Using the bootstrapping method to compare research achievements for authors who worked for Departments of Ophthalmology in United States since 2010

Yu-Tsen Yeh  
London Metropolitan University Libraries

Wei-Chih Kan  
Chi Mei Medical Center

Chien Tsai Wei (rasch.smile@gmail.com)  
Chi Mei Medical Center  https://orcid.org/0000-0003-1329-0679

Shu-Chun Kuo  
Chi Mei Medical Center

Research

Keywords: authorship weighted scheme, Kano model, pyramid plot, research achievement, hx-index, ophthalmology, PubMed Central

DOI: https://doi.org/10.21203/rs.3.rs-22502/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background: The h-index has its popularity in the global scientific community. Despite the h-index being used as an indicator of individual research achievement (IRA), two main disadvantages have not been solved: (1) all coauthors contributing equally to article bylines and (2) the integer nature of the h-index making it difficult to differentiate the IRAs among entities. This article evaluated the most cited authors, institutes, and states in United States in ophthalmology in recent ten years using a proposed hx-index.

Methods: Authors who worked for departments of ophthalmology in United States were selected for identifying their IRAs in Pubmed Central (PMC) since 2010. Using the PubMed search engine, we conducted an observational study of citation analyses in affiliated research institutes and states of all authors who worked for departments of ophthalmology since 2010. A total of 18,289 published articles from 46,121 authors related to departments of ophthalmology from 50 states were analyzed. The bootstrapping method was applied with an estimated 95% confidence interval (CI) to distinguish the differences in IRAs among states and institutes. The x-index and the Kano model were complemental to the h-index for identifying the group IRA characteristics and rankings. A pyramid plot was used to illustrate the importance of the author-weighted scheme (AWS) used for evaluating IRAs in academics. The hx-index combined both advantages of h-/x-index was proposed to assess IRAs for each facility. A significant difference was identified by observing two bands of estimated 95% CIs that were not overlapped. Furthermore, we drew a choropleth map on Google Maps to visualize the differences of IRA among states.

Results: There is a significant rise over time in the number of publications. The top-ranking states in hx-index based on publications and citations were Massachusetts (42.28), California (39.24), and Massachusetts (42.28). If only the top 100 authors were included for calculating the median hx-index, the top three would be California (6.45), Massachusetts (3.97), and New York (3.07) with no significant difference found among these three using the bootstrapping method. The institute and author with the highest hx-index were Harvard Medical School (Massachusetts) and Felipe A Medeiros (California), respectively. We demonstrated that Dr. Medeiros from California published 213 articles in PMC and used the example to elucidate the importance of AWS when IRAs were assessed.

Conclusions: With an overall increase in publications in the field of ophthalmology, IRAs assessed by these (1) hx-index, (2) the bootstrapping method, and (3) AWS should be emphasized and promoted more in the future.

Highlights

- We emphasized the importance of the bootstapping methods in statistics, particularly in use for comparing research achievements with 95% CI in institutes instead of rankings only.
- The research achievements in the US for authors who worked for department of ophthalmology are rare seen in literature because difficulties are to extract data for authors who were only in department of ophthalmology.
- The choropleth maps were plotted in this study. The needs for complementing choropleth maps are those including the Kano diagrams, 95% CI on a dashboard, and pyramid plot being introduced and demonstrated in this study.

Background

A few bibliometric studies have been done in the field of ophthalmology in recent years[1-3]. The h-index has its popularity in the global scientific community[5]. Despite the h-index being used popularly as an indicator of individual research achievement (IRA)[4], two main disadvantages including (1) all coauthors contributing equally to article bylines[6-8] and (2) the integer nature of the h-index making it difficult to differentiate the IRAs among entities[9] have been pointed out in the past. The fair author-weighted scheme (AWS) and the modified h-index were required accordingly for illustration and application.

