Assessment of the NIH-supported relative citation ratio as a measure of research productivity among 1687 academic neurological surgeons

Vamsi Reddy, BS,1 Arjun Gupta,2 Michael D. White, BS,1 Raghav Gupta, BS,2 Prateek Agarwal, AB,3 Arpan V. Prabhu, MD,4 Bryan Lieber, MD,5 Yue-Fang Chang, PhD,1 and Nitin Agarwal, MD1

1Department of Neurological Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; 2Department of Neurological Surgery, Rutgers New Jersey Medical School, Newark, New Jersey; 3Department of Neurological Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; 4Department of Radiation Oncology, University of Arkansas for Medical Sciences, Little Rock, Arkansas; and 5Department of Neurological Surgery, Detroit Medical Center, Wayne State University School of Medicine, Detroit, Michigan

OBJECTIVE Publication metrics such as the Hirsch index (h-index) are often used to evaluate and compare research productivity in academia. The h-index is not a field-normalized statistic and can therefore be dependent on overall rates of publication and citation within specific fields. Thus, a metric that adjusts for this while measuring individual contributions would be preferable. The National Institutes of Health (NIH) has developed a new, field-normalized, article-level metric called the “relative citation ratio” (RCR) that can be used to more accurately compare author productivity between fields. The mean RCR is calculated as the total number of citations per year of a publication divided by the average field-specific citations per year, whereas the weighted RCR is the sum of all article-level RCR scores over an author’s career. The present study was performed to determine how various factors, such as academic rank, career duration, a Doctor of Philosophy (PhD) degree, and sex, impact the RCR to analyze research productivity among academic neurosurgeons.

METHODS A retrospective data analysis was performed using the iCite database. All physician faculty affiliated with Accreditation Council for Graduate Medical Education (ACGME)—accredited neurological surgery programs were eligible for analysis. Sex, career duration, academic rank, additional degrees, total publications, mean RCR, and weighted RCR were collected for each individual. Mean RCR and weighted RCR were compared between variables to assess patterns of analysis by using SAS software version 9.4.

RESULTS A total of 1687 neurosurgery faculty members from 125 institutions were included in the analysis. Advanced academic rank, longer career duration, and PhD acquisition were all associated with increased mean and weighted RCRs. Male sex was associated with having an increased weighted RCR but not an increased mean RCR score. Overall, neurological surgeons were highly productive, with a median RCR of 1.37 (IQR 0.93–1.97) and a median weighted RCR of 28.56 (IQR 7.99–85.65).

CONCLUSIONS The RCR and its derivatives are new metrics that help fill in the gaps of other indices for research output. Here, the authors found that advanced academic rank, longer career duration, and PhD acquisition were all associated with increased mean and weighted RCRs. Male sex was associated with having an increased weighted RCR but not an increased mean RCR score. Overall, neurological surgeons were highly productive, with a median RCR of 1.37 (IQR 0.93–1.97) and a median weighted RCR of 28.56 (IQR 7.99–85.65).

KEYWORDS research productivity; citations; relative citation ratio; neurological surgery; National Institutes of Health
Bibliometric analysis of a scientific author’s research output is a valuable tool that allows for quantification of both productivity and impact. Currently, the most common metric for such analysis has been the Hirsch index (h-index). The h-index is an author-level metric that is defined as follows: “A scientist has index h if h of his or her Np papers have at least h citations each and the other (Np − h) papers have ≤ h citations each.” Although the h-index was widely adopted at the time of its conception, the simplicity of this metric lends itself to several major shortcomings. Its main criticism is that combining the frequency of publication and the frequency of citation into a single metric can lead to misleading characterization of productivity and impact. An interesting example demonstrating this limitation was published by the Institute of Mathematical Statistics: “Think of two scientists, each with 10 papers with 10 citations, but one with an additional 90 papers with 9 citations each; or suppose one has exactly 10 papers of 10 citations and the other exactly 10 papers of 100 each. Would anyone think them equivalent?” For this reason, it has been noted that the h-index may also unfairly disadvantage younger authors when measuring their impact in the field, such that an older author with numerous low-impact publications may have a higher h-index than a newcomer with only a handful of high-impact papers. Additionally, the h-index is not a field-normalized statistic in that the citation potential of a publication is heavily dependent on the size of the audience of the field in which it is published. Thus, the h-index has been criticized as inappropriate for comparisons of authors from different fields and authors with different career lengths.

