Thai J Ophthalmol 2563;34(1):30-38.

Original Article/นิพนธ์ต้นฉบับ

Bibliometric Differences between Scopus and Google Scholar for Ophthalmology Academics in Thailand: A Comparative Study

การศึกษาเปรียบเทียบค่าบรรณมิติระหว่างฐานข้อมูล Scopus และ Google Scholar ของอาจารย์ ในภาควิชาจักษุวิทยา คณะแพทยศาสตร์ศิริราชพยาบาล

Supattra Sawangkul, BBA[1] Warinyupa Pinitpuwadol, MD[1] Darin Sakiyalak, MD[1]

สุพัตรา สว่างกุล, บธ.บ.[1] วรินยุพา พินิตภูวดล, พ.บ.[1] ดารินทร์ สากิยลักษณ์, พ.บ.[1]



Pitipol Choopong, MD[1]

ปิติพล ชูพงศ์, พ.บ.[1]

Abstract

Objective: To compare the bibliometric indices between Scopus and Google Scholar databases using data from the Department of Ophthalmology, Siriraj Hospital, Mahidol University, Thailand.

Methods: This cross-sectional study was conducted in March, 2020. Bibliometric parameters of each academic from Scopus and Google Scholar databases, including those individuals' publications, citations, and *h*-index, were collected. Descriptive and comparative statistics analyses were performed to identify

Corresponding author: Pitipol Choopong, MD

Department of Ophthalmology, Faculty of Medicine Siriraj Hospital, Mahidol University, 2 Wanglang Road, Bangkoknoi, Bangkok 10700, Thailand

Tel: (+66) 2-419-8037; Fax: (+66) 2-414-1232; E-mail: pitipol.cho@mahidol.edu

[1]*Department of Ophthalmology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.*

[1]*ภาควิชาจักษุวิทยา คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล*

the differences between the two databases.

Results: The data included the bibliometrics from 28 academic staff members. A total of 500 publications was identified from Scopus and 566 from Google Scholar. There were 8,793 citations recorded in Scopus vs. 13,903 in Google Scholar. The median *h*-index was 6 (range: 1-20) and 7 (range: 1-21) in Scopus and Google Scholar, respectively. All indices from Google Scholar were statistically significantly higher than those from Scopus. There were high concordances between both databases in the number of publications and *h*-index. However, the patterns of agreement were different. Moreover, there were increasing trends of bibliometric indices according to the higher academic ranks of the staff members.

Conclusion: Scopus and Google Scholar databases were parallel in displaying bibliographic indices; however, Google Scholar showed higher values. Regarding the comparability of the scholars' productivity, it is very important to identify which database was used to calculate the bibliometric indices.

Keywords: bibliometrics, Scopus, Google Scholar, ophthalmology, *h*-index

บทคัดย่อ

วัตถุประสงค์: เพื่อวิเคราะห์เปรียบเทียบค่าดัชนีบรรณมิติในการอ้างอิงผลงานทางวิชาการของคณาจารย์ ในภาควิชาจักษุวิทยา คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล ที่ปรากฏอยู่ในฐานข้อมูล Scopus กับ Google Scholar

วิธีการศึกษา: เก็บข้อมูลบรรณมิติของคณาจารย์ ในเดือนมีนาคม 2563 ได้แก่ จำนวนบทความวิจัยที่ได้รับการตีพิมพ์ และจำนวนครั้งที่ได้รับการอ้างอิง รวมถึงข้อมูลค่า *h*-index จากฐานข้อมูล Scopus และ Google Scholar แล้วนำไปวิเคราะห์ข้อมูลเชิงพรรณนาและเชิงเปรียบเทียบ เพื่อหาความแตกต่างของค่าบรรณมิติระหว่างฐานข้อมูลทั้งสอง

ผลการศึกษา: พบจำนวนอาจารย์ในภาควิชาจักษุวิทยา ที่อยู่ในช่วงระยะเวลาที่เก็บข้อมูล ทั้งหมด 28 คน โดยมีผลงานตีพิมพ์จำนวนรวม 500 เรื่องและได้รับการอ้างอิงรวม 8,793 ครั้ง ในฐานข้อมูล Scopus ส่วนในฐานข้อมูล Google Scholar พบจำนวนผลงานตีพิมพ์รวม 566 เรื่อง และได้รับการอ้างอิงรวม 13,903 ครั้ง ค่ามัธยฐานของ *h*-index ในฐานข้อมูล Scopus =6 (พิสัย 1-20) และในฐานข้อมูล Google Scholar =7 (พิสัย 1-21) จากข้อมูลดังกล่าวพบว่า ฐานข้อมูล Google Scholar มีค่าดัชนีบรรณมิติมากกว่า Scopus อย่างมีนัยสำคัญ แต่ก็พบว่ามีความสอดคล้องกันมากระหว่างฐานข้อมูลทั้งสอง ในด้านจำนวนผลงานตีพิมพ์และค่า *h*-index อย่างไรก็ตามรูปแบบของการสอดคล้องของดัชนีทั้งสองมีความแตกต่างกัน นอกจากนี้ยังพบว่า มีแนวโน้มของค่าบรรณมิติที่เพิ่มขึ้นตามความก้าวหน้าในตำแหน่งวิชาการ

สรุป: ฐานข้อมูล Scopus และ Google Scholar มีค่าดัชนีบรรณมิติ ไปในแนวทางเดียวกัน แต่ฐานข้อมูล Google Scholar มีการเข้าถึงข้อมูลได้ครอบคลุมมากกว่าฐานข้อมูล Scopus จึงได้ผลลัพธ์ที่สูงกว่า ดังนั้นการเลือกฐานข้อมูลในการคำนวณค่าบรรณมิติ จึงเป็นสิ่งสำคัญในการประเมินประสิทธิภาพของการผลิตผลงานทางวิชาการของอาจารย์

คำสำคัญ: บรรณมิติ, ฐานข้อมูล Scopus, ฐานข้อมูล Google Scholar, จักษุวิทยา, *h*-index

Introduction

Bibliographic records have become important tools in appraising a scholar on the basis of his or her academic activity.[1,2] The records are known as bibliometrics, which can be measured in many different ways to evaluate researchers' ability, productivity, and the impact of their publications. The traditional method was simply to count a person's academic publications. However, the quantity of published works cannot determine their quality. Instead, research citations emphasize the quality and impact of a publication, which then determine the author's academic influence. In 2005, the *h*-index was introduced by Jorge Hirsch to balance the scores pertaining to the quantity and quality of a researcher's works.[3]

A researcher's *h*-index is *h* when their *h* publications reached at least *h* number of citations. Publications, citations, and *h*-index are the three main indices to evaluate the productivity of an academic. With the goals of research and overall academic excellence in mind, these bibliometric data are increasingly considered to be reliable, efficient assessing tools of both researchers and their institutes.[4,5]

With the evolution of the digital era, many medical databases have been developed and then presented on the World Wide Web offering information on scholars' publications and citation data. These records are used to calculate the bibliometric parameters and to do citation analyses. Among the notable databases, Scopus and Google Scholar are famous for collecting broad ranges of publications and providing the key bibliometric parameters, including the number of publications, citations, and the *h*-index.[2,6,7]

Scopus is a well-known citation database made by Elsevier, a major publishing company. Scopus contains more than 23,000 peer-reviewed journals focusing on life sciences, physical sciences, health sciences, and social sciences from over 5,000 publishers. The content types on Scopus include scientific journals, books, and conference proceedings that are indexed in the database. However, it does not provide data outside its indexed publications, which are mostly limited to the English language. Moreover, Scopus is available only with a paid annual subscription.

In contrast, Google Scholar is a free database provided by Google. With its data-mining ability, it includes wide ranges of online academic documents, such as journal articles, books, dissertations/theses, and conference papers, including articles in many languages. Additionally, based on the Google search engine, it updates data as soon as a new academic publication is released. Nevertheless, Google does not inform users about its search algorithm and coverage.

Although Scopus and Google Scholar deliver overlapping data and parameters, there are substantial differences in their patterns of coverage and the counts of the content. Most studies have found that Google Scholar covered more articles and citation references in English and other languages.[8] However Bakkalbasi and colleagues found that the results differed according to the publication year and the discipline.[9] For local academics, many works were published in their local language. Most of those articles are not indexed in the Scopus database. Thai researchers,

as well as others around the world, may suffer from this limitation in assessments of their productivity.

Therefore, understanding the different results of bibliometric indices from different databases is very useful in evaluating a scholar's productivity. The objective of this study is to compare the bibliometric indices between Scopus and Google Scholar using data from academics in the Department of Ophthalmology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand.

Methods

Ethical Approval The study protocol was approved by the Ethics Committee of the Siriraj Institutional Review Board (COA no. Si 044/2018).