In addition, numerous studies have applied the h-index to evaluate IRAs in academic rankings[10-12]. However, all those studies[10-12] were problematic in assuming h-indices following normal distribution and then using 95% confidence intervals (CI = mean +/- 2*SD) to identify the difference between groups. The assumption is not evident and the CI should be calculated by
mean +/- 2*SE, using standard error instead of standard deviation, unlike the study in ref[12]. The bootstrapping method allows estimation of the sampling distribution for almost any statistic data using random sampling methods with accuracy (as defined by 95% CI) and is practical in comparing h-indices between groups.

We were thus motivated to use three novel approaches (i.e., AWS, modified h-index, and bootstrapping method) to evaluate IRAs for authors, institutes, and the US states. All authors were affiliated to departments of ophthalmology in the US.

Three questions are raised in this study: (1) which states and institutes have significantly higher IRA, (2) which individual author has substantially higher IRA, and (3) how the AWS plays an important role in assessing IRA.

**Methods**

**Data sources**

We included only original research articles and eliminated those such as letters, editorials, and comments. A total of 18,289 results since 2010 were obtained from PubMed Central (PMC) by searching keywords ("Ophthalmology[Affiliation] and (United States[Affiliation])") on January 12, 2020. A total number of 182,241 citing articles were successfully linked to the 13,145 cited papers on the PMC. The other 5,144 articles have not been cited by any as of January 12, 2020. Study data were included in Additional File 1.

Because this study did not involve any examination or treatment on patients or reviews of patient records, it was exempt from review and approval by our research ethics committees.

**Three novel approaches used in this study**

(1) An authorship-weighted scheme (AWS) was proposed in the previous articles [8,14,15] for quantifying author contributions with the definition of all weights equal to 1.0 for each paper. Accordingly, more importance is given to the first (primary) and the last (corresponding or supervisory) authors, while it is assumed that the others (the middle authors) have made smaller contributions[1,16-18].

(2) To equip the h-index with higher discrimination power in decimal places[9], the hx-index is defined in the formula (1), where rh is the ratio (=citations/publications on the x-core of the x-index = , where all of the cited papers are denoted by ci] based on the x-core publications at i[19].

\[ hx\text{-index} = h + \frac{rh}{(1+ rh)}, \]

For instance, an author whose h-index=2 and x-index with the x-core publications and citations at 3 and 1 has the hx-index at 2.25 (=2 + 1/3(1 + 1/3)).

(3) The bootstrapping method [20] was performed to verify the difference in hx-indices among groups. A total of 1,000 median metrics were retrieved from random samples of 100 repetitions of median values for each hx-index in each group. Next, the mean and 95% confidence intervals (CI) were obtained to compare differences in hx-index among departments by inspecting whether two 95% CI bands were not overlapped. In this part of the study, the criterion of sampling authors to be included is only those with the top 100 hx-index in descending order for each group (i.e. regions or institutes).

**Representations of the research results**

1) Choropleth maps

The term “choropleth map” was coined by John Kirtland Wright in 1938[21]. The most famous example of CM used was the results of the 2000 US presidential election[22]. Other examples include showing the disparities in health outcomes across areas related to dengue outbreaks[23=24], disease hotspots[25], and the Global Health Observatory (GHO) maps on major health topics[26].
2) Kano diagrams

The Kano Model is based on a theory of product development and customer satisfaction developed in 1984 by Professor Noriaki Kano[27], who classified products or items into three main categories of quality: basic requirement, one-dimensional quality, and excitement(or, say wow) feature. These are diagrammed according to the satisfaction perceived by customers on Axis Y and the effort achieved by providers on Axis X. In this study, we attempt to demonstrate the use of the Kano diagram to characterize the feature of authors using the hx-index in bubble and x-core publications and citations on axes X and Y, respectively.

3) Pyramid plot for comparing IRAs with AWS or non-AWS in comparison

A productive author will be illustrated using the pyramid plot to present the feature of IRAs with AWS or non-AWS.