The National Institutes of Health (NIH) released a new article-level metric in 2015 called the “relative citation ratio” (RCR), which resolves many of the aforementioned limitations of the h-index (https://icite.od.nih.gov/analysis). The RCR for a given publication is calculated as the total number of citations per year of that publication divided by the average field-specific citations per year received by NIH-funded papers in the same field (as determined by a co-citation network). This dynamic field normalization is a defining feature of the RCR and allows for more accurate comparison of author productivity and impact across scientific fields. Author-level derivatives of the RCR, such as the mean RCR and the weighted RCR, are calculated as the average and the sum, respectively, of all the article-level RCR scores earned by publications produced by an individual author. By averaging an author’s article-level RCR scores, the mean RCR allows for more fair and accurate comparison of impact by authors from different age groups without letting the total number of publications put younger authors at a disadvantage. In contrast, the weighted RCR is calculated by adding the article-level RCR scores for each one of an author’s publications, allowing the quantity of publications to influence the score. Thus, weighted RCR is a more appropriate metric when measuring the total number of publications or any other time-dependent factor; however, when measuring how often, on average, an author’s publications are cited within the field, the mean RCR is the more appropriate metric. We classify these two characteristics as “productivity” versus “impact,” respectively. Weighted RCR can be used to measure total productivity over the entire duration of an author’s career, whereas mean RCR will show how impactful the author’s publications are in comparison to NIH-funded papers in the same field.

Currently, there are no benchmark data for RCR scores within the field of neurological surgery. Therefore, in the present study, we conducted an analysis of RCR scores for 1687 academic neurosurgeons in the United States, including the mean and weighted RCRs. We further assessed the impact of sex, career duration, academic rank, and acquisition of a Doctor of Philosophy (PhD) degree on the RCR scores of United States academic neurosurgeons.

Methods

Departmental and Faculty Inclusion Criteria

Academic neurosurgeons (Doctor of Medicine [MD] or Doctor of Osteopathic Medicine [DO]) employed as faculty at Accreditation Council for Graduate Medical Education (ACGME)–accredited neurological surgery programs were included in our analysis. Individual departmental websites for each accredited residency program were accessed in January 2019 (https://apps.acgme.org/ads/Public/Programs/Search). Sex, degrees, and academic rank were determined using physician profiles on departmental websites. Additional information, including residency start year, was obtained through American Association of Neurological Surgeons (AANS) membership database self-entries (https://www.aans.org/en/Trainees/Residency-Directory).

Bibliometric Analysis

Article-level RCR is the total number of citations per year of a publication divided by the average field-specific citations per year received specifically by NIH-funded papers in the same field (as determined by a co-citation network). By definition, an RCR score of 1.0 indicates the median for all NIH-funded publications. The aggregate article-level RCR scores for all of an individual’s publications can be used to derive author-level RCR scores for any particular author. The mean RCR, often referred to simply as an author’s “RCR,” is the statistical average of the RCR scores earned by all publications produced by that author. The weighted RCR, which is more heavily influenced by the number of publications produced by a particular individual, is calculated as the sum of all the RCR scores earned by every publication produced by that author.

Academic neurosurgeons were individually indexed using the NIH iCite website (https://icite.od.nih.gov/analysis), although their nonarticles (i.e., letters to the editor) were excluded from analysis. Total number of publications, mean RCR, and weighted RCR were collected in January 2019, which included PubMed-listed articles from 1995 to 2018.