We conducted a cross-sectional study in March, 2020. Bibliometric measurements of the international publications of all academics in the department were included in the study. We assessed the bibliometric data from the Scopus database through its official website (https://www.scopus.com). For the Google Scholar database, we extracted researchers' bibliometric data using Publish or Perish® version 6 (Anne-Wil Harzing, Middlesex University, London).

The search term consisted of first and last names of the academics in the Department of Ophthalmology, Siriraj Hospital, Mahidol University, Bangkok, Thailand appeared as first, co-, or corresponding author. All publications in English and local languages were included. One reviewer (SS) performed searching and screening of publications.

Bibliometric indices for each scholar included publication counts, citation counts, and the *h*-index (*h* number of publications with an at least *h* number of citations). All duplicated publications and citations were excluded from the calculation of all indices. Demographic information, including gender, academic rank, and subspecialty, was also collected. Academic rank was classified as follows: professor, associate professor, assistant professor, and instructor. The subspecialty was classified as Retina, Cornea and Refractive Surgery, Pediatric Ophthalmology, Orbit and Oculoplastic Surgery, Glaucoma, Ocular immunology/uveitis, and Neuro-Ophthalmology.

Descriptive statistics were shown as frequency and percentage for categorical data and as median and range for continuous data. The differences in the bibliometric indices between the two databases was determined using the Wilcoxon signed-rank test. A p-value of less than 0.05 indicated a statistical significance. A Bland-Altman plot and concordance-correlation coefficient were conducted to evaluate the level of agreement between both databases. All statistical analyses were performed using *Stata Statistical Software: release 16* (StataCorp., Texus, USA).

Results

A total of 28 active academics was identified in March, 2020. There were 18 females and 10 males. The majority, in terms of academic rank, were associate professors (15 individuals, or 53.6%). Retina, Cornea and Refractive Surgery comprised the highest proportion among the subspecialties (Table 1).

Table 1. Demographic distributions of academics in the Department of Ophthalmology, Siriraj Hospital, Mahidol University in Bangkok, Thailand

Characteristics	n (%)
Gender	
Female	18 (64.3)
Male	10 (35.7)
Academic rank	
Professor	3 (10.7)
Associate professor	15 (53.6)
Assistant professor	4 (14.3)
Instructor	6 (21.4)
Subspecialty	
Retina	6 (21.4)
Cornea and refractive surgery	6 (21.4)
Pediatric ophthalmology	5 (17.9)
Glaucoma	4 (14.3)
Ocular immunology/uveitis	3 (10.7)
Neuro-ophthalmology	2 (7.1)
Orbit and oculoplastic surgery	2 (7.1)
Total	**28 (100)**

A total of 500 publications was identified from Scopus and 566 from Google Scholar for 28 researchers. All were research articles published in English language except five Thai language articles from Google Scholar. Of these, there were 453 overlapping publications found in both databases. Forty-seven articles were recorded only in Scopus, while 112 were found only in Google Scholar. The median number of individual publications was 14 in Scopus vs. 15.5 in Google Scholar ($p = 0.004$). The total number of citation counts was 8,793 in the Scopus database and 13,903 in the Google Scholar database. There was a statistically significant difference in the individual citation count between Scopus (128.5) and Google Scholar (226.5), $p < 0.001$. The sum *h*-index was 188 and 225 from Scopus and Google Scholar, respectively. The individual researcher's *h*-index differed significantly between Scopus[6] and Google Scholar[7], $p < 0.001$. The details are shown in Table 2.

A series of Bland-Altman plots are illustrated in Fig. 1, which assesses the levels of agreement in the publication count, citation count, and *h*-index between Scopus and Google Scholar. The mean value was greater for all indices from Google Scholar. The average difference in the publication count was 2.36. The differences varied when the number of publications increased (Figure 1a). The mean differences for the citation count and *h*-index were 1.85 and 1.32, respectively. There was a positive trend of difference proportionately to the magnitude of the citation count and *h*-index (Figures 1b and 1c). The concordance-correlation coefficients (CCC) between the two databases for publications, citations, and the *h*-index and their 95% confidence interval were 0.94 (0.88-0.97), 0.86 (0.82-0.99), and 0.95 (0.91-0.97), respectively.

Distributions of bibliometric indices from Scopus and Google Scholar by academic ranks are shown in Figure 2. The subgroup comparisons demonstrated increasing trends for all indices when the scholar's rank progressed from instructor to professor. Indices from Google Scholar were higher than those from Scopus in all academic ranks, except for the *h*-index of instructor, which was equal at the value of 1. High variations were observed in the number of publications of associate professors, the citations of instructors,

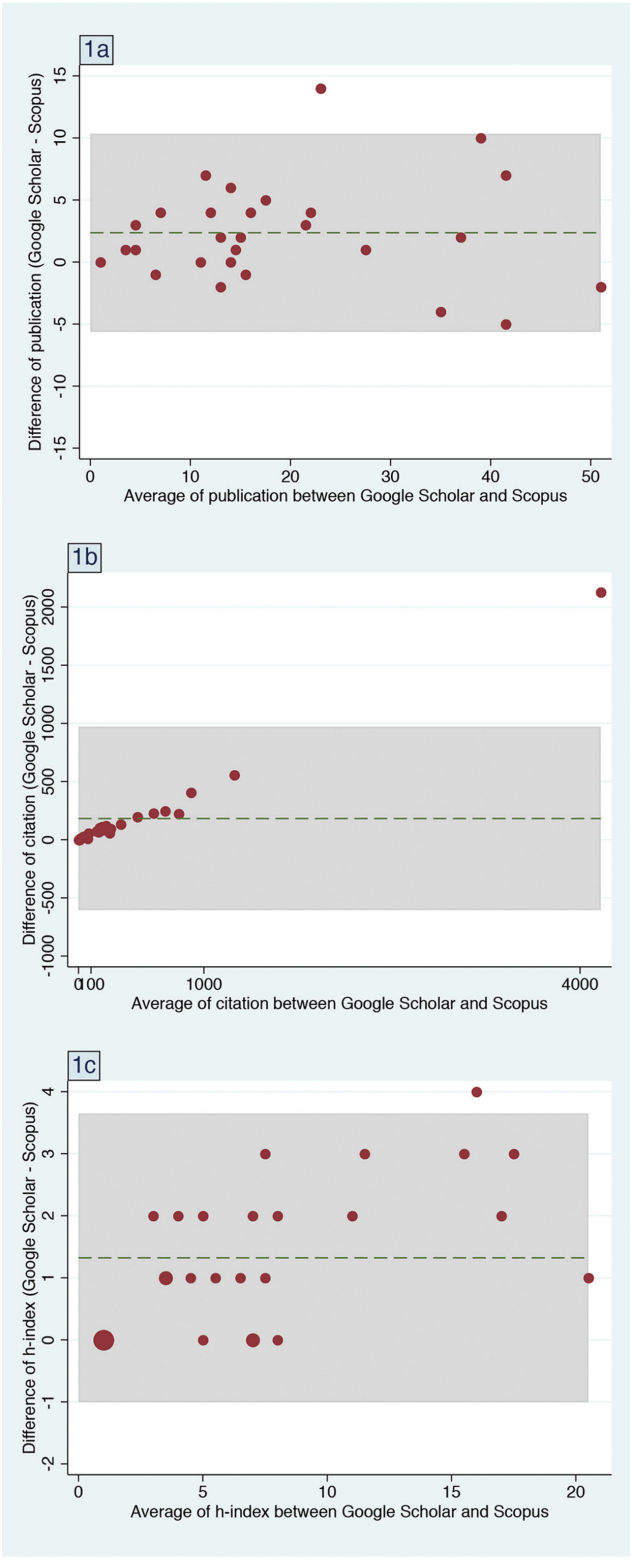

Figure 1. Bland-Altman plots demonstrating the agreement of bibliometric indices between Scopus and Google Scholar: 1a) Publication counts 1b) Citation counts 1c) *h*-index

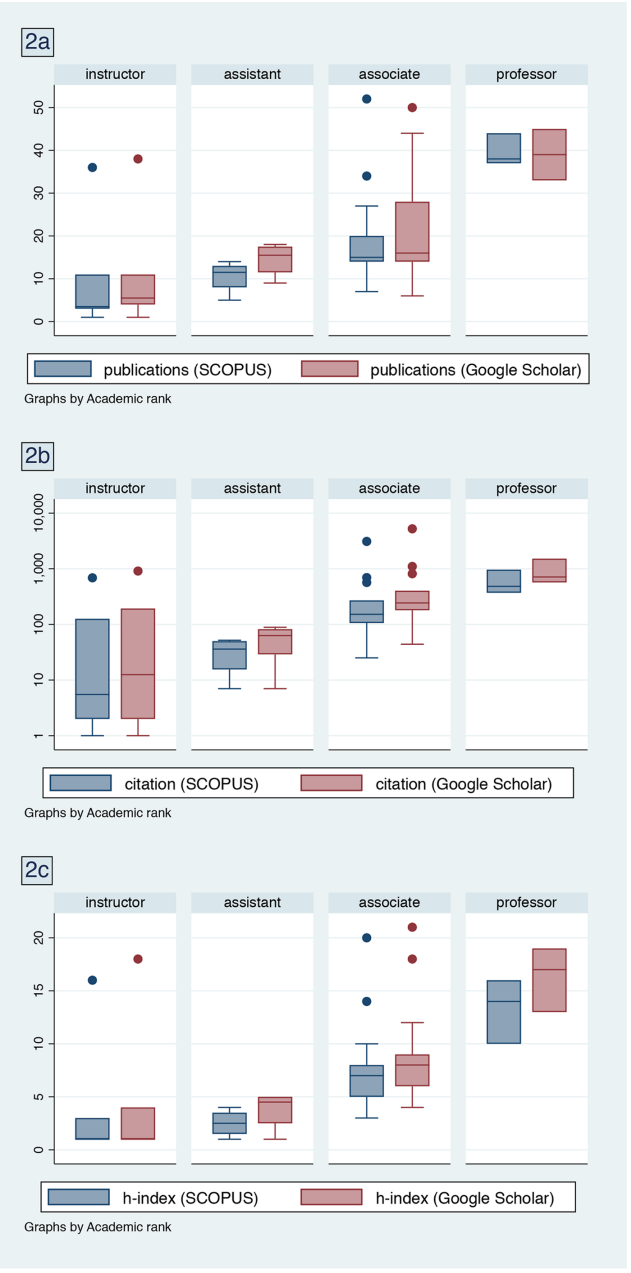

Figure 2. Distribution of bibliometric indices from Scopus and Google Scholar by academic rank: 2a) Publication counts 2b) Citation counts 2c) *h*-index

and the *h*-index of professors.

Discussion

According to the bibliometric records of our academics, we found high agreements

in the indices between the two databases, including publication count, citation count, and *h*-index. Notably, the publication count and *h*-index demonstrated a very high concordance. Nonetheless, there were biases against Scopus. All indices from Google Scholar were higher than for Scopus. The greater article coverage of Google Scholar may be the main reason for these biases. In the subgroup analysis, all indices increased with the progression of academic rank. This finding reflected the academic growth path of individuals, as one of the criteria to obtain an academic rank is a specified number and quality of publications. Therefore, a researcher with higher productivity has more chance to progress to a higher rank.

Our findings are different from those in a report by Bakkalbasi et al. in 2006.[9] They compared citation counts for articles from oncology and condensed-matter physics among Scopus, Web of Science, and Google Scholar. In their study, Google Scholar returned the lowest number of citing references for both subjects. The differences were more extensive in older publications until 1993. This result suggests that the coverage of Google Scholar may not include articles published in an earlier era when the internet and electronic data were hardly as essential as now. The majority of our staff started their career after the year 2000, when electronic data had become widely used and in fact ubiquitous.

Studies in recent years have demonstrated similar results to our own. Google Scholar has consistently demonstrated higher indices compared to Scopus. The differences in the number of citations were very clear in the area of general medicine, mathematics and computer sciences, as well as the humanities.[8,10,11] Nevertheless, the pattern of differences varied among disciplines and regions.[7,8,10-12]

Bar-Ilan and colleagues reported higher differences among almost all Israeli mathematicians and computer scientists, while much lower differences were interestingly observed for more highly cited researchers. Furthermore, a study from South Africa also found that Google Scholar retrieved higher citations than did Scopus. However, Google Scholar contained articles with multiple copies and showed the most inconsistencies, such as authors' name and sequence, as well as the volume and issue number of the publication. On the other hand, Scopus delivered no duplicates and the least inconsistencies.[7] Additionally, the most recent study found that the larger number of citations from Google Scholar included most citations found in Scopus.[8] The counts and correlations between these databases nevertheless varied according to the coverage of each database in those disciplines.

In the era of the Internet of Things, both databases have progressively increased their citation coverage, leading to more accurate bibliometrics. Scopus includes content systematically with links to an individual author, making it a reliable source for bibliometric data. However, cited references in the journals that are indexed in other databases are often not visible in Scopus.[13] In contrast, Google Scholar has advantages over the basic Google search engine, in that it can provide more comprehensive data

such as the articles published online in local languages and various types of citing references. However, Google Scholar includes non-academic content from websites, grey literature, non-peer-reviewed articles, and dissertations/theses.[6] Thus, the results may be inflated over the actual value of the metrics.

