January 2021

Citations of the Top 100 Most-cited Papers of the Journal Scientometrics in Web of Science and its Association and Correlation with Scopus and Google Scholar Citations

Renjith V R
renjithliber@gmail.com

Pradeepkumar A P
geo.pradeep@keralauniversity.ac.in

Citations of the Top 100 Most-cited Papers of the Journal Scientometrics in Web of Science and its Association and Correlation with Scopus and Google Scholar Citations

Renjith. V. R.
Reference Assistant
Campus Library, Kariavattom Campus
University of Kerala, Thiruvananthapuram, India
ORCID: 0000-0002-2389-7338
renjithliber@gmail.com

A. P. Pradeepkumar
Professor
Department of Geology
University of Kerala, Thiruvananthapuram, India
ORCID: 0000-0002-4629-509X
geo.pradeep@keralauniversity.ac.in

ABSTRACT

This study is conducted to analyse the citations of the top 100 most-cited papers of the journal ***Scientometrics*** *in Web of Science (WoS) and its association and correlation with corresponding citations in Scopus and Google Scholar (GS). Chi-square and Spearman's rank rho are used to ascertain the association and correlation among these citations in different platforms. GS citations for the papers are comparatively higher than the citations in the other two databases. Scopus citations are slightly higher than WoS citations. The study found that there is significant association among level of citations of top 100 cited papers of the journal Scientometrics in WoS and its corresponding level of citations in Scopus and Google Scholar and also high positive correlation among citations in the three databases.*

Key Words: Scientometrics, Library and Information Science Journal, Citations, Highly Cited Papers, Top 100 Cited Papers, Scopus, Google Scholar, Web of Science

1. Introduction

Scientific and scholarly writers usually cite other scholars' publications while writing research papers, as part of bibliographical reference to other scholarly documents in the text and also elaborate them in bibliographical form in the reference list of their papers. These are commonly

termed as cited references. Thus published papers receive citations. Authors adopt different referencing styles like the MLA style, APA style and Chicago style etc. which contain a set of standardized information about the cited documents to enable its tracing. A citation index is a paper based or electronic database that provides citation links between documents. There are several citation indexes like Web of Science (WoS), Scopus, Google Scholar (GS), Microsoft Academic, Crossref, Dimensions etc. The first modern citation index was proposed by renowned information scientist Eugene Garfield in 1955 and made practical by him in 1964. This is considered as an innovative step in knowledge organisation and information retrieval.

The WoS and Scopus are the two multidisciplinary subscription based citation indexes used to rank journals in a particular discipline to measure them in terms of productivity, total citations received so as to indicate the journal impact, influence or prestige within the subject discipline. WoS is a platform created in 1997 and renamed Web of Science Core Collection in 2014 consisting of databases designed to support scientific and scholarly research. Web of Science (WoS) Core Collection especially covers its three classical journal citation indexes, i.e. Science Citation Index Expanded (SCIE), Social Science Citation Index (SSCI), and Arts and Humanities Citation Index (A&HCI), are well known and widely used among academic scholarly community. The renowned international publisher Elsevier released Scopus in 2004 as a newcomer in the citation indexing field. Although it is a newcomer it is considered as a powerful competitor of Web of Science and is attempting to challenge the dominating role of WoS and as part of that various studies have been conducted to compare these two databases from different perspectives. Google Scholar (GS) was launched in November 2004 and was originally intended as a tool for researchers to find and retrieve the full text of documents. Its outstanding feature is that it is a free academic search engine and citation index, indexing full text and metadata of scholarly literature, across disciplines. GS is an altmetric journal citation-based indicator, and it covers a wider variety of document types and sources than Scopus and WoS (Thelwall & Kousha, 2015).

Scientometrics is an international open access journal jointly published by Academia Kiado (Budapest) and Springer publishing original studies, short communications, preliminary reports, review papers, letters to the editor and book reviews on scientometrics. The journal is concerned with the qualitative features and characteristics of science and scientific research and emphasis is placed on investigations in which the development and mechanism of science are studied by statistical mathematical methods (springer.com/journal/11192). The journal is indexed in Web of

Science and has an impact factor value 2.867 (2019) and its five-year impact factor is 2.710 (2018) (springer.com/journal/11192). Scopus also indexes this journal, which is included in quartile number one with SJR value 1.210 and h index 106 for the year 2019. The journal's cite score for the same year is 5.6 (scimagojr.com). It has high visibility and discoverability of authors and papers in GS. Thus ***Scientometrics*** is one of the most influential or important ones in the area, which is being indexed by WoS, Scopus and GS. In this background, the present study analyses the citations of the top 100 cited papers of ***Scientometrics*** in WoS and its association and correlation with corresponding Scopus and GS citations.

2. Related Studies

Yang and Meho (2006) presented a case study comparing citations found in Scopus and GS with those found in WoS for items published by two full-time faculty members of the School of Library and Information Science (LIS) at Indiana University, USA. The paper also presented a brief overview of a prototype system called CiteSearch. Combined data from multiple citation databases are analysed by CiteSearch, generating citation-based quality evaluation factors. The study showed that WoS should not be used alone for locating citations to an author or title. Scopus and GS can help in identifying a considerable number of valuable citations not found in WoS. Scopus and GS can help in identifying a considerable number of citations in document types not covered by ISI citation databases.

Martell (2009) conducted a search of 217 articles in College and Research Libraries from 2000 to 2006. The search was conducted by using the title on Yahoo, Google, GS, and ISI Web of Knowledge to find out the frequency with which articles are cited, thereby assessing the effectiveness of the four search services. The results showed that Yahoo, Google and ISI Web of Knowledge averaged between 2.8 and 3.5 citations per title for the period covered and GS averaged 6.4.

In a paper in ***Scientometrics,*** Abrizah et al. (2012) compared the coverage, ranking, impact and subject categorisation of LIS journals, using 79 titles based on data from WoS and 128 titles from Scopus. The study found that a total of 45 titles covered in both databases with normalised impact factors being higher for titles covered in Scopus. Furthermore, Scopus covered more unique titles (n=72) than did WoS (n=23). This study showed that the two databases differ in the number of journals covered and the impact factor is higher in Scopus than in WOS, due to wider coverage of LIS journals in the former.

Renjith (2018) attempted to make an assessment of the visibility and impact of Indian LIS journals on the basis of scientometric indicators using data from GS with the help of 'Publish or Perish (PoP)' software. The contents in the selected journals published during the period 2010-2015 and citations received to these contents during 2010-2018 (up to July) were subjected to analysis. The visibility of journals is estimated in terms of the number of papers actually published in the journals, computed manually, and the number of postings available in GS and the number of citations received by the papers, estimated using PoP. The study showed that Indian LIS journals have visibility and citation impact in GS.

Renjith (2019) in another paper attempted to highlight the authorship pattern and citation level of i10 cited research articles in DESIDOC Journal of Library and Information Technology (DJLIT) based on GS data. The study established that citations of i10 cited papers are equally distributed in its different authorship pattern; there is no association between authorship pattern and level of citations. The study also showed that there is an association between period of publication and level of citations.

3. Objectives

The objectives of the present study are (a) to trace the association among level of citations of the top 100 cited papers of the ***Scientometrics*** in WoS with its corresponding level of citations in Scopus and GS; (b) to determine the correlation among citations of the top 100 cited papers of ***Scientometrics*** in WoS with its corresponding citations in Scopus and GS.

4. Null Hypotheses

H01: There is no association among level of citations of the top 100 cited papers of ***Scientometrics*** in WoS and its corresponding level of citations in Scopus and GS.

H02: There is no relationship among citations of the top 100 cited papers of ***Scientometrics*** in WoS and its corresponding citations in Scopus and GS.

5. Method

The most cited papers in ***Scientometrics*** was identified by a search in the database of WoS, using the search term "SO=scientometrics". The search was conducted in WoS Core Collections on 18 June 2020. A total of 5627 items were retrieved and then ordered by the most cited first option. The top 100 articles were thus identified based on their citation counts. These articles were then cross-matched with data from Scopus and GS for its corresponding citation counts in those

databases. For every selected papers its title and citation counts in WoS, Scopus and GS were extracted and entered in the Excel datasheet for further analysis. The statistical program SPSS version 22.0 was used for analysis. The Shapiro-Wilk test was employed to detect departures from normality. The Spearman rank test was used to determine correlations between citation counts.

6. Analysis and Interpretation

The top 100 most cited articles received a total of 23,015 (WoS), 25,052 (Scopus) and 46,425 (GS) citations with a citation range of 112-1273 (WoS), 118-1476 (Scopus) and 174-2451 (GS) citations. Table 1 gives the top 10 most cited research papers of ***Scientometrics***in WoS and its corresponding citations in Scopus and GS. The first ranked most cited paper with 1273 (WoS), 1476 (Scopus) and 2451 (GS) was "Software survey: VOSviewer, a computer program for bibliometric mapping" authored by Van Eck and Waltman. The second ranked most cited article with 1054 (WoS), 1148 (Scopus) and 1535 (GS) citations was "Citation review of Lagergren kinetic rate equation on adsorption reactions" by Ho. The third ranked most cited article with 919 (WoS), 1036 (Scopus) and 2053 (GS) citations was "Theory and practice of the g-index" by Egghe. To consider an article as a "classic article", it must have at least 100 citations. Hence, the articles considered for the present study can be considered as classic articles because all of them received more than 100 citations in each database.

Table 1.
Top 10 Most Cited Research Papers of *Scientometrics* in WoS and its Corresponding Citations in Scopus and GS

Sl.No.	Title of the Article	Citation Count (WoS)	Citation Count (Scopus)	Citation Count (GS)
1	Van Eck, Nees Jan & Waltman, Ludo (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. 84(2). 523-538.	1273	1476	2451
2	Ho,Y. S. (2004). Citation review of Lagergren kinetic rate equation on adsorption reactions. 59(1). 171-177	1054	1148	1535
3	Egghe, Leo. (2006). Theory and practice of the g-index. 69(1). 131-152.	919	1036	2053
4	Boyak, K. W., Klavans, R &Borner, K. (2005). Mapping backbone of science. 64(3). 351-374.	446	485	895
5	Nederhof, A.J. (2006). Bibliographic			

	monitoring of research performance in the social sciences and the humanities: A review. 66(1). 81-100.	404	429	805
6	Van Raan, A.F.J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. 62(1). 133-143.	404	423	880
7	Fanelli, Daniele. (2012). Negative results are disappearing from most disciplines and countries. 90(3). 891-904.	401	426	806
8	Glanzel, W. (2001). National characteristics in international scientific co-authorship relations. 51(1). 69-115.	399	418	748
9	Callon, M., Courtial, J. P. &Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research- the case of polymer chemistry. 22(1). 155-205.	386	466	945
10	Mongeon, Philippe, Paul-Hus, Adele. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis.	379	417	672

6.1. Association among Level of Citations in WoS and its Corresponding Level of Citations in Scopus and GS.

The level of citations is arrived at transforming the citation counts in to a new variable by finding out each database citations' quartile values. Accordingly the values below and up to first quartile are designated as low level citations, values above and equal to third quartile are high level citations and values in between first and third quartiles are moderate level citations.

Table 2
Level of WoS Citations vs Level of Scopus Citations

Level of WoS Citations	**Level of Scopus Citations**			**Total**	**Chi-square**	**p value**
	Low	**Moderate**	**High**			
Low	20 (80%)	5 (20%)	0 (0%)	25 (100%)		

	[76.9%]	[10.4%]	[0%]	[25%]		
Moderate	6 (12%) [23.1%]	42 (84%) [87.5%]	2 (4%) [7.7%]	50 (100%) [50%]	128.897	0.000
High	0 (0%) [0%]	1 (4%) [2.1%]	24 (96%) [92.3%]	25 (100%) [24.5%]		
Total	26 (26%) [100%]	48 (48%) [100%]	26 (26%) [100%]	100 (100%) [100%]		

1) The value within ( ) refers to row percentage
2) The value within [ ] refers to column percentage

Table 3
Level of WoS Citations vs Level of GS Citations

Level of WoS Citations	**Level of GS Citations**			**Total**	**Chi-square**	**p value**
	Low	**Moderate**	**High**			
Low	17 (68%) [68%]	8 (32%) [16%]	0 (0%) [0%]	25 (100%) [25%]	105.760	0.000
Moderate	8 (16%) [32%]	40 (80%) [80%]	2 (4%) [8%]	50 (100%) [50%]		
High	0 (0%) [0%]	2 (8%) [4%]	23 (92%) [92%]	25 (100%) [24.5%]		
Total	25 (25%) [100%]	50 (50%) [100%]	25 (25%) [100%]	100 (100%) [100%]		

1) The value within ( ) refers to row percentage
2) The value within [ ] refers to column percentage

Table 4
Level of Scopus Citations vs Level of GS Citations

Level of Scopus Citations	Level of GS Citations			Total	Chi-square	p value
	Low	Moderate	High			
Low	21 (80.8%) [84%]	5 (19.2%) [10%]	0 (0%) [0%]	26 (100%) [26%]	105.760	0.000
Moderate	4 (8.3%) [16%]	42 (87.5%) [84%]	2 (4.2%) [8%]	48 (100%) [48%]		
High	0 (0%) [0%]	3 (11.5%) [6%]	23 (88.5%) [92%]	26 (100%) [26%]		
Total	25 (25%) [100%]	50 (50%) [100%]	25 (25%) [100%]	100 (100%) [100%]		

1) The value within ( ) refers to row percentage
2) The value within [ ] refers to column percentage

Based on row percentage, majority of WoS low level citations (80%) are at Scopus low level citations, only 20% at moderate level and no percentage of low level citations of WoS is included in the high level citations of Scopus. Similar to the case above, majority of the WoS moderate level citations (84%) are at moderate level, 12% at low level, and only 4% at high level of citations of Scopus. Majority (96%) of the high level citations of WoS are at high level, only 4% at moderate level and no percentage at low level of Scopus citations (Table 2).

Majority of WoS low level citations (68%) are at GS low level citations, only 32% at moderate level and no percentage of low level citations at the high level citations of GS. Majority of the WoS moderate level citations (80%) are at moderate level, 16% at low level, and only 4% at high level citations of GS. Majority (92%) of the high level citations of WoS are at high level of GS citations and only 8% at moderate level and no percentage at low level of GS citations (Table 3).

Majority of Scopus low level citations (80.8%) are at GS low level citations, only 19.2% at moderate level and no percentage of low level citations at the high level citations of GS. Majority of

the Scopus moderate level citations (87.5%) are at moderate level, 8.3% at low level, and only 4.2% at high level citations of GS. Majority (88.5%) of the high level citations of Scopus are at high level of GS citations and only 11.5% at moderate level and no percentage at low level of GS citations (Table 4).

Since $p<0.01$ in all the cases, the null hypothesis is rejected at 1% level of significance. Hence it is concluded that there is significant association among level of citations of top 100 cited papers of ***Scientometrics*** in Web of Science and its corresponding level of citations in Scopus and Google Scholar.

6.2.Correlation among Citations in WoS, Scopus and GS

The normality of citation counts in different databases was checked using the Shapiro-Wilk test and it was found that all the citation counts do not follow normality. Since the distribution of the citation counts are not normal, the correlation was found out using the non-parametric Spearman rank test. The test result is given in Table 5.

Table 5
Correlations

			WoS Citations	Scopus Citations	GS Citations
Spearman's rho	WoS Citations	Correlation Coefficient	1.000	.963**	.923**
		Sig. (2-tailed)	.	.000	.000
		N	100	100	100
	Scopus Citations	Correlation Coefficient	.963**	1.000	.941**
		Sig. (2-tailed)	.000	.	.000
		N	100	100	100
	GS Citations	Correlation Coefficient	.923**	.941**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	100	100	100

**. Correlation is significant at the 0.01 level (2-tailed).

The correlation coefficient between WoS citations and Scopus citations is 0.963 (Spearman's rank correlation coefficient) which indicates 96.3% positive relationship between WoS citations and Scopus citations at 1% level. Spearman's rho statistical correlation revealed a strong

(r>0.6) correlation between WoS citations and Scopus citations. Figure 1 is a scatter plot examining the relationship between WoS citations and Scopus citations.

A strong correlation could be found between the numbers of citations in WoS and GS (r = 0.923, p = 0). The clear linear correlation between the numbers of citations obtained by the papers in both the databases can be appreciated in the scatterplot (Fig. 2). Similarly Spearman's rho statistical correlation revealed a high correlation between Scopus and GS citations (coefficient value 0.941). Figure 3 is the scatterplot depicting the same.

The null hypothesis was 'there is no relationship among citations of the top 100 cited papers of ***Scientometrics*** in Web of Science and its corresponding citations in Scopus and Google Scholar'. As not hypothesised, there exists a strong positive correlation among citation counts in WoS, Scopus and Google Scholar. So the null hypothesis is rejected.

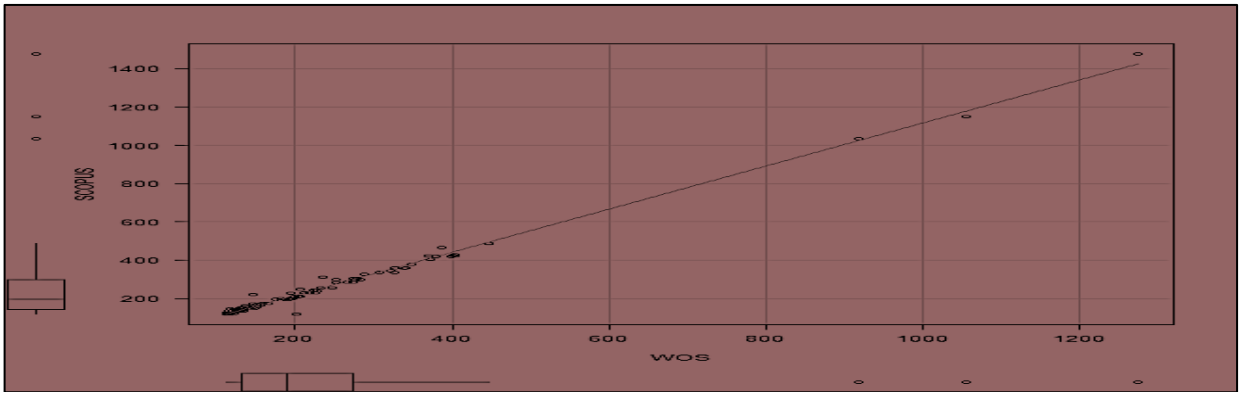

Figure 1.
Scatter plot examining the relationship between WoS citations and Scopus citations; the line represents the trend line

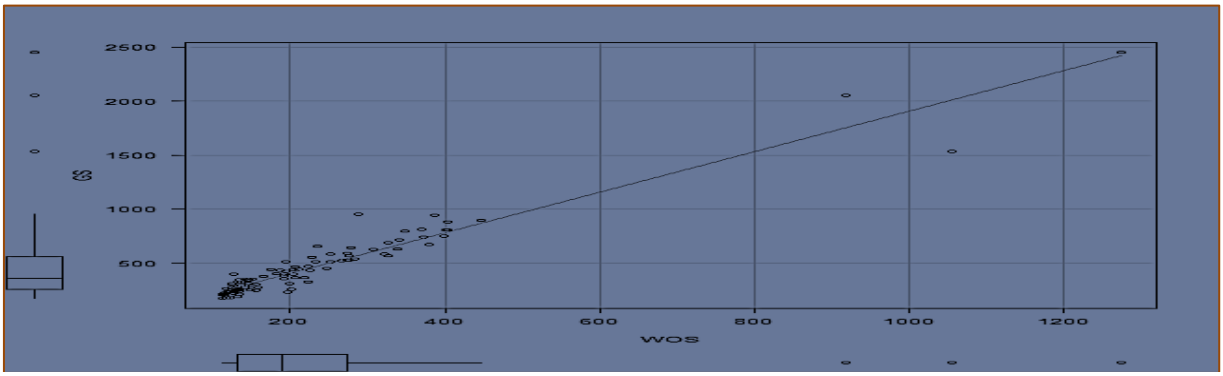

Figure 2.
Scatter plot examining the relationship between WoS citations and Google Scholar citations; the line represents the trend line

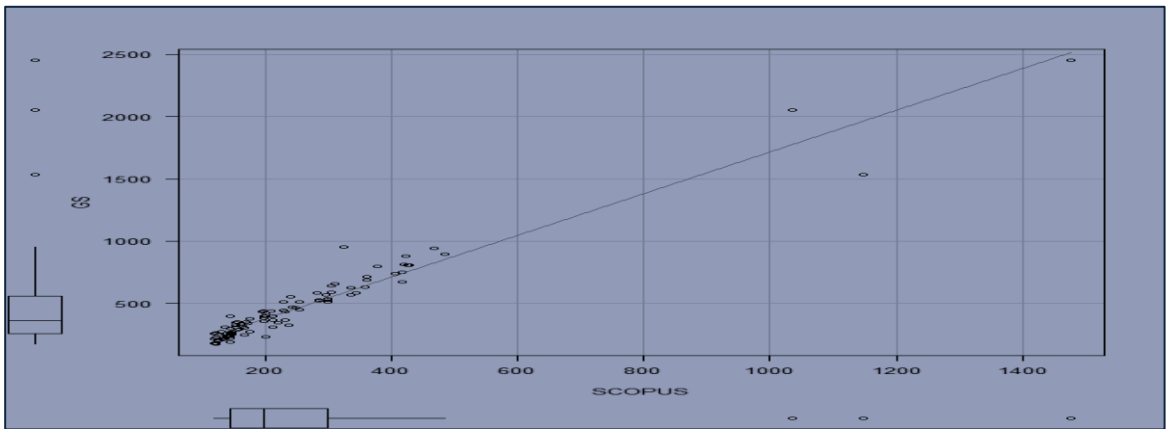

Figure 3.
Scatter plot examining the relationship between Scopus citations and Google Scholar citations; the line represents the trend line

7. Discussion

The aim of this study was to identify the association and relationship among the citations of the top 100 most cited papers of ***Scientometrics*** in WoS and its corresponding citations in Scopus and GS. The study results shows that there is significant association among level of citations of the top 100 cited papers of ***Scientometrics*** in WoS and its corresponding level of citations in Scopus and GS. It shows that even though the citation counts are indifferent in three databases for each paper, the level of citations is almost same in all the three databases for the top most cited papers.

The study also showed that there exists a strong positive correlation among citations of top 100 cited papers of ***Scientometrics*** in WoS and its corresponding citations in Scopus and GS. The citations in Scopus are slightly higher than the WoS citations and citations in GS are comparatively higher than the citations in both WoS and Scopus databases. Both Scopus and Web of Science index primarily refereed journal articles whereas GS indexes several refereed and non-refereed types of documents in addition to journal articles. GS, in contrast to WoS and Scopus, also covers material like preprints, course notes, assignments, word documents, technical reports, Bachelor's, Master's and Doctoral theses and dissertations, abstracts, conference proceedings volumes, newsletters, product brochures, blogs, (Yang & Meho, 2006) and even predatory journal papers, reviews and Twitter feeds. In fact GS has helped some such journal articles gain citations, and though the journals are predatory in nature, occasional papers sometimes find resonance with the scientific community across the world. This is especially true in the case of studies that are of very local

interest, and which often do not make it to the regular peer-reviewed journals. Predatory journal papers also sometimes get indexed in major databases like PubMed (Cortegiani et al. 2019). GS has the most extensive coverage of conference proceedings and non-English language journals. The inclusion of citations from non-English speaking nations has been viewed as one of the Google Scholar's advantages (Martel, 2009). This coverage will result in generating high citation counts in GS. Though WoS and Scopus include some proceedings volumes and books, they mostly cover journal articles.

GS and Scopus cover journals published outside the USA than does WoS. WoS covers only "high-influence" publications. Scopus and WoS databases are each grounded by certain principles to cover selective important journals in all knowledge fields. The information generated by these two databases can provide pointers to the journals that cover relevant and current research in an area and which would be influential in shaping future research endeavours.Scopus citations are slightly higher than WoS citations. This is mainly due to the fact that there are more LIS sources in Scopus which generate higher citations in this database and confirms that Scopus is the world's largest multidisciplinary database in terms of more recent scholarly literature. More over top impact LIS journals could be identified in Scopus, which were not reported in WoS (Abrizah et al. 2012; Moya-Anegon et al. 2007; Leydesdorff et al. 2010). But another study conducted by Meho and Sugimoto (2009) indicated that when assessing the smaller citing entities such as journals, institutional and conference proceedings, both databases produce significantly different results. However when assessing larger citing entities such as research domains and countries, they produce similar scholarly impact. Martín-Martín et al. (2020) did a document-level comparison using Scopus, WoS, Dimensions, OpenCitations, Microsoft Academic, and Google Scholar by selecting highly cited documents, and analysing the overlap between the databases in terms of documents that cite the selected highly cited documents. They found that each had its own unique advantages.

8. Conclusion

Citation counts are the total number of citations an article receives. This type of service is offered by citation databases like WoS, Scopus and GS. In general, the higher the number of citations, the greater the perception of quality for that article. Majority of the highly cited articles of ***Scientometrics*** have received a good number of citations in these three databases. WoS had been the sole tool for citation analysis until the creation of Scopus and GS. GS is a feasible, free-of-cost

alternative to the well-known citation databases WoS and Scopus. However, the low data quality found in GS raises questions about its suitability for research evaluation. Thus, WoS and Scopus remain today the main sources for citation data. Moreover indisciplinary coverage of these databases represents a significant strength for the study and comparison of different scientific fields. Scopus and GS have similarity to WoS. These databases can be used for citation searching and also for bibliographic searching. Scopus and GS are the major competitors to WoS in the field of citation analysis and bibliometrics (Yang & Meho, 2006). ***Scientometrics*** being one of the top influential scientific journal has high visibility and citation impact in theses databases. For identifying citations to an article in ***Scientometrics***, WoS should not be used alone. Scopus and GS should also be searched as it helps in identifying considerable number of citations not found in WoS. Thus the coverage of these databases determines the citation counts in general.

References

1. Abrizah, A., Zainab, A.N., Kiran, K. & Raj, R. G. (2012). LIS journals scientific impact and subject categorization: A comparison between Web of Science and Scopus. *Scientometrics*, 94, 721-740. https://doi.org 10.1007/s1192-012-0813-7
2. Cortegiani, A., Sanfilippo, F., Tramarin, J, &Giarratano, A. (2019). Predatory open-access publishing in critical care medicine, *Journal of Critical Care,* 50, 247-249. https://doi.org/10.1016/j.jcrc.2018.12.016.
3. Leydesdorff, L., Moya-Anegon, F. & Guerrero-Bote, V. P. (2010). Journal maps on the basis of Scopus data: A comparison with the Journal Citation Reports of the ISI. *Journal of American Society for Information Science and Technology*, 61(2), 352-369.
4. Martell, C. (2009). A citation analysis of College and Research Libraries comparing Yahoo, Google, Google Scholar, and ISI Web of Knowledge with implications for promotion and tenure. *College & Research Libraries*. 70(5), 460-472.
5. Martín-Martín, A., Thelwall, M., Orduna-Malea, E., &López-Cózar, E.D. (2020) Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. arXiv:2004.14329v1.

6. Meho, L.I. & Sugimoto, C. R. (2009). Assessing the scholarly impact of information studies: A tale of two citation databases- Scopus and Web of Science. *Journal of American Society for Information Science and Technology*, 60(12), 2499-2508. http://dx.doi.org/10.1002/asi.21165
7. Moya-Anegon, F. et al. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 73(1), 53-78.
8. Renjith, V. R. (2018). Visibility and citation impact of Indian journals in Library and Information Science: A study based on Google Scholar using 'Publish or Perish' software. *ILIS Journal of Librarianship and Informatics*. 1(2), 176-183.
9. Renjith, V. R. (2019). Authorship pattern and citation level of i10 cited research papers of DESIDOC Journal of Library and Information Technology (DLIT) in Google Scholar. *Library Philosophy and Practice (e-journal)*. 2679. http://digitalcommons.unl.edu/libphilprac/2679.
10. Thelwall, M & Kousha, K. (2015). Web indicators for research evaluation. Part1: Citations and links to academic articles from the web. *EI Profesional de La Information*, 24(5), 1699-2407.
11. http://www.springer.com/journal/11192
12. http://www.scimagojr.com
13. Yang, K &Meho, L. I. (2006). Citation analysis: A comparison of Google Scholar, Scopus and Web of Science. *Proceedings of the American Society for Information Science and Technology*, 43(1), 1-15. https://doi:10.1002/meet.14504301185