Statistical Analysis

Mean RCR (hereafter referred to simply as “RCR” in reference to the author-level metric) and weighted RCR were calculated for all academic neurosurgeons and com-
pared by sex, degree, academic rank (assistant, associate, and full professors), and career duration by residency start date (e.g., ≤ 1980, 1981–1990, 1991–2000, > 2000). Since RCR and weighted RCR data were highly skewed, in addition to using actual scores with the Wilcoxon rank-sum test, analyses were also performed using the quartiles of RCR and weighted RCR data. The data herein are primarily presented as median and interquartile range values, with p values representing results from the Wilcoxon rank-sum test. Analyses were done on the entire sample, as well as within each group based on the year in which its residency training started. Analyses were conducted using SAS 9.4 (SAS Institute).

Results

A total of 1687 academic neurosurgeons were included in the data analysis (Table 1). Table 2 provides weighted RCR data and Table 3 provides mean RCR data, both with detailed breakdowns by sex, PhD acquisition, academic ranking, and residency start year. Figure 1 shows an overview of the mean RCR data for all academic neurosurgeons, whereas Fig. 2 illustrates an overview of weighted RCR data for all academic neurosurgeons.

Academic Rank

The majority of academic neurosurgeons in our study were male (n = 1533 [91.03%]), and approximately 15% had a PhD (n = 245). Assistant professor, which includes clinical assistant professor, instructor, and lecturer, was the most populated academic rank at 566 members (41.65%), with associate professors composing approximately one-fourth of the total population (23.25%) and full professors making up the remaining 35.10%. Overall, academic neurosurgeons had high but widely variable RCR scores, with a median RCR of 1.37 (IQR 0.93–1.97) and a median weighted RCR of 28.56 (IQR 7.99–85.65). The median number of total publications produced by an individual neurosurgeon was 21 (IQR 7–52), with a median publications per year rate of 1.67 (IQR 0.92–3.4).

Both RCR and weighted RCR were positively correlated with successive academic rank for our overall sample (p < 0.0001). This trend persisted when the population was stratified by residency start date to control for career duration (data not shown). It is worth noting that the absence of this trend among individuals starting residency training...
prior to 1980 is most likely attributable to a lack of sufficient data for assistant and associate professors in that subgroup (n = 6 and 4, respectively). Full professors were the most productive subgroup included in the study with a median RCR score of 1.56 (IQR 1.16–2.16) and a median weighted RCR of 64.34 (IQR 25.51–152.73).

Sex
Female neurosurgeons had a median RCR score of 1.34 (IQR 0.91–1.97), and males had a median RCR score of 1.37 (IQR 0.93–1.97). Although there was no significant difference in RCRs (p = 0.822) between the sexes, the weighted RCR was significantly higher (p = 0.003) for males than females (medians 30.51, IQR 8.16–88.66; and 19.76, IQR 5.51–49.41, respectively).

PhD Acquisition
Acquisition of a PhD degree showed a significant increase in both RCR and weighted RCR (p = 0.008 and p < 0.0001, respectively). This was clear in the median weighted RCR, which was 50.62 (IQR 16.52–123.10) for individuals with a PhD compared to 26.23 (IQR 6.94–78.43) for individuals without a PhD. Interestingly, upon stratifying by career duration, statistically significant differences between these two groups disappeared in all residency-start-date subgroups, except among individuals who started residency training after the year 2000. This finding occurred for both RCR and weighted RCR (p = 0.004 and p < 0.0001, respectively, in > 2000 subgroup).

Career Duration
Longer career duration, calculated via residency start date, had a significant impact on both RCR and weighted RCR scores (p = 0.031 and p = 0.006, respectively). For both RCR metrics, the subgroup with the longest career duration (residency start date ≤ 1980) had the highest median scores, and the shortest career duration subgroup (residency start date > 2000) had the lowest median scores.

