

doi: 10.3325/cmj.2010.51.99

Croatian Medical Journal Citation Score in Web of Science, Scopus, and Google Scholar

Marijan Šember¹, Ana
Utrobičić², Jelka Petrak¹

¹Central Medical Library, University
of Zagreb, School of Medicine,
Zagreb, Croatia

²Central Medical Library, University
of Split, School of Medicine, Split,
Croatia

Aim To analyze the 2007 citation count of articles published by the *Croatian Medical Journal* in 2005-2006 based on data from the Web of Science, Scopus, and Google Scholar.

Methods Web of Science and Scopus were searched for the articles published in 2005-2006. As all articles returned by Scopus were included in Web of Science, the latter list was the sample for further analysis. Total citation counts for each article on the list were retrieved from Web of Science, Scopus, and Google Scholar. The overlap and unique citations were compared and analyzed. Proportions were compared using χ^2 -test.

Results Google Scholar returned the greatest proportion of articles with citations (45%), followed by Scopus (42%), and Web of Science (38%). Almost a half (49%) of articles had no citations and 11% had an equal number of identical citations in all 3 databases. The greatest overlap was found between Web of Science and Scopus (54%), followed by Scopus and Google Scholar (51%), and Web of Science and Google Scholar (44%). The greatest number of unique citations was found by Google Scholar ($n=86$). The majority of these citations (64%) came from journals, followed by books and PhD theses. Approximately 55% of all citing documents were full-text resources in open access. The language of citing documents was mostly English, but as many as 25 citing documents (29%) were in Chinese.

Conclusion Google Scholar shares a total of 42% citations returned by two others, more influential, bibliographic resources. The list of unique citations in Google Scholar is predominantly journal based, but these journals are mainly of local character. Citations received by internationally recognized medical journals are crucial for increasing the visibility of small medical journals but Google Scholar may serve as an alternative bibliometric tool for an orientational citation insight.

Received: June 24, 2009

Accepted: March 3, 2010

Correspondence to:

Marijan Šember
Central Medical Library
University of Zagreb School of
Medicine
Šalata 3
10000 Zagreb, Croatia
seember@mef.hr

Despite many controversies, bibliometric scores are widely used in measuring research output and impact at individual and collective levels, as well as in measuring the performance of scientific journals (1). The citation rate of scientific publications has been measured for many years using the citation indexes of the Institute for Scientific Information (now Thomson Reuters) (2). In general medical literature, virtually all citation analysis studies have exclusively used these databases (3). Scholarly communication has been changing rapidly and new means of sharing and archiving results have emerged (digital repositories, open-access journals, etc) (4). Several other citation databases have also become available, including Scopus (5) and Google Scholar (6), both attracting much interest in academic community. Scopus is a multidisciplinary subscription-based bibliographic resource covering more than 16 500 peer-reviewed journals, among them many titles published in less-developed and developing countries and over 1200 open access journals (7). Falagas et al found, for example, that for citation analysis Scopus offers about 20% more coverage than Web of Science (8). Google Scholar probably has the widest coverage because it indexes traditional scientific literature, as well as preprints, institutional repositories, and conference proceedings and it is freely available (9). Both databases were introduced in 2004. Many authors have compared various aspects of citation analyzes based on the data provided by these three resources (10-13). A recent study has compared the citation counts of articles published in 3 most prestigious general medical journals using all 3 databases (3).

The *Croatian Medical Journal* is a general medical journal regularly indexed by both subscription-based bibliographic databases, Web of Science (Science Citation Index Expanded) and Scopus. It is a free-access journal, with full-texts available without any restriction via PubMed Central (14), national open-access repository Hrčak (15), and the journal's Web site. For a small journal from a small country, citation rate is not only a matter of its scientific visibility, but it could also be a matter of local financial support and manuscript inflow. By exploring the *Croatian Medical Journal's* citation profile in all 3 resources, our aim was to find: a) whether the number of citations differed significantly between Web of Science and Scopus since the latter covers more regional (among them many Croatian medical journals) and open access-journals, b) whether the Google Scholar citation score differed significantly from the numbers returned by Web of Science and Scopus, especially regarding citations deriving from peer-reviewed journals, and c) whether Google Scholar can be a qualitative replacement for expensive databases, especially in low-income communities.

METHODS

In the *Croatian Medical Journal* citation analysis, we used the method for calculation of Thompson impact factor (16). We collected two sets of data: a) articles published in 2005-2006 and b) citations they received in 2007.

The first round search was conducted in Web of Science and Scopus by the publication name (*Croatian Medical Journal*) limited to the 2005-2006 period. Web of Science returned 294 and Scopus 286 articles. The comparison showed that all articles returned by Scopus were included in Web of Science database. The list of Web of Science-indexed articles (n=294) was our sample for further analysis.

The number of citations was checked in all 3 databases. Every sample item was checked for citations in Web of Science database, using the "Times Cited" field of the respective bibliographic record. Scopus was searched by the full title of each sample item, and citations were checked using the "Cited By" field of the respective bibliographic record. The Google Scholar was searched by the full title of each sample item (taken from journal Web site tables of contents), and the citations of all retrieved items were analyzed using the "Cited By" function. If an item was not found by the full title search, then it was retrieved by the author, journal title, and publication year, or title word. If not retrieved at all, it was counted as the Google Scholar not-indexed article.

Three lists of captured citations were then checked for the citations received in 2007. The separation was simple for the Web of Science and Scopus citations. However, many records found by Google Scholar did not contain the date of publication and all of them had to be verified further. The identified citation duplicates (eg, title of cited article in various languages) were eliminated.

The 2007 citations returned by all 3 databases were then analyzed and compared. Unique citations, defined as those retrieved by 1 database only and not by the other 2, were registered, as well as those found in all 3 databases. The overlap between databases was also identified and analyzed. All searches were done manually between January and March 2009.

The citations were sorted as follows (10): 1) overlap between all 3 resources; 2) overlap between Web of Science and Scopus; 3) overlap between Web of Science and Google Scholar; 4) overlap between Scopus and Google

Scholar; 5) unique citations from Web of Science; 6) unique citations from Scopus; 7) unique citations from Google Scholar. Proportions were compared using χ^2 -test.

RESULTS

Table 1 presents the number of articles indexed from the *Croatian Medical Journal*, total number of citations returned by each of the 3 analyzed databases, and the number of unique citations. Minimal difference in the number of indexed articles was due to different indexing practices for non-research material (book reviews, obituaries, etc.). Google Scholar returned the greatest share of articles with citations (45%), followed by Scopus (42%), and Web of Science (38%). We identified a group of 145 articles (49%) that had no citations, while a group of 32 articles (11%) returned an equal number of identical citations by all 3 databases. The difference between Scopus and Google Scholar in the average number of citations per indexed article was minimal (1.01 vs 1.03), while the average number of citations received by a Web of Science-indexed article was 0.80. The greatest number of unique citations was found by Google Scholar.

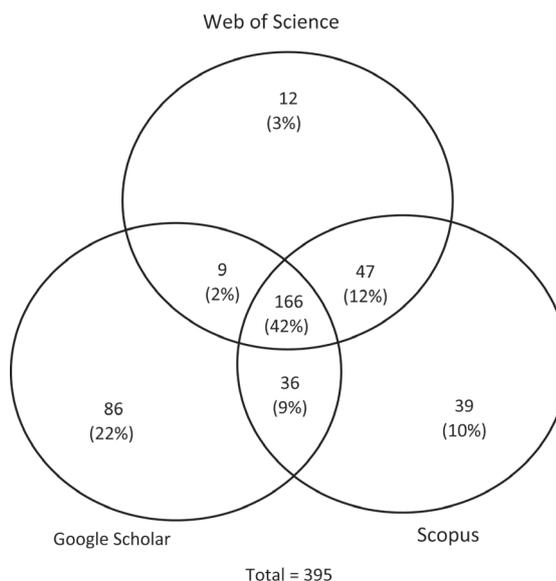
TABLE 1. Croatian Medical Journal indexed articles (in 2005-2006) and citations (in 2007) in the Web of Science, Scopus, and Google Scholar

	Web of Science	Scopus	Google Scholar
Total index articles, No.	294	286	288
Total cited articles, No. (%)	112 (38)	120 (42)	131 (45)
Total citations, No.	234	288	297
Citations per index article, No.	0.80	1.01	1.03
Articles with 1-5 citations	108	111	125
Articles with 6-10 citations	4	7	5
Articles with >10 citations	0	2	1
Unique citations, No.	12	39	86

The analysis of the distribution of unique and overlapping citations as returned by the 3 databases (Figure 1) showed that among 395 citations returned, 166 (42%) were tracked in all 3 of the databases. The greatest overlap was found between Web of Science and Scopus (213 or 54%), followed by Scopus and Google Scholar (202 or 51%) and Web of Science and Google Scholar (175 or 44%). There was a significant difference in unique citations between the 3 databases ($\chi^2 = 56.5, P < 0.001$), as well as between the 3 pairs of databases ($P < 0.05$).

We examined unique citations in Web of Science and Scopus only randomly and it seems that the coverage makes

Figure 1.



The distribution of the unique and overlapped citing articles.

TABLE 2. Google Scholar unique citations

Type of citing publication	Citations, No.	Open access
Journals	55	28
PhD theses	11	11
Books	12	0
Others (conference proceedings, technical reports, project documentation, etc.)	8	8

only a part (though major) of the difference found. For example, some of the Scopus unique citations originated from the Croatian and regional journals, not indexed by Web of Science. But, in Scopus we found several citations not returned by Web of Science because of an error in the process of citing or linking. Further investigation is needed to come upon a valid conclusion.

The list of Google Scholar unique citations regarding the types of citing documents, their web accessibility, and language is presented in Table 2.

The qualitative analysis of Google Scholar unique citations revealed that the majority (64%) came from journals, followed by books and PhD theses (Table 2). Approximately 55% of all citing documents were full-text resources in open access. The language of citing documents was

mostly English, but as many as 25 citing documents (29%) were in Chinese.

DISCUSSION

Our study demonstrated that the Web of Science databases covered the highest-impact scientific journals as the source of citation for the *Croatian Medical Journal*, but that the coverage of Scopus, and especially of Google Scholar was broader and included additional local sources. It has been shown that the Web of Science is a selective source of publication citations (11). On the other hand, for a sample of high-profile general medicine articles, Google Scholar and Scopus may retrieve a greater number of citations than Web of Science (3). In our study, the difference in the number of retrieved citations was 19% between Scopus and Web of Science, 21% between Web of Science and Google Scholar, and 3% between Scopus and Google Scholar.

Previous studies have shown that the degree of citation overlap between the 3 databases varied by field of study (10,11) with no more than 31% of citations overlapping in all 3 databases (10). Our results showed that 42% of the citations could be tracked in all 3 databases. The greatest overlap was found between Web of Science and Scopus (54%), followed by Scopus and Google Scholar (51%) and Web of Science and Google Scholar (44%). Meho and Yang have found the overlap between Web of Science and Scopus of 58.2% (12), while Kousha and Thelwall (11) have found the overlap between Web of Science and Google Scholar of 57%.

There were 86 unique citations from Google Scholar that did not occur either within Scopus or Web of Science. Google Scholar produced more than twice as many unique citations than Scopus. Bakkalbasi et al (10) have also shown that Google Scholar returned the greatest number of unique citing documents for a group of oncology journals, but the difference between Scopus and Google Scholar has been significantly smaller (12% vs 13%). The qualitative analysis of the Google Scholar unique citations revealed that most of them (64%) were from scholarly journals, half of them being in open-access. These findings are typical for the medical field in two aspects, ie, in predominant importance of scientific journals and continuing rapid growth of publicly accessible electronic biomedical information. Our results also confirmed the findings of Kousha and Thelwall on Chinese as a fast-rising language of scientific literature (11).

