US. These values will provide an initial basis for use within neurosurgery by which faculty members can evaluate the impact of their own publications relative to that of their colleagues.

Despite the suggested advantages of the $h$ index in providing an objective evaluation of valued academic productivity, every new statistic must be interpreted within context. Potential drawbacks of the $h$ index may affect the value that can be derived from these calculations. For instance, the context of citation may unduly skew the $h$ index, such as when there is no differentiation between positive and negative citations of data (for example, a highly criticized paper would still be highly cited). In addition, the work of renowned scientists will be disproportionately referenced more often than the work of those authors who are less widely known, even if the content of the citation could be attributed to any number of papers. This so-called “Matthew effect”\(^22\) may bias the $h$ index of those more widely known authors. Conversely, scientists outside of the English-speaking world are at a disadvantage.\(^19\) These are general problems with the $h$ index and other citation metrics that cannot be easily resolved. On the other hand, many concerns about the $h$ index may be overcome or may be less of a concern with the $h$ index than other authors may suggest. These include multiple coauthors receiving equal credit, self-citation to increase one’s own $h$ index,\(^2\) preferential bias to greater number of years of practice, and possibly even sex bias.\(^15\)

Although stricter guidelines for authorship are becoming more widely adopted by peer-reviewed journals,\(^2,7,8,11,12,18\) gratuitous authorship remains a known practice. The accepted convention now recognizes the first and last authors as the primary contributors with secondary contributors typically listed in order of their contribution. Our admittedly small test suggests that, at least with our weighting system, position of authorship does not significantly affect the $h$ index of an individual. Thus, the robustness of the $h$ index appears to overcome this potential drawback.

Self-citation has also been a wide criticism of the $h$ index,\(^4,5\) although this potential problem would also impact simple counting of the total number of citations. Should an individual be aware of which of his or her papers are on the borderline of impacting the $h$ index and then choose to preferentially cite those papers, this may artificially inflate or manipulate the $h$ index.\(^4\) For those authors with a small $h$ index, such as those in the beginning stages of their careers, self-citation could have a large impact, and $h$ index inflation could more easily be accomplished. Engqvist and Frommen\(^4\) studied this dilemma in a group of 40 scientists and found only a minimal impact on the $h$ index after excluding all self-citations. Therefore, over time, the robustness of the $h$ index appears to overcome this potential drawback as well.

As with most measures of productivity, the $h$ index is positively correlated with time spent publishing within a field. We found a positive correlation between the length of time the department chairpersons spent publishing within neurosurgery and the $h$ index. Because the $h$ index may only increase with time, even posthumously, there is a built-in bias for those who have been practicing longer in a given field even if their current productivity is lower than that of their newer counterparts. Nevertheless, the $h$ index...
index is still useful in determining an individual’s contributions over his or her career.

The issue of sex bias has been raised as a potential problem with the use of the $h$ index for evaluation of productivity in at least some academic fields. Kelly and Jennions noted that female researchers in ecology and evolutionary science publish fewer papers and thus have lower $h$ indices than their male counterparts. They noted that the sex differences disappear when controlling for productivity, and therefore the $h$ index could be misinterpreted. Because of the small numbers of women in academic neurosurgery, we are unable to determine whether a sex bias also exists. We will only be able to evaluate this potential bias as more women pursue academic neurosurgery and we can examine their publication records.

One of the identified limitations of the $h$ index is the difficulty in comparing the statistic across fields. Hirsch calculated the $h$ index for physicists and found the high value was 110, whereas a typical value for advancement to tenure (associate professor) would be ~12 and the value for promotion to full tenure would be ~18. Hirsch recognized the potential for differences in publication and citation rates between fields and reported that the $h$ indices of researchers in the life sciences were generally higher. Kelly and Jennions identified a mean $h$ index of 45 for “highly cited” evolutionists and ecologists. reported values between 12 and 39 for editorial board members of the journal Retrovirology. Thus, the $h$ index will likely have the most value when it can be used within a single field or closely related fields.

Comparing $h$ indices among related researchers within the same fields places these indices within the appropriate context. We therefore compared a mean $h$ index for both surgical and nonsurgical medical specialties. Our sample supports the assertion that the $h$ index may not be comparable, even across closely related fields. We found a >15-point difference in the calculated $h$ index between specialties. As might be expected from the size of the field and the variety of interests of the practitioners, general surgery had the highest mean $h$ index at 32.8. This high $h$ index was followed by other widely practiced medical specialties such as oncology (28.5) and cardiology (28.3). Neurosurgery (27.2) had a relatively high mean $h$ index for a specialty of a rather small size. Certainly factors such as the number of clinical practitioners versus researchers and the cross-pollination between certain fields will also affect the mean $h$ index for a specific specialty.

Other measures of academic productivity, namely the g index and the AWCR, attempt to address potential limitations of the $h$ index. Given the strong correlation between the $h$ index and both of these measures, we argue that the $h$ index itself may be superior to either of these indices alone because it incorporates not only numbers of citations but also numbers of papers. The $h$ index has also been more widely studied and appears to overcome many potential drawbacks.

These data represent the first attempt to quantify the $h$ index within the field of neurosurgery and place the $h$ index in context relative to other medical specialties. A cohort of neurosurgeons was randomly selected, thereby minimizing selection bias. We believe this cohort is representative of academic neurosurgery as a whole, although this has not yet been validated in an independent sample. Relative to other specialties, neurosurgery is a young specialty with only ~3000 Board-certified neurosurgeons in the US for the past decade. In comparison, general surgery is one of the oldest specialties, with ~25,000 practicing general surgeons in the US. We provided a simple comparison between different medical specialties by taking a sample of individuals on journal editorial boards within each specialty, but a more rigorous methodology will need to be applied to these specialties to determine a more accurate representation of each medical specialty as a whole.

**Conclusions**

Overall, the robustness of the $h$ index appears to overcome many of its potential drawbacks and provide an accurate assessment of an individual’s contribution to a given field. Within the field of academic neurosurgery, clear differences of $h$ indices between academic ranks exist. On average, an increase of the $h$ index by 5 appears to correspond to the next highest academic rank, with the exception of chairperson. It is likely that this difference results from the limited number of these positions relative to other academic positions and the difficulty of attaining this rank. Although they certainly should be balanced by other evaluations of individual productivity, as a simple and easily determined value, the averages we calculated for each rank may be helpful in evaluating an individual’s productivity in the academic advancement process within the field of neurosurgery.

**Disclaimer**

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

**Acknowledgments**

The authors thank Dr. John Kestle for his assistance in statistical evaluations and Elsevier Scopus for providing a short-term complimentary subscription to their database for review purposes.

**References**


Accepted October 29, 2008.
Please include this information when citing this paper: published online April 24, 2009; DOI: 10.3171/2008.10.JNS08978.
Address correspondence to: William T. Couldwell, M.D., Ph.D., Department of Neurosurgery, University of Utah, 175 North Medical Drive East, Salt Lake City, Utah 84132. email: neuropub@hsc.utah.edu.