### Table 3. Mean RCR by sex, PhD acquisition, academic ranking, and residency start year

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>No.</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>25th Percentile</th>
<th>75th Percentile</th>
<th>p Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>151</td>
<td>1.88</td>
<td>2.65</td>
<td>1.34</td>
<td>0.91</td>
<td>1.97</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1533</td>
<td>1.76</td>
<td>3.10</td>
<td>1.37</td>
<td>0.93</td>
<td>1.97</td>
<td>0.822</td>
</tr>
<tr>
<td>PhD degree</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1442</td>
<td>1.77</td>
<td>3.27</td>
<td>1.36</td>
<td>0.91</td>
<td>1.96</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>245</td>
<td>1.77</td>
<td>1.21</td>
<td>1.48</td>
<td>1.09</td>
<td>2.01</td>
<td>0.008</td>
</tr>
<tr>
<td>Academic ranking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistant professor</td>
<td>566</td>
<td>1.60</td>
<td>1.74</td>
<td>1.20</td>
<td>0.85</td>
<td>1.78</td>
<td></td>
</tr>
<tr>
<td>Associate professor</td>
<td>316</td>
<td>1.62</td>
<td>2.98</td>
<td>1.32</td>
<td>0.92</td>
<td>1.86</td>
<td></td>
</tr>
<tr>
<td>Professor</td>
<td>477</td>
<td>2.02</td>
<td>3.76</td>
<td>1.56</td>
<td>1.16</td>
<td>2.16</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Residency start yr</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤1980</td>
<td>92</td>
<td>1.73</td>
<td>1.30</td>
<td>1.53</td>
<td>1.10</td>
<td>2.00</td>
<td></td>
</tr>
<tr>
<td>1981–1990</td>
<td>247</td>
<td>1.65</td>
<td>1.21</td>
<td>1.42</td>
<td>0.97</td>
<td>2.01</td>
<td></td>
</tr>
<tr>
<td>1991–2000</td>
<td>381</td>
<td>1.96</td>
<td>3.88</td>
<td>1.48</td>
<td>1.01</td>
<td>2.10</td>
<td></td>
</tr>
<tr>
<td>&gt;2000</td>
<td>587</td>
<td>1.66</td>
<td>2.36</td>
<td>1.31</td>
<td>0.92</td>
<td>1.82</td>
<td>0.031</td>
</tr>
</tbody>
</table>

* The p values are based on the Wilcoxon rank-sum test.

### Discussion
It is useful to have a reliable field-normalized publication metric to assess academic physicians for a variety of reasons, including improvement of grant outcomes, promotions as well as offers for tenure, and continued evaluation of research productivity both within a field and between disciplines. The data presented in this study offer a more accurate means of self-evaluation by academic neurological surgeons as well as evaluation of faculty by institutional and departmental leaders. In this benchmark analysis of RCR among academic neurosurgeons, we found a strong correlation between longer career duration, having a PhD, or advanced academic rank and an increase in both RCR and weighted RCR. Male neurosurgeons were positively correlated with an increased weighted, but not mean, RCR. Overall, these findings are consistent with prior studies of the older h-index in similar subgroups of academic neurosurgeons, indicating the validity of the RCR for use in the evaluation of faculty for hiring, academic promotion, and awarding of grants.
posed of full professors progressively increased as career duration increased—that is, the longest-career subgroup (residency start date ≤ 1980, n_{total} = 92) had the highest proportion of full professors (n = 69) to assistant/associate professors. This finding reveals one of the many reasons why RCR metrics were positively correlated with career duration.

It is important to note, however, that although the RCR was positively correlated with longer career duration, the mean RCR does not unfairly disadvantage younger authors on the basis of the total number of articles published. One of the two primary criticisms of the h-index originates from the older metric’s tendency to unfairly disadvantage younger authors, who likely have not published as many articles as their older counterparts in the field. Thus, it is important for newer metrics to be able to eliminate the influence of total publication number, which will naturally increase along with career duration and age, when strictly measuring relative research productivity and impact. The design of the mean RCR is well qualified to meet this criterion, as it is calculated as an average of the RCR scores of all publications produced by an individual author. Conversely, the weighted RCR, which is calculated as a sum, is designed to take into account the total number of publications produced by an author, if such analysis is desired.

The positive correlation between increasing academic rank and higher research productivity is a finding consistent with nearly all prior literature regarding bibliometric analysis using the h-index, RCR, and other metrics in evaluating academic neurosurgeons, among other specialists. Importantly, RCR and weighted RCR scores increased significantly with a more advanced academic rank, both for the overall sample and when stratified by career duration. These data suggest that although causality cannot be assigned in one direction over the other, there is a clear correlation between academic rank and higher RCR metrics. This indicates that the RCR can be used as an accurate measure of research productivity among academic neurosurgeons and taken into consideration by teaching institutions for offers of tenure and promotion.

Sex-specific data analysis revealed no significant difference in mean RCR (p = 0.822) between male and female neurosurgeons, although males had significantly higher weighted RCR scores (p = 0.003). Similar findings regarding sex and RCR were noted in an analysis of productivity among academic radiation oncologists. As the weighted RCR is calculated as a sum over time, this correlation between male neurosurgeons and higher weighted RCR scores is likely attributable to factors such as career duration and academic rank rather than elements contributing to research productivity (e.g., frequency of publication). A study of 9952 academic physicians across medical disciplines revealed that women are underrepresented at the professor level, a subgroup that our data showed is corre-

**FIG. 1.** Overview of mean RCR data for all academic neurosurgeons. n = number of neurosurgeons. Figure is available in color online only.
lated with significantly higher RCR metrics (p < 0.0001). It is likely that the research output of female neurosurgeons would be more accurately measured when controlling for academic rank, as has been demonstrated by findings from previous studies of research productivity in neurosurgery using the h-index.12 Furthermore, because of historic biases, females have only recently entered the field of neurosurgery in statistically relevant numbers and thus represent a much smaller fraction of neurosurgeons than males. Given that weighted RCR is significantly correlated with career duration, it is expected that female neurosurgeons will have lower weighted RCR scores than males because males dominate subgroups characterized by longer career duration.8 It is also important to note that female neurosurgeons composed only 8.97% of our sample population. This disparity should be combated over the next decade, as there is a pressing need to resolve this sex disproportion and ensure equal representation and opportunities for women within this field.

Overall, academic neurosurgeons with a PhD were more productive and impactful than those without a PhD if research impact is defined as the frequency with which an author’s publications are cited by colleagues in the same field. To this end, this subset of neurosurgeons undergoes specific, multiyear training devoted to research and academic exploration within their field. The correlation between PhD acquisition and research productivity is not limited to academic neurosurgeons and has been observed in numerous other medical specialties.4,16 Interestingly, however, upon controlling for career duration, statistically significant differences between these two groups disappeared in all residency-start-date subgroups, except among individuals who started residency training after the year 2000. We can attribute this partially to the competitive nature of the field of neurosurgery and the expectation to publish extensively during both medical school and residency training. As the field becomes more and more selective, research output weighs heavily into both residency and fellowship selection processes.

Our findings also suggest that the research impact of academic neurosurgeons is relatively influential compared to the general scientific literature as evidenced by RCR values well above the benchmark ratio of 1.

Our findings also suggest that current academic neurosurgeons are exceptionally productive compared to both physicians in other specialties and the general scientific community. Overall, academic neurosurgeons had a median RCR score of 1.37 (IQR 0.93–1.97). By comparison, a study of 1299 academic radiation oncologists found a median RCR score of 1.32 (IQR 0.87–1.94).16 Among academic neurosurgeons, the median weighted RCR was 28.56 (IQR 7.99–85.65), compared to 18.0 (IQR 4.5–65.9) among radiation oncologists.16 The median RCR score for all publications currently listed in the iCite database is

FIG. 2. Overview of weighted RCR data for all academic neurosurgeons. Figure is available in color online only.
outside of this review, there exists only one other field-require extensive research before validating its use. 

**Future Directions**

Another discipline.16 Other published paper that analyzes RCR metrics within surgery and other fields of medicine, as there is only limited in the scope of comparison between neurological neurosurgeons.” Therefore, RCR values using this metric may be complex to understand and compare between various fields, which may or may not align perfectly with the aforementioned co-citation network. Finally, we are also limited in the scope of comparison between neurological surgery and other fields of medicine, as there is only one other published paper that analyzes RCR metrics within another discipline.16

**Limitations**

The RCR does suffer from some limitations. Like the h-index and all other bibliometric indicators currently in use, the RCR is unable to make distinctions between various levels of author seniority. Additionally, while field normalization is critical for accurate comparison of research output between academic fields, practical implications of the RCR’s co-citation network are not yet clear. With greater emphasis on subspecialization, particularly in medicine, distinctions between subdisciplines even within classically defined fields of academia are becoming increasingly prevalent. Field normalization by means of a co-citation network allows the mean RCR scores of individual articles to be normalized to potentially highly specific areas of study. From a practical perspective, this may mean that a high-impact paper in a popular subspecialty, such as spine surgery, may have a similar mean RCR score to a high-impact paper published for a niche subspecialty. Researchers are encouraged to investigate trends in RCR data between neurosurgical subspecialties to shed additional light on this characteristic of the new metric.

The iCite website itself does not differentiate between individuals with the same name, which can lead to errors. Authors combat this by including middle initials and excluding names for which there are an unreasonable number of publications (e.g., > 1000 publications within the last 15 years), which we could not narrow. Furthermore, only PubMed-listed publications from 1995 to 2018 are currently included in iCite, which may underrepresent the RCR for physicians who have published a large volume of work prior to 1995. Another limitation is the various degree of information and rankings listed on individual program websites. While many websites did not list ranking, some included neurosurgeon ranks such as clinical assistant professor, instructor, and lecturer, and we included their analysis under the “assistant professor” rank. For any information that was not explicitly provided via online query of department websites or via the AANS, we excluded the data from particular analyses, as seen in Tables 1–3. Furthermore, the co-citation network in which this metric is measured cannot be fine-tuned to include only authors within a specific field, for example, “academic neurosurgeons.” Therefore, RCR values using this metric may be complex to understand and compare between various fields, which may or may not align perfectly with the aforementioned co-citation network. Finally, we are also limited in the scope of comparison between neurological surgery and other fields of medicine, as there is only one other published paper that analyzes RCR metrics within another discipline.16

**Conclusions**

The RCR and its derivatives are new metrics that help fill in the gaps of other indices for research output. We found that advanced academic rank, longer career duration, and PhD acquisition were all associated with increased mean and weighted RCRs. Male sex was associated with having an increased weighted, but not mean, RCR score. Furthermore, our data showed that current academic neurosurgeons are exceptionally productive compared to both physicians in other specialties and the general scientific community. We hope that these data will be useful for individuals to perform self-evaluation through the iCite website as well as for departmental and institutional leadership seeking to evaluate current and potential faculty.

**Acknowledgments**

We thank the AANS for providing residency, fellowship, and career data on current members.

**References**

5. Bornmann L, Daniel HD: The state of h index research. Is the h index the ideal way to measure research performance? EMBO Rep 10:2–6, 2009
11. Hutchins B, Yuan X, Anderson JM, Santangelo GM: Relative citation ratio (RCR): a new metric that uses citation

specific review in radiation oncology.16 Future investigations of trends in RCR metrics within additional medical disciplines are encouraged. Future directions for RCR-related research in the field of neurosurgery include the following: RCR versus h-index analysis, RCR comparison within the various subspecialties of neurosurgery, and RCR comparison between academic institutions.

**Future Directions**

RCR is a relatively new tool and, as with any novel metric, requires extensive research before validating its use. Outside of this review, there exists only one other field-specific review in radiation oncology.16 Future investigations of trends in RCR metrics within additional medical disciplines are encouraged. Future directions for RCR-related research in the field of neurosurgery include the following: RCR versus h-index analysis, RCR comparison within the various subspecialties of neurosurgery, and RCR comparison between academic institutions.
rates to measure influence at the article level. PLoS Biol 14:e1002541, 2016

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

Author Contributions

Correspondence
Nitin Agarwal: University of Pittsburgh Medical Center, Pittsburgh, PA. agarwaln@upmc.edu.