Creating Dashboards on Google Maps

All figures but the pyramid are shown by author-made modules in Excel (Microsoft Corp). We created pages of HTML used for Google Maps. All relevant h-index information on the entities can be linked to dashboards on Google Maps. The absolute advantage coefficient(ACC) was applied to assess the strength for the first entity(i.e., A1) in comparison with the next two(i.e., A2 and A3. The ACC is determined by the formula(=W/(1+W), where W=V1/V2, and V1=A1/A2, V2=A2/A3), Cutting points is set at 0.70[28].

Results

The overall impact factor is 10.38(=137345/13236) as shown in Table 1. There was a significant rise over time in the number of publications, referring to Table 1 statistics of each US states. The top three states in publication number were California, Massachusetts, and New York with proportions of 14.06%, 9.30%, and 8.39%, respectively. The top three states in hx-index were Massachusetts(42.28),California(39.24), and Massachusetts(42.28) shown in Figure 1. The ACC for Massachusetts is 0.48, indicating a weakly dominant role in IRAs among the US states. By selecting only the top 100 authors for calculating the median hx-index, the top three were California(6.45), Massachusetts(3.97). and New York(3.07), no significant difference was found among these three using the bootstrapping method, see Figure 2. The most prestigious institute and author with highest hx-index were Harvard Medical School(Massachusetts) and Felipe A Medeiros (California), respectively. The top 50 institutes and authors are shown in Figures 3 and 4, respectively. It is worth noting that the article[29] cited by 353 times since 2014 was authored by Weinreb and Medeiros from California) highlighted in Figure 4. Interested authors are invited to scan the QR-code in Figure 4. Moreover, we noticed an individual author Dr. Medeiros (California ) who published 213 articles in PMC. The importance of AWS when IRAs were assessed is present in Figure 5, where we can see the author’s impact factor increased from 20.5 to 28.3. The hx-indexes identified by either AWS or non-AWS are substantially different because the weighted hx-index will be smaller due to both weighted publications and citations less than the non-AWS ones.

Discussion

In this study, there was a significant rise in publication number over time in original research published by authors who worked for departments of ophthalmology. Citations were not included in the case.

Although choropleth maps provide a better interpretation of the disparities in health outcomes across countries/areas[31], the bootstrapping method we used in Figures 2 and 3 helps us understand the difference between entities. For instance, by examining whether the 95% CI bands overlaps can be complemental to the choropleth maps in a study.

There are other features in this study. First, the hx-index with decimal places can complement the original h-index to increase the discrimination power[9] for identifying the group IRA characteristics and rankings.
The second feature is the application of the Kano model[27] to interpret the attributes for one's IRA using a visual representation. Particularly, the dashboard-type display allows us to see details of the bibliometric indices while scanning the QR-code and clicking the entity of interest on Google Maps. Interested readers are invited to read the number of citations for each author in Figure 4 when the author bubble has been clicked.

The reasons for using x-index on two axes in Figures 4 are (1) to closely correlate to h-index [19]; (2) it was newly developed in 2018. Despite being a simple and easy-to-use metric in comparison to others, the disadvantage of using h-index has been addressed [19]. For example, an h-index of 5 means that a scientist has published five papers that each has at least five citations[10]. The h-index is little affected by those with a high volume of low-impact papers or only a few high-impact articles[10]. Excessive citations or publications were totally neglected[31,32]. The x-index was thus used to complement h-index to clearly interpret one's feature toward the citation-oriented or the productively-oriented. The hx-index shown on the Kano diagram(Figure 4) is novel and innovative in showing the author's IRAs on a dashboard.

The third feature is the dashboard of the Kano diagram combined with the hx-index on Google Maps, which is harder to create using the traditional method[33].

The fourth feature is the rankings based on the 95% CIs comparison, which provides us a deeper insight into the knowledge and information we need. The bootstrapping procedure and the dashboard creation are detailed in an MP4 video in Additional File 2.

Finally, the ACC was proposed to present the dominant role in members, which is unique to inspect whether a role is overwhelmingly played or concentrated in a field or discipline. The way using Gini coefficient for evaluating the research domain of an author[34,35] is challengeable and comparable in future studies.