In conclusion, significant difference in citation rate between Web of Science and Scopus was a result of the difference in coverage. Since Web of Science has recently introduced the policy of wider coverage for "regional scholarship," we may expect that the difference in citation return would not be significant in the near future (17). Our results showed that Google Scholar shared a total of 42% citations returned by the 2 other, more influential, bibliographic resources. Google Scholar list of unique citations is also predominantly journal based, but these journals are mainly of peripheral and/or local character. Though citations received by internationally recognized medical journals are crucial for increasing the visibility of small medical journals, but it is also useful to follow their visibility in journals of their size and importance in the global scientific community. For these small journals, the open question – whether extra citations, especially from non-journal resources, would improve or over-value journal visibility (11), is, therefore, of minor importance.

In the times of various financial constraints, expensive databases are not affordable to many smaller and low-income communities. Several studies (12,18) have confirmed that, despite many open questions raised by non-transparent indexing policies and quality of covered material, Google Scholar may serve as a complementary tool for accessing citation data. In our opinion, it may also serve as an alternative resource for the quick orientational citation insight. Its use in evaluative bibliometric analysis is a matter of further research.

References

- 1 Moed HF. Citation analysis in research evaluation. Dordrecht (The Netherlands): Springer; 2005.
- 2 Thomson Reuters-Science. Scholarly research, publishing and analysis. Available from: http://thomsonreuters.com/products_services/science/academic/. Accessed: January 22, 2010.
- 3 Kulkarni AV, Aziz B, Shams I, Busse JW. Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. JAMA. 2009;302:1092-6. Medline:19738094 doi:10.1001/jama.2009.1307
- 4 Kljakovic-Gaspic M, Petrak J, Rudan I, Biloglav Z. For free or for fee? Dilemma of small scientific journals. Croat Med J. 2007;48:292-9. Medline:17589971
- 5 About Scopus. Available from: <http://info.scopus.com/about/>. Accessed: January 22, 2010.
- 6 About Google Scholar. Available from: <http://scholar.google.com/intl/en/scholar/about.html>. Accessed: January 22, 2010.
- 7 Scopus: What does it cover? Available from: <http://info.scopus>.

- [com/scopus-in-detail/facts/?url=overview/what](#). Accessed: January 22, 2010.
- 8 Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J*. 2008;22:338-42. [Medline:17884971](#) [doi:10.1096/fj.07-9492LSF](#)
 - 9 Hull D, Pettifer SR, Kell DB. Defrosting the digital library: bibliographic tools for the next generation web. *PLoS Comput Biol*. 2008;4:e1000204. [Medline:18974831](#) [doi:10.1371/journal.pcbi.1000204](#)
 - 10 Bakkalbasi N, Bauer K, Glover J, Wang L. Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomed Digit Libr*. 2006;3:7. [Medline:16805916](#) [doi:10.1186/1742-5581-3-7](#)
 - 11 Kousha K, Thelwall M. Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics*. 2008;74:273-94. [doi:10.1007/s11192-008-0217-x](#)
 - 12 Meho LI, Yang K. Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *J Am Soc Inf Sci Technol*. 2007;58:2105-25. [doi:10.1002/asi.20677](#)
 - 13 Bornmann L, Marx W, Schier H, Rahm E, Thor A, Daniel HD. Convergent validity of bibliometric Google Scholar data in the field of chemistry – Citation counts for articles that were accepted by *Angewandte Chemie International Edition* or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. *Journal of Informetrics*. 2009;3:27-35. [doi:10.1016/j.joi.2008.11.001](#)
 - 14 Overview PMC. Available from: <http://www.pubmedcentral.nih.gov/about/intro.html>. Accessed: January 22, 2010.
 - 15 Stojanovski J, Petrak J, Macan B. The Croatian national open access journal platform. *Learn Publ*. 2009;22:263-73. [doi:10.1087/20090402](#)
 - 16 Thomson Reuters-Science. Introducing the impact factor. Available from: http://thomsonreuters.com/products_services/science/academic/impact_factor/. Accessed: January 22, 2010.
 - 17 Testa J. Regional content expansion in Web of Science: opening borders to exploration. Available from: http://thomsonreuters.com/products_services/science/free/essays/regional_content_expansion_wos/. Accessed: January 22, 2010.
 - 18 Bauer K, Bakkalbasi N. An examination of citation counts in a new scholarly communication environment. *D-Lib*. 2005;11. [doi:10.1045/september2005-bauer](#)