Web Citations vs. ISI and Scopus Citations: Can Web be Better Impact Indicator for Persian Medical Science Research

Mahshid Abdoli[1]
Kayvan Kousha[2]

Abstract

Many medical science research articles are published in Persian language journals which are not indexed by the ISI and Scopus citation databases. Thus, they have been invisible for previous citation analysis studies. Google Scholar also contains citation information, but includes publications from different types of Web documents. We selected 56 Persian language medical science journals and searched their exact English titles in the ISI Web of Science, Scopus and Google Scholar citation databases. We compared number, mean and median of the cited items to the selected journals in order to examine the extent of differences between the three citation databases and if the Web extracted citations would be better indicator for citation tracking or impact assessment. The result showed that the number, mean and median of Google Scholar citations is two times more than both ISI and Scopus citations. This suggests that Google Scholar has a good coverage for citation counting of the Persian language medical science research. We found stronger correlation between Scopus and Google Scholar citation than ISI and Google Scholar citations for the studied journals, indicating that both Scopus and Google Scholar citation counts measure the same aspect of formal patterns from a wide range of non-journal scholarly documents (i.e., conference papers) which are not covered by ISI databases. Finally, although we found practical methods

1. Faculty Member and Head of the Periodical Department, National Library and Archives of Iran
mahshid_abdoli@hotmail.com
2. Department of Library and Information Science, University of Tehran, Iran, kkoosha@ut.ac.ir

for citation counting of the Persian language scholarly journals, we should cautiously use Web extracted citation particularly for research evaluation.

Keywords: Web citation, ISI, Scopus, Google Scholar, Persian medical journals

Introduction

For a long time bibliometrics and ISI[3] citations have been widely used for research evaluation and impact assessment of scientific works in many medical science disciplines (**3**: 281-300; **17**). However, there has been discussion about using ISI citation data for research evaluation especially for non-English language/non-ISI journals (i.e., Persian academic journals). For practical reasons, bibliographic databases can only contain a subset of the scientific literature. In other words, the huge volume of national Persian medical science research in Iran is not indexed by ISI or other citation indexes like Scopus (www.scopus.com). Recently, Google Scholar (http://scholar.google.com) is the largest new multi-disciplinary tool that can be used for accessing and measuring 'hidden' citations, especially from non-journal Web documents which are not traced by conventional bibliographic and citation databases like ISI or Scopus. Google Scholar contains citation information from many publishers, but does not apply any quality control procedure for its collection of journal publications and includes substantial coverage of non-journal documents, such as conference papers, theses, books, research/technical reports, and preprint repositories (**6**).

Many have discussed the advantages of automatic Web-based citation indexes (i.e., in Google Scholar) which may provide less biased citation searching in broader disciplines and types of documents (**16**: 67-71; **8**). Since "the vast majority of authors" would willing to deposit copies of their articles in an institutional or subject-based repository (**20**) and over 90% of journals "have given their official green light to author self-archiving" (**7**), an increasingly amount of citation information from Web documents (open access or non/open aces) seems to be available and it is reasonable to use Web citation extraction techniques for locating relevant documents.

In this paper we compare citation counts from three citation databases including ISI Web of Science,

3. ISI is now known as Thomson Scientific

Scopus and Google Scholar in order to examine which database retrieve more citations to 56 Persian language medical journals. The main question of this study is to explore quantitative aspects of Google Scholar citations vs. ISI and Scopus citations to Persian language medical journals. Hence, we intend to compare the number, mean and median of ISI, Scopus and Google Scholar citations and to explore the extent of differences between three citation databases. For this purpose, we searched the journal names as required (described in method). Several studies used Web citations to determine the potential use of Web sources for impact assessment and research evaluation (**12**: 1055-1065; **13**: 1-14; **14**; **21**: 1313-1324; **22**: 1075-1087). Yet, there is less evidence on online impact measures and Web extracted citation indicators in medical sciences. This approach helps us to understand the extent of citations available on the Web and their functionality as a useful or trivial source for medical sciences research evaluation.

Related studies

Although recent studies have found that there is some commonality between traditional and Web-extracted citations from commercial search engines, no studies have been conducted on Persian language medical journals comparing conventional citations (e.g. ISI or Scopus citations) with citations from Web-based citation indexes (**6**). Goodrum, et al. (2001), for instance, compared ISI and CiteSeer citations from online computer science papers. They found more citations from conference papers than ISI articles, indicating the importance of conference papers among computer professional for disseminating research results (**5**: 666-676). Zhao & Logan (2002) compared online and offline citation patterns for the XML research area, finding more citations from CiteSeer than the ISI counterpart (**24**: 449-472). Zhao (2005) found a 10% overlap between citing papers by CiteSeer and the ISI in the XML research field, indicating that XML articles published in journals were not mainly open access and papers published on the Web were not well represented in journals indexed by the ISI (**23**: 1403-1418).

With advent of Google Scholar several reviews and papers have discussed its pitfalls and capabilities (**4**; **9**; **10**; **11**; **18**), but few articles have compared Google Scholar with ISI citation data. Bauer & Backlash

(2005), for instance, compared the citation counts from the ISI Web of Science, Elsevier's Scopus, and Google Scholar to articles published in 1985 and 2000 in just the *Journal of the American Society for Information Science and Technology (JASIST)*, finding significantly higher citation counts from Google Scholar than either the Web of Science or Scopus for the year 2000 (**1**). Pauly & Stergiou (2005) also compared citations from the ISI and Google Scholar to 99 papers in 11 disciplines as well as 15 highly-cited articles, finding nearly equal patterns for articles published after 1990 (**19**: 33-35). Belew (2005) selected six academics at random and compared citations to publications by these authors indexed by the ISI with those reported by Google Scholar, finding a small overlap between two databases (**2**).

Kousha & Thelwall (2007) examined correlations between ISI and Google Scholar citations to a sample of 1,650 articles from 108 Open Access journals published in 2001 in four science and four social science disciplines, finding significant relationship in all cases. They also examined the overlap between ISI citations and Google Scholar citations in four science disciplines, finding that the overall relative overlap percentage for the ISI (overlapping ISI citations with Google Scholar citations divided by total ISI citations) was 57%. They also found that the percentage for ISI citations was about 12% and for Google Scholar 22% within a four-month period (**12**: 1055-1065). In another study, Kousha & Thelwall (2007) examined the publication type of sources of citations from Google Scholar that were not present in the ISI index in four science disciplines, finding that in computer science conference or workshop papers (43.2%) were the most common type. The results indicate that the type of unique citations from web documents varies significantly between hard sciences (**15**).

Research questions

We addressed two questions below to compare ISI, Scopus and Google Scholar citation counts at the individual journal level. We use these questions to validate web as useful or trivial source of information for research evaluation.

1. What is the number, mean and median of Google Scholar, ISI and Scopus citations to Parisian language academic medical journals?

2. Can web extracted citations be

better quantitative value for evaluation and impact assessment of Persian medical research?

Methods

Journal Selection

For the purpose of this study, medical journals that had been published by academic institution or research center in Persian language. We used several medical directories to locate the exact English journal names as appeared in the title page. If a journal name was too general we ignored it to avoid false matches during search process in three citation databases. In fact some English names for Persian journals were very general and there are no practical methods for retrieving correct citation counts to this kind of journal. Ultimately, our selected journals for citation analysis included 56 academic journals (Table 2).

ISI, Scopus and Google Scholar Citation Counts

For ISI citation counts, we searched the English titles of the 56 Persian journals as appeared in the original journal title page in the "Cited Reference Search" field in the ISI Web of Science to find the possible number of citations received. Searches conducted during a week on August 2007. Since all Persian language journals were not indexed in the ISI; the aim was to find the number of citations to them in the reference sections of journals indexed by the ISI. A limitation in using this method related to different names (abbreviations) of cited sources in the ISI databases entered in the citation information. For example, several different abbreviations were used for *IRANIAN JOURNAL OF NUCLEAR MEDICINE* including:

IRAN J OF NUCLEAR MEDICINE
IRAN J OF NUCL MED
IRANIAN J OF NUCLEAR MEDICINE

Consequently, using truncation (*) other possible abbreviations were searched and through a manual checking process unrelated names or abbreviations were excluded. In order to prevent possible similarity between abbreviations for different journals in different fields, we manually checked the results. For Scopus citation counts, we searched the exact journal names or possible abbreviations in the reference field through the basic Scopus search page. We sometimes conducted several searches for the same journal articles in order to retrieve possible journal abbreviations in the references. For Google Scholar citations, we searched

journal titles in the option "return articles published in" in advanced search page. Again we conducted several searches to get different journal abbreviation.

We then recorded the total number of ISI, Scopus and Google Scholar citing sources, as shown by each databases. Note that we manually checked the search results against the journal names to avoid false matches and also removed any duplicated citing documents. Ultimately, we have citation counts for each 56 Persian language medical journals from three different citation databases.

Results

Comparing descriptive statistic for citations

Table 1 shows the number, mean and median of citation to 56 Persian language medical journals from ISI, Scopus and Google Scholar citation databases. It shows that the number, mean and median of Google Scholar citations is much higher than both ISI and Scopus citations. Moreover, the mean and median number of ISI citations is lower than Scopus citation. The results suggest that Google Scholar have particularly good coverage of sources for citations in the medical sciences targeting Persian language journal articles. In fact, there is wide range of non-ISI citing sources; especially from non-journal documents (i.e., conference papers, research reports, e-prints, dissertations) are accessible by Google Scholar. Some of these documents might never have been indexed by ISI and Scopus, so they would be 'invisible' for impact assessment and measuring performance of higher education.

Table 2 gives citation statistic for each 56 studies journals. As shown below, many journals have received more citations on the web based upon

Table 1. Number, mean and median of ISI, Scopus and Google Scholar Citations for Persian language medical journals

	ISI citation	Scopus Citation	Google Scholar Citation
Number of Citation (% of ISI citation)	99 (100%)	130 (131%)	301 (304%)
Mean	1.76	2.32	5.37
Median	0.00	2.00	4.00

the Google Scholar search results. For instance, *Journal of the Shaheed Beheshti University of Medical Sciences and Health Services* has received 19 citations on the web. However, 3 and 5 citations have recorded in ISI and Scopus receptively. We can almost identify similar citation pattern in three databases, indicating that "**web citation**" is better quantitative indicator for impact assessment and measuring research communication for Persian language medical journals. Nevertheless, few journal titles have received more ISI than Google Scholar (i.e., *Kowsar Medical Journal*).

Correlation between the traditional and Web-extracted citations

In order to examine the relationship between traditional citations (ISI and Scopus) and Web-based citation patterns (Google Scholar) we calculate the correlation tests (Table 3) at the journal level as the data points. Spearman correlation tests were preformed since we found that the frequency distributions were skewed. As shown in Table 3, there is a significant correlation (r=0.408**) between the ISI and Google Scholar citation counts for 56 journals as the data points ($p < 0.01$). We also found relatively higher correlation between Scopus and Google Scholar citations (r=718**; $p < 0.01$) than ISI and Scopus citation (r=0.473**; $p < 0.01$), indicating that both Scopus and Google Scholar citation counts measure the

Table 2. ISI, Scopus and Google Scholar citation counts for 56 Persian journal titles

Journal titles (in English)	GS	ISI	Scopus
Journal of the Shaheed Beheshti University of Medical and Health Services	19	3	5
Quarterly Journal of Andeesheh VA Raftar	18	0	6
Hakim Research Journal	16	11	8
Journal of Kerman University of Medical Sciences	14	11	9
Iranian Journal of Diabetes and Lipid Disorders	14	8	4
Armaghane-Danesh, Journal of Yasuj University of Medical Sciences	13	0	2
Kowsar Medical Journal	13	20	5
Feyz, Kashan University of Medical Sciences & Health Services	12	6	9
The Iranian Journal of Nuclear Medicine	12	5	4
Journal of Hormozgan University of Medical Sciences	11	0	0
Journal of Isfahan Medical School	11	0	3
Medical Journal of Tabriz University of Medical Sciences & Health Services	10	2	3
Iranian South Medical Journal	9	0	2
The Journal of Qazvin University of Medical Sciences & Health Services	9	0	5

Journal of Shahid Sadoughi University of Medical Sciences and Health Services	8	1	5
Journal of Dentistry Tehran University of Medical Sciences	8	0	0
Modarres Journal of Medical Sciences	7	3	3
Daneshvar, Scientific-Research Journal of Shahed University	6	2	1
Iranian Journal of Infectious Diseases & Tropical Medicine	6	0	4
Journal of Mazandaran University of Medical Sciences	6	1	3
Payesh, Journal of The Iranian Institute for Health Sciences Research	5	4	7
The Iranian Journal of Urology	5	0	1
Shahrekord University of the Iranian of Medical Sciences	5	2	0
Scientific Journal of Hamadan University of Medical Sciences & Health Services	5	0	2
Pejournal Quarterl Reserch Journal	4	2	4
The Journal of Tehran Faculty of Medicine	4	0	4
Journal of Medical Faculty Guilan University of Medical Sciences	4	0	3
Journal of Iran University of Medical Sciences	4	0	0
Journal of Babol University of Medical Sciences	4	0	5
Tabib-E-Shargh, Journal of Zahedan University of Medical Sciences and Health Services	3	0	5
Blood, The Scientific Journal of Iranian Blood Transfusion Organization-Research Center	3	0	3
Shahid Beheshti Medical Sciences University Journal of The Dental School	3	1	2
Journal of Gorgan University of Medical Sciences	3	1	2
Journal of Zanjan University of Medical Sciences & Health Services	3	2	2
Yakhteh Medical Journal	3	0	0
Asrar, Journal of Sabzevar School of Medical Scinces	2	0	0
Behood, The Scientific Quarterly	2	1	1
Koomesh, Journal of Semnan University of Medical Sciences	2	0	0
The Journal of the Iranian Society of Ophthalmology	2	0	0
The Iranian Journal of Obstetrics, Gynecology and Infertility	2	1	2
Journal of Birjand University of Medical Sciences	2	1	3
Scientific Journal of Kurdistan University of Medical Sciences	2	0	1
Iranian Journal of Medical Physics	2	1	1
Strides in Development of Medical Education, Journal of Medical Education Development Center of Kerman University of Medical Sciences	1	0	1
Urmia Medical Journal	1	1	0
Medical Journal of Mashad University of Medical Sciences	1	0	0
Journal of Mashad Dental School	1	0	0
The Iranian Journal of Otorhinolaryngology	1	0	0
Iranian Journal of Dermatology	0	9	0
Journal of Urmia Unrsing and Midwifery Faculty	0	0	0
Pahavard Danesh, Journal of Arak University of Medical Sciences	0	0	0
Iranian Journal of Pathology	0	0	0
Journal of Iranian Society of Anaesthesiology and Intensive Care	0	0	0
Iranian Journal of Pediatrics	0	0	0
The Journal of Islamic Dental Association of Iran	0	0	0
Journal of Iranian Anatomical Sciences	0	0	0

Table 3. Correlations between ISI, Scopus and Google Scholar citation counts to 56 Persian language medical journals

ISI and Google Scholar citations	Scopus and Google Scholar	ISI and Scopus
0.408**	718**	0.473**

same aspect of formal patterns from a wide range of scholarly documents (i.e., conference papers) which are not covered by ISI databases.

Conclusions and Discussion

In answer to first question, we found that the Google Scholar citation is much higher than both ISI and Scopus. The results suggest that web contains more citation data targeting Persian medical science research. The main cause may be that Google Scholar contains a wide range of citation data (conference papers, research reports, e-prints, dissertations) which aren't covered by ISI and Scopus. Moreover, we relatively found more citations from Scopus than ISI. The reason might be that Scopus also cover citations from conferences and workshops.

Although it is reasonable to use citation statistic derived from Google Scholar (cautiously) for impact calculations, when ISI or Scopus citation data is not available for Persian language medical research, ISI citations are from highest impact scientific research. In fact Google Scholar includes a less quality controlled collection of publications from different types of web documents.

A major practical problem of our method was related to different names (abbreviations) of cited sources in the citation databases, although we tired to cover possible abbreviations to gain the most useful results. Since, a considerable body of Persian language academic journals are not indexed by conventional citation databases (i.e., ISI and Scopus) our method may shed light on application of web citations for research communication.

It can be expected that a considerable number of researchers in non-English language countries would be willing to deposit preprints/postprints of their papers online, then web citations can be more important for citation tracking and impact assessment.

References

1. Bauer, K.; Backkalbasi, N. "An

examination of citation counts in a new scholarly communication environment". ***D-Lib Magazine***, Vol.11, No.9 (2005). [on-line]. Available: http://www.dlib.org/dlib/september05/bauer/09bauer.html. [23Dec.2005].

2. Belew, R. "Scientific impact quantity and quality: analysis of two sources of bibliographic data". 2005. [on-line]. http://arxiv.org/abs/cs.IR/0504036. [3May 2006].

3. Cole, J. "A short history of the use of citations as a measure of the impact of scientific and scholarly work". In B. Cronin & H. B. Atkins (Eds.), ***The web of knowledge: A festschrift in honor of Eugene Garfield***. Medford, NJ: Information Today Inc, 2000, PP.281-300.