**Limitations and suggestions**

Although our analysis provides quantified findings, several potential limitations were noticed and may encourage further research. First, this study only focused on authors in China, Hong Kong, or Taiwan. The results cannot be generalized to other areas and countries.

Second, certain biases might have occurred during citation extraction because of the increasing number of citations with the elapsed date. Namely, the IRA might differ if the time periods and the citation sources of the data are disparate.

Third, many authors work for multiple departments, such as in university, medical school, and affiliated hospital. Only the unit listed first was included in this study. The results of hx-index in comparison in Figures 3 might be biased if we were to apply the author-weighted scheme[8,14,15].

Fourth, although our proposed bootstrapping method was to display the 95% CIs for comparison of IRAs, many statistical software does not provide such an approach to users. The Microsoft Excel version of the bootstrapping method is provided on request to the corresponding author of this study.

Finally, despite the hx-index is considered not normally distributed in nature, the bootstrapping method we used in this study is merely selecting the mean from random samples of 100 repetitions of median values for each hx-index in each group. The 95% CI would be different if the median was retrieved from 100 repetitions of median values for each group. Future studies are recommended to compare the difference between the methods of using the mean and the median in this bootstrapping method.

**Conclusions**

The hx-index and the Kano models can complement the h-index for identifying the various IRA characteristics among groups. The bootstrapping method allows estimation of the sampling distribution for almost any statistic data using random sampling methods with accuracy (as defined by 95% CI). We highly recommend these three features used in this study, AWS, hx-index, and bootstrapping method, to be applied in future studies.
List Of Abbreviations

AIF: author impact factor
AWS: authorship-weighted scheme
JCR: journal citation report
MESH: medical subject headings
PMC: Pubmed Center
RSM: rating scale model
SCI: Scientific Citation Index
SNA: Social network analysis
VBA: visual basic for application

Declarations

Ethics approval and consent to participate
Not applicable.
All data were downloaded from MEDLINE database at pubmed.com.

Consent to publish
Not applicable.

Availability of data and materials
All data used in this study is available in Additional File files.

Competing interests
The authors declare that they have no competing interests.

Funding
There are no sources of funding to be declared.

Authors' Contributions
YT developed the study concept and design. TWC, WC and SC analyzed and interpreted the data. SBS monitored the process of this study and helped in responding to the reviewers’ advice and comments. TWC drafted the manuscript, and all authors provided critical revisions for important intellectual content. The study was supervised by SC. All authors read and approved the final manuscript.

Acknowledgements
We thank Enago (www.enago.tw) for the English language review of this manuscript.

References
1. Mimouni M, Zayit-Soudry S, Segal O, Barak Y, Nemet AY, Shulman S, Geffen N. Trends in Authorship of Articles in Major Ophthalmology Journals by Gender, 2002-2014.Ophthalmology. 2016 Aug;123(8):1824-1828.
2. Pekel E, Pekel G.Publication trends in corneal transplantation: a bibliometric analysis.BMC Ophthalmol. 2016 Nov 8;16(1):194.
3. Zhao F, Shi B, Liu R, Zhou W, Shi D, Zhang J.Theme trends and knowledge structure on choroidal neovascularization: a quantitative and co-word analysis.BMC Ophthalmol. 2018 Apr 3;18(1):86.
4. Hirsch JE. An index to quantify an individual’s scientific research output. Proc Natl Acad Sci U S A. 2005;102:16569–72. doi: 10.1073/pnas.0507655102.
5. Nowak JK, Lubarski K, Kowalik LM, Walkowiak J.H-index in medicine is driven by original research.Croat Med J. 2018 Feb 28;59(1):25-32.
6. Vavryčuk V. Fair ranking of researchers and research teams. PLoS One. 2018 Apr 5;13(4):e0195509.
7. Sekercioglu CH. Quantifying coauthor contributions. Science 2008; 322 (5900): 371.
8. Chien TW, Wang HY, Lai FJ. Applying an author-weighted scheme to identify the most influential countries in research achievements on skin cancer: observational study.JMIR Dermatol.2019; 2(1):e11015
9. Huang MH, Chi PS. A comparative analysis of the application of h-index, g-index, and a-index in institutional-level research evaluation. Journal of Library and Information Studies 2010; 8(2):1-10.
10. Doja A, Eady K, Horsley T, Bould MD, Victor JC, Sampson M.The h-index in medical education: an analysis of medical education journal editorial boards.BMC Med Educ. 2014 Nov 28;14:251
11. Panduro A, Ojeda-Granados C, Rivera H, Roman S. h Index and scientific output of researchers in medicine from the University of Guadalajara, Mexico.Rev Med Inst Mex Seguro Soc. 2015 Nov-Dec;53(6):716-21.
12. Sharma B, Boet S, Grantcharov T, Shin E, Barrowman NJ, Bould MD. The h-index outperforms other bibliometrics in the assessment of research performance in general surgery: A province-wide study. Surgery. 2011;153:493–501.
13. Lee J, Kristin KL, Couldwell WT. Use of the h index in neurosurgery. J Neurosurg. 2009;111:387–392.
14. Chien TW, Wang HY, Kan WC, Su SB.Whether article types of a scholarly journal are different in cited metrics using cluster analysis of MeSH terms to display: A bibliometric analysis.Medicine (Baltimore). 2019 Oct;98(43):e17631.
15. Lin CH, Chou PH, Chou W, Chien TW.Using the Kano model to display the most cited authors and affiliated countries in schizophrenia research.
16. Mimouni M, Zayit-Soudry S, Segal O, Barak Y, Nemet AY, Shulman S, Geffen N. Trends in Authorship of Articles in Major Ophthalmology Journals by Gender, 2002-2014.Ophthalmology. 2016 Aug;123(8):1824-1828.
17. Lindsey D. Further evidence for adjusting for multiple authorship. Scientometrics. 1982; 4(5): 389–395.
18. Chien TW, Wang HY, Chang Y, Kan WC. Using Google Maps to display the pattern of coauthor collaborations on the topic of schizophrenia: A systematic review between 1937 and 2017. Schizophr Res. 2018; 204, 206-213
19. Fenner T, Harris M, Levene M, Bar-Ilan J. A novel bibliometric index with a simple geometric interpretation. PLoS One. 2018;13(7):e0200098.
20. Efron B. Bootstrap methods: Another look at the jackknife. The Annals of Statistics 1979; 7(1): 1-26.
21. Wright JK . Problems in Population Mapping in Notes on statistical mapping, with special reference to the mapping of population phenomena. American Geographical Society; Washington, Population Association of America, NY, 1938. 2019/4/4 available at https://www.worldcat.org/title/notes-on-statistical-mapping-with-special-reference-to-the-mapping-of-population-phenomena/oclc/5160537/
22. Newman ME, Gastner MT. Diffusion-based method for producing density-equalizing maps.Proc Natl Acad Sci U S A 2004;101(20): 7499–7504.
23. Chen WJ. Dengue outbreaks and the geographic distribution of dengue vectors in Taiwan: A 20-year epidemiological analysis.Biomed J 2018;41(5):283-289.