Our study demonstrates that both databases have had a high concordance, especially in terms of the number of publications and the *h*-index. Also, the metrics of individual scholar seemed to be consistent. So, for a comparison of productivity among local researchers, either database is suitable. Google Scholar has benefits in its coverage of non-English articles. However, when compared across databases, bias corrections should be done because indices from Google Scholar were generally higher than those from Scopus. Notably, the differences were broader with the increasing bibliometrics values. The correlations may also change, depending on the subgroup of researchers. Tschudy and colleagues also reported a correlation coefficient of the *h*-index between Scopus and Google Scholar of as low as 0.79 in the group of pediatric chairpersons.[14] As a result, comparisons across databases are not recommended, especially regarding researchers with high indices.

We moreover found a wide difference in the citation counts between the databases (Table 2). The variation is much higher than for other indices and is highest in the instructor group (Figure2). Apparently, Google Scholar covers local or non-English-language articles more than Scopus did. Another explanation could be that researchers in their early career prefer to publish their articles in their native language. As a consequence, those articles may only be cited in other local language articles. For accuracy, a detailed exploration and data-cleaning should be done. Further searching in the local databases will help to identify additional publications in local languages.

The main limitation of our study is that our sample size was small. Nonetheless, with the level of significance detected in our results, it should be concluded that there were differences in the bibliometric indices between Scopus and Google Scholar. Additionally, since we sampled our department's scholars, selection bias was unavoidable. This limitation might obscure the generalizability of our study to other ophthalmology departments. Future research may be done on a larger scale, especially country-wide, with comparisons at the researcher, journal, and institutional levels.

In conclusion, bibliometrics indices, including publication counts, citation counts, and the *h*-index differed between the Scopus and Google Scholar databases. As Google Scholar contained a broader range of documents, higher metrics were found from this resource; however, both databases demonstrated a high agreement. For comparability, any corrections should be adjusted carefully, especially for an academic with a high value of indices.

Acknowledgements

We would like to thank Nutnicha Neti, MD and Ms. Wilawan Sanpan for preparing this manuscript for submission.

Conflicts of Interest: None

References

1. Mering M. Bibliometrics: understanding author-, article- and journal-level metrics. Serials Rev. 2017; 43:41-5.
2. Choudhri AF, Siddiqui A, Khan NR, et al. Understanding bibliometric parameters and analysis. Radiographics. 2015;35:736-46.
3. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci USA. 2005;102:16569-72.
4. Agarwal A, Durairajanayagam D, Tatagari S, et al. Bibliometrics: tracking research impact by selecting the appropriate metrics. Asian J Androl. 2016;18:296-309.
5. Garner RM, Hirsch JA, Albuquerque FC, et al. Bibliometric indices: defining academic productivity and citation rates of researchers, departments and journals. J Neurointerv Surg. 2018;10:102-6.
6. Falagas ME, Pitsouni EI, Malietzis GA, et al. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. FASEB J. 2008;22:338-42.
7. Adriaanse L. Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison. Electron Libr. 2013;31:727-44.
8. Martín-Martín A, Orduna-Malea E, Thelwall M, et al. Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. J Informetr. 2018;12:1160-77.
9. Bakkalbasi N, Bauer K, Glover J, et al. Three options for citation tracking: Google Scholar, Scopus and Web of Science. Biomed Digit Libr. 2006;3:7.
10. Bar-Ilan J. Which h-index? - A comparison of WoS, Scopus and Google Scholar. Scientometrics. 2008; 74:257-71.
11. Kulkarni AV, Aziz B, Shams I, et al. Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. JAMA. 2009;302:1092-6.
12. Khan NR, Thompson CJ, Taylor DR, et al. An analysis of publication productivity for 1225 academic neurosurgeons and 99 departments in the United States. J Neurosurg. 2014;120:746-55.
13. Li J, Burnham JF, Lemley T, et al. Citation Analysis: Comparison of Web of Science®, Scopus™, SciFinder®, and Google Scholar. J Electron Resour Med Libr. 2010;7:196-217.
14. Tschudy MM, Rowe TL, Dover GJ, et al. Pediatric Academic Productivity: Pediatric Benchmarks for the h- and g-Indices. J Pediatr. 2016;169:272-6.