4. Friend. F. "Google Scholar: Potentially good for users of academic information". ***The Journal of Electronic Publishing***, 2006. [on-line]. Available: http:// hdl.handle.net/2027/spo.3336451.0009.105. [28 Apr.2006].

5. Goodrum, A.A. … [et al]. "Scholarly publishing in the Internet age: a citation analysis of computer science literature". ***Information Processing & Management***, Vol.37, No.5 (2001): 661-676.

6. "Google Scholar". 2007. [on-line]. http://scholar.google.com/ scholar/ about.html. [12Jun.2007].

7. Harnad, S. … [et al]. "The access/impact problem and the green and gold roads to open access". ***Serials Review***, Vol.30, No.4 (2004). [on-line]. Available: http://eprints.ecs.soton.ac.uk/10209/01/impact.html. [5 May 2006].

8. Hitchcock, S. … [et al]. "Open citation linking: the way forward". ***D-Lib Magazine***, Vol.8, No.10 (2002). [on-line]. Available: http://www.dlib.org/dlib/october02/ hitchcock/10hitchcock.html. [10Jan.2006].

9. Jacso, P. "Google Scholar: the pros and the cons". ***Online Information Review***, Vol.29, No.2 (2005): 208-214.

10. Ibid. "Google Scholar Beta". Péter's Digital Reference Shelf, 2004. [on-line]. Available: http://snipurl.com/dwco [10Jan.2004].

11. Ibid. "As we may search: comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases". ***Current Science***, Vol.89, No.9 (2005): 1537-1547. [on-line]. Available: http://www.ias.ac.in/currsci/ nov102005/1537.pdf. [28Apr.2006].

12. Kousha, K.; Thelwall, M. "Google Scholar citations and Google Web/URL citations: a multi-discipline exploratory analysis". ***Journal of the American Society for Information Science and Technology***, Vol.58, No.7 (2007): 1055-1065.

13. Ibid. "How is science cited on the

web? a classification of Google unique web citations". ***Journal of the American Society for Information Science and Technology***, Vol.58, No.11 (2007):1–14.

14. Ibid. "The online impact of open access social science research". ***Library and Information Science Research***. 2007. [on-line]. Available: http://www.koosha.tripod. com/impact_of_ss_res.doc.

15. Ibid. "Sources of Google Scholar citations outside the science citation index: a comparison between four science disciplines". ***Scientometrics***, Vol.74, No.8 (2007). [on-line]. Available: http://www.koosha.tripod.com/GSoutsideSSC.doc.

16. Lawrence, S.; Giles, C. L.; Bollacker, K. "Digital libraries and autonomous citation indexing". ***IEEE Computer***, Vol.32, No.6 (1999): 67-71. [on-line]. Available: http://csdl.computer.org/dl/mags/co/1999/06/r6067.pdf. [1Apr.2006].

17. Moed, H. F. ***Citation analysis in research evaluation***. New York: Springer, 2005.

18. Notess, G. R. "Scholarly Web Searching: Google Scholar and Scirus". ***Online***, Vol.29, No.4 (2005). [on-line]. Available: http://www.infotoday.com/Online/jul05/OnTheNet.shtml. [6May 2006].

19. Pauly, D.; Stergiou, K. "Equivalence of results from two citation Thomson ISI's Citation Index and Google's Scholar service". ***Ethics in Science and Environmental Politics***, (Dec.2005): 33-35. [on-line]. Available: http://www.int-res.com/articles/esep/2005/E65.pdf. [25Apr.2006].

20. Swan, A.; Brown, S. "Open access self-archiving: an author study". 2005. [on-line]. Available: http://eprints.ecs.soton.ac.uk/10999/01/jisc2.pdf. [20Apr.2006].

21. Vaughan, L.; Shaw, D. "Bibliographic and Web citations: what is the difference?". ***Journal of the American Society for Information Science and Technology***, Vol.54, No.14 (2003):1313-1324.

22. Ibid. "Web citation data for impact assessment: a comparison of four science disciplines". ***Journal of the American Society for Information Science and Technology***, Vol.56, No.10 (2005): 1075–1087.

23. Zhao,D. "Challenges of scholarly publications on the Web to the evaluation of science A comparison of author visibility on the Web and in print journals". ***Information Processing and Management***, Vol.41, No.6 (2005): 1403-1418.

24. Zhao, D.; Logan, E. "Citation analysis using scientific publications on the Web as data source: a case study in the XML research area". ***Scientometrics***, Vol.54, No.3 (2002): 449-472.