24. Lai WT, Chen CH, Hung H, Chen RB, Shete S, Wu CC. Recognizing spatial and temporal clustering patterns of dengue outbreaks in Taiwan.BMC Infect Dis. 2018;18(1):256.
Table 1 Publications with first authors in China, Hong Kong, and Taiwan
| State          | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | Total | % Citations | IF  |
|----------------|------|------|------|------|------|------|------|------|------|------|-------|-------------|-----|
| California     | 102  | 106  | 109  | 127  | 194  | 240  | 228  | 240  | 270  | 245  | 1861  | 14.06       | 1815 9.76 |
| Massachusetts  | 69   | 77   | 82   | 109  | 142  | 132  | 155  | 173  | 151  | 141  | 1231  | 9.30        | 1526 12.40 |
| New York       | 53   | 41   | 51   | 77   | 120  | 155  | 159  | 153  | 175  | 126  | 1110  | 8.39        | 1014 9.14  |
| Pennsylvania   | 34   | 43   | 52   | 63   | 73   | 81   | 98   | 99   | 77   | 84   | 704   | 5.32        | 7597 10.79 |
| Michigan       | 38   | 43   | 40   | 40   | 61   | 76   | 99   | 107  | 108  | 87   | 699   | 5.28        | 7669 10.97 |
| Illinois       | 48   | 51   | 57   | 44   | 66   | 96   | 74   | 85   | 78   | 82   | 681   | 5.15        | 5637 8.28  |
| Maryland       | 22   | 34   | 27   | 35   | 92   | 86   | 98   | 89   | 82   | 66   | 631   | 4.77        | 7001 11.10 |
| Texas          | 50   | 50   | 43   | 42   | 64   | 77   | 85   | 85   | 56   | 59   | 611   | 4.62        | 5784 9.47  |
| Wisconsin      | 44   | 30   | 38   | 54   | 57   | 58   | 52   | 66   | 56   | 67   | 522   | 3.94        | 4506 8.63  |
| Florida        | 32   | 30   | 43   | 34   | 61   | 50   | 56   | 60   | 63   | 59   | 488   | 3.69        | 4547 9.32  |
| Washington     | 21   | 20   | 22   | 23   | 48   | 79   | 59   | 82   | 73   | 55   | 482   | 3.64        | 7015 14.55 |
| North Carolina | 21   | 21   | 26   | 27   | 43   | 56   | 58   | 55   | 54   | 63   | 424   | 3.20        | 4388 10.35 |
| Ohio           | 30   | 25   | 22   | 44   | 41   | 45   | 42   | 40   | 41   | 39   | 369   | 2.79        | 4468 12.11 |
| Iowa           | 21   | 27   | 22   | 22   | 35   | 39   | 43   | 47   | 29   | 36   | 321   | 2.43        | 3928 12.24 |
| Georgia        | 23   | 18   | 19   | 29   | 35   | 50   | 34   | 42   | 37   | 29   | 316   | 2.39        | 2597 8.22  |
| Utah           | 9    | 23   | 30   | 17   | 30   | 34   | 35   | 33   | 37   | 31   | 279   | 2.11        | 2982 10.69 |
| Minnesota      | 28   | 28   | 20   | 20   | 29   | 26   | 32   | 32   | 31   | 31   | 277   | 2.09        | 2430 8.77  |
| Tennessee      | 10   | 17   | 14   | 20   | 31   | 27   | 25   | 22   | 24   | 26   | 216   | 1.63        | 1935 8.96  |
| Oklahoma       | 13   | 9    | 17   | 13   | 30   | 25   | 31   | 29   | 23   | 22   | 212   | 1.60        | 1693 7.99  |
| Alabama        | 8    | 8    | 10   | 13   | 23   | 27   | 32   | 19   | 25   | 23   | 188   | 1.42        | 2237 11.90 |
| Others         | 109  | 105  | 95   | 111  | 175  | 210  | 196  | 189  | 221  | 203  | 1614  | 12.19       | 1736 10.76 |
| **Total**      | 785  | 806  | 839  | 964  | 1450 | 1669 | 1691 | 1747 | 1711 | 1574 | 1323  | 1.00        | 1373 10.38 |

**Figures**
Figure 1

The top three states in hx-index were Massachusetts(42.28), California(39.24), and Massachusetts(42.28)
Figure 2

By selecting only the top 100 authors for calculating the median hx-index, the top three were California(6.45), Massachusetts(3.97), and New York(3.07), no significant difference was found among these three using the bootstrapping method.
Figure 3

No legend.
Figure 4

It is worth noting that the article[29] cited by 353 times since 2014 was authored by Weinreb and Medeiros from California) highlighted
The importance of AWS when IRAs were assessed is present in Figure 5, where we can see the author's impact factor increased from 20.5 to 28.3

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile2.docx
- datasetd.xlsx