

Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses

Matthew E. Falagas,^{*,†,1} Eleni I. Pitsouni,^{*} George A. Malietzis,^{*} and Georgios Pappas[‡]

^{*}Alfa Institute of Biomedical Sciences, Athens, Greece; [†]Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts, USA; and [‡]Institute of Continuing Medical Education of Ioannina, Ioannina, Greece

ABSTRACT The evolution of the electronic age has led to the development of numerous medical databases on the World Wide Web, offering search facilities on a particular subject and the ability to perform citation analysis. We compared the content coverage and practical utility of PubMed, Scopus, Web of Science, and Google Scholar. The official Web pages of the databases were used to extract information on the range of journals covered, search facilities and restrictions, and update frequency. We used the example of a keyword search to evaluate the usefulness of these databases in biomedical information retrieval and a specific published article to evaluate their utility in performing citation analysis. All databases were practical in use and offered numerous search facilities. PubMed and Google Scholar are accessed for free. The keyword search with PubMed offers optimal update frequency and includes online early articles; other databases can rate articles by number of citations, as an index of importance. For citation analysis, Scopus offers about 20% more coverage than Web of Science, whereas Google Scholar offers results of inconsistent accuracy. PubMed remains an optimal tool in biomedical electronic research. Scopus covers a wider journal range, of help both in keyword searching and citation analysis, but it is currently limited to recent articles (published after 1995) compared with Web of Science. Google Scholar, as for the Web in general, can help in the retrieval of even the most obscure information but its use is marred by inadequate, less often updated, citation information.—Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., and Pappas, G. Comparison of Pub Med, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J.* 22, 338–342 (2008)

Key Words: citation analysis • open access • Medline • Institute of Scientific Information (ISI) • education • electronic databases

THE DEVELOPMENT ALONG WITH THE SPREAD of the World Wide Web (WWW) represents an informational revolution, with rapid, practical distribution and storage of data available worldwide. One of the most prominent examples of enhanced storage and distribution of important information is the development of scientific databases, the significance of which was rec-

ognized early. Specifically in the field of medicine, the National Library of Medicine (NLM) in the United States introduced the first interactive searchable database (Medline) in 1971 (1) and subsequently in 1996 (1) added the “Old Medline” database with coverage of publications between 1950 and 1965. In 1997, PubMed (a combination of both Old Medline and Medline) was launched to the Internet by NLM and has become the most popular and one of the most reliable WWW resources for clinicians and researchers.

Another acknowledged source of scientific information is the Institute of Scientific Information (ISI) of Thomson Scientific, which has been serving as a data provider since the early 1960s (2), especially for citation analyses. In recent years electronic database searching has become the *de facto* mode of medical information retrieval, as shown by numerous studies highlighting the utility of the WWW in medicine today (3–7). As expected, numerous efforts have focused on refining the mode of information retrieval and augmenting citation analysis. In that vein in 2004, Scopus and Google Scholar databases were also launched to the Internet. Given that the various scientific databases have their own characteristics, we aimed in this article to compare the utility of the current most popular sources of scientific information in biomedical sciences, namely PubMed, Scopus, Web of Science, and Google Scholar, in retrieval of information on a specific biomedical subject and in up-to-date citation analysis.

MATERIALS AND METHODS

We searched the official home pages of PubMed, Scopus, Web of Science, and Google Scholar to identify and extract information regarding the various characteristics of these databases. We focused on the date of the official inauguration, content, coverage, number of keywords allowed for each search, uses, updating, owner, and characteristics and quality of citations in our analysis of PubMed, Scopus, Web of Science, and Google Scholar.

¹ Correspondence: Alfa Institute of Biomedical Sciences (AIBS), 9 Neapoleos St., 151 23 Marousi, Greece. E-mail: m.falagas@aibs.gr

doi: 10.1096/fj.07-9492LSF

Furthermore, we evaluated the utility of these databases in retrieving information on a particular subject 1) by using a specific keyword referring to a well-defined medical condition (we chose the term “brucellosis” as being specific enough as a condition and a medical concept that is not too vague) and 2) by attempting to perform a citation analysis for a specific recent article. A recent article from a highly cited journal was chosen to assure that referencing to the article would be constant in the present period (the article used was Pappas, G., Akritidis, N., Bosilkovski, M., and Tsianos, E. (2005) Brucellosis. *N. Engl. J. Med.* **352**, 2325–2336). The keyword search was repeated daily for all databases to estimate update speed. The article’s citation analysis was followed through Google Scholar, Scopus, and Web of Science for a period of 2 months.

The search for the identification of relevant information and the extraction of data was performed by two of the authors independently (E.I.P. and G.A.M.). Any discrepancies were discussed in meetings with the senior author (M.E.F.). The evaluation of utility of the databases for a specific keyword search and a specific article was performed by G.P.

RESULTS

General characteristics of the databases

In **Table 1** we present data regarding various characteristics of PubMed, Scopus, Web of Science, and Google Scholar. Scopus is the database that indexes a larger number of journals than the other three databases studied. Web of Science does not provide any data regarding open access articles that it includes (if any).

PubMed, Google Scholar, and Web of Science originate from the United States, whereas Scopus originates from Europe. PubMed and Google Scholar are free and provide open access to all interested clinicians, researchers, and trainees and also to the public in general. Scopus and Web of Science are databases that belong to commercial providers and require an access fee. Regarding Google Scholar, although relevant data are not summarized anywhere, the database is essentially a part of a popular WWW search engine, which means that there are no limits on the languages covered, keywords allowed per search, and list of covered journals, provided for the latter that an electronic edition exists. Similarly there are no data for the frequency of Google Scholar updates (see later discussion of this topic).

PubMed focuses mainly on medicine and biomedical sciences, whereas Scopus, Web of Science, and Google Scholar cover most scientific fields. Web of Science covers the oldest publications, because its indexed and archived records go back to 1900. PubMed allows the larger number of keywords per search but is the only database of the four that does not provide citation analysis. Scopus includes articles published from 1966 on, but information regarding citation analysis is available only for articles published after 1996.

PubMed was developed by the NLM, a division of the National Institutes of Health, and rapidly became synonymous with medical literature research worldwide. It

offers a quick free search with numerous keywords as well as limited searching with various criteria [*i.e.*, search by authors, journal, date of publication, date of addition to PubMed, or type of article]. The results of a search can be displayed in a listing including from 5–500 items per page or as a summary [in which the full title, the names of the authors, the source and PubMed identification (PMID) of each article are presented], and the list can also be presented with abstracts, if available. Information on whether an abstract is available, and free text access is comprehensively represented by a displayed icon. A search can easily be sent to text, a file, the clipboard, e-mail, an RSS feed, and an order. PubMed also allows for direct use of other search engines developed by the NLM, such as GENSAT, OMIM, and PMC, the latter allowing for free full text access to a wide array of previous decades’ publications from numerous journals. Thus, PubMed now offers >1 million freely available articles of which a significant number come from digitized back issues.

One major advantage of PubMed, not reproduced by Scopus or Web of Science, is that it is readily updated not only with printed literature but also with literature that has been presented online in an early version before print publication by various journals. In contrast, Scopus and Web of Science are readily updated for printed literature but do not include online early versions.

The Scopus database was developed by Elsevier, combining the characteristics of both PubMed and Web of Science. These combined characteristics allow for enhanced utility, both for medical literature research and academic needs (citation analysis), yet access to the database is not free, although reviewers for numerous Elsevier medical journals are entitled to 1 month of free use. It offers a quick search, a basic search, an author search, an advanced search, and a source search. In the basic search the results for the keywords chosen can be limited by date of publishing, by addition to Scopus, by document type, and by subject areas, whereas the author search is based only on author names. The advanced search combines the basic search without the limits and the author search, and more operators and codes are allowed. The source search is confined to selection of a subject area, a source type (*i.e.*, trade publication or conference proceedings), a source title, the ISSN number, and the publisher.

The search results in Scopus can be displayed as a listing of 20–200 items per page, and documents can be saved to a list and/or can be exported, printed, or e-mailed. The results can be refined by source title, author name, year of publication, document type, and/or subject area, and a new search can be initiated within the results. The presence of an abstract, references, and free full text is noted under each article title, in addition to where these can be found. When abstracts are displayed, the keywords are highlighted. The fields that can be included in the output are optional (*i.e.*, citation information, bibliographical information,

TABLE 1. *Characteristics of databases*

Characteristic	Pub Med	Scopus	Web of Science	Google Scholar
Date of official inauguration	06/1997 ^a	11/2004	2004 ^b	11/2004
Content				
No. of journals	6000 (827 open access)	12,850 (500 open access)	8700	No data provided (theoretically all electronic resources)
Languages	English (plus 56 other languages)	English (plus more than 30 other languages)	English (plus 45 other languages)	English (plus any language)
Focus (field)	Core clinical journals, dental journals, nursing journals, biomedicine, medicine, history of medicine, bioethics, space, life sciences	Physical sciences, health sciences, life sciences, social sciences	Science, technology, social sciences, arts and humanities	Biology, life sciences and environmental sciences, business, administration, finance and economics, chemistry and materials science, engineering, pharmacology, veterinary science, social sciences, arts and humanities
Period covered	1950–present	1966–present	1900–present	Theoretically all available electronically
Databases covered	Medline (1966–present), old Medline (1950–1965), PubMed Central, linked to other, more specialized, NLM databases	100% Medline, Embase, Compendex, World textile index, Fluidex, Geobase, Biobase	Science citation index expanded, social sciences citation index, arts and humanities citation index, index chemistry, current chemical reactions	PubMed, OCLC First Search
No. of keywords allowed	No limit	30	15	Theoretically no limit
Search				
Abstracts	(+)	(+)	(+)	(+)
Authors	(+)	(+)	(+)	(+)
Citations	(–)	(+)	(+)	(+)
Patents	(–)	(+)	(+)	(–)
Uses	Links to related articles, links to full-text (5426 journals), links to free full text articles for a subset of journals (827 open access journals)	Links to full-text articles and other library resources	Links to full-text, links to related articles	Links to full-text articles, free full-text articles, links to journals, links to related articles, links to libraries
Updating	Daily	1–2 times weekly	Weekly	Monthly on average
Developer/owner (country)	National Center for Biotechnology Information (NCBI), NLM (US)	Elsevier (Netherlands)	Thomson Scientific and Health Care Corporation (US)	Google Inc. (US)
Citation analysis	None	Total number of articles citing work on a topic or by an individual author	As for Web of Science plus the total number of articles on a topic or by an individual author cited in other articles	Next to each paper listed is a “cited by” link; clicking on this link shows the citation analysis

^aPubMed was created by the NLM of the United States with the intention of making the content of Medline accessible *via* the Internet.

^bWeb of Science was created by Thomson Scientific to make citation indices (that E. Garfield assessed since the early 1960s) accessible *via* the Internet.

abstract, and keywords). The citation analysis that Scopus performs is presented as a table with numbers of cited articles for individual years, as well as the total number of cited references for all years. The articles cited can be accessed by simply clicking on the number

of citations. In addition, Scopus has search tips written in 10 languages.

Web of Science was developed by Thomson Scientific, a part of the Thomson Corporation, another private company, and has dominated the field of aca-

demographic reference, mainly through the annual release of the journal impact factor, a tool for evaluating the importance and influence of specific publications. The impact factor has been highly criticized but remains the most widely used of the indexes available. It has a quick search (by entering a topic), an advanced search, a general search, and a cited reference search. Help is offered for all types of searches of author, of group author, and of full source title, as well as of abbreviations. In the cited reference search the search can be limited by cited author, cited work, and cited years, whereas the cited author index and the cited work index can be presented, if the researcher requires it.

The results of a search can be displayed as a listing of 10–50 items per page. The full title, author names, and source are provided. When the full text is available, the option of “view free full text” is present. Related records can be found, sorted by latest date, times cited, relevance, first author, publication year, and source title. The results can be analyzed (*i.e.*, by author, country/territory, or document type), and the citation report is presented with a label bar chart. The results can be refined, and the researcher can view or exclude records.

Google Scholar was developed by Google Inc., another private company, but it is freely accessible and aims to summarize all electronic references on a subject. There is no journal frame/list available for Google Scholar, because it presumably lists all publications that have emerged from the electronic search. Being essentially a Web search engine, its aim is to reach the widest audience available. It allows a quick search and an advanced search. In the advanced search the results can be limited by title words, authors, source, date of publication, and subject areas. The languages of the interface and of the search are optional. The results can be displayed as a listing of 10–300 items per page. Each retrieved article is represented by title, authors, and source, but the abstract and information on free full text availability are not provided by Google Scholar. Under each retrieved article the number of cited articles is noted and can be retrieved by clicking on the relevant link. By clicking on the article title, Google Scholar leads you to a list of possible links to the article, usually on the journal’s site, but for older articles the link is directed to the PubMed citation. In addition, Google Scholar provides links to relevant articles and allows for a general Google Web search, using self-selected keywords from the article and the author name.

Utility trial

A search on the word “brucellosis” elicits thousands of results by all of the databases. PubMed’s simple search elicits the newest ones first and PubMed is updated daily, including online early articles, thus allowing for a comprehensive follow-up of a specific subject. On the other hand, some of the results returned (roughly 5%) were of peripheral relevance to the subject (a kind of false-positive result). Relevant articles (as categorized

by PubMed) can also be assessed. Unfortunately, the relevance is inconsistent. Updates to Scopus and Web of Science were less frequent, generally on a weekly basis. The results produced by Scopus corresponded to its extended listing of included journals with a greater number of citations. False-positive results in Scopus could be eliminated if one is searching for articles including the keyword in the title only, but that search omitted a few relevant articles (an analog to false-negative results). By clicking on the head of the relevant column, articles can be rearranged by most cited in declining order, thus allowing the uninitiated searcher to familiarize himself or herself with the outstanding articles on the subject. PubMed does not possess such a facility. Google Scholar presents results with the most cited first. Although online early articles are included, updates are less frequent (in a period certainly exceeding 1 month).

Searching for citations of a specific article can be a difficult task for academic candidates, and the task is even more difficult when the question of which journal citations are eligible is raised. The use of Web of Science has been the standard, yet Scopus does offer more citation analyses; in our case, Scopus listed 20% more articles referencing our example in any given period than did Web of Science. Admittedly, some of these additional citations were derived from vague journals in non-English languages, yet even Web of Science lists similar journals. The inclusion criteria of Web of Science, similar to the criteria used for calculating the impact factor, have repeatedly been the subject of dispute (8). Both databases include only published articles and not online early ones. One major factor that may bias these results though is the selection of a recent article as an example. If an older article were chosen, Scopus would offer limited citation information, because it covers a significantly shorter period than Web of Science. This observation has also been confirmed in a previous comparison of the three citation databases (9). The use of Google Scholar to determine citations for the particular article was disappointing. The reference list was much shorter than those for the other databases and, as mentioned earlier, updates was less frequent. It was obvious that Google Scholar only cited articles that were accessible electronically. When we assessed articles for other examples, duplicate references (false-positive citations) were a common occurrence.

A final interesting observation was that citation analysis could be adequately updated when Google was used, with article year/volume/page numbers as keyword. In that way, most online early articles were rapidly identified, although the vast number of results prohibits any adequate citation analysis to be performed.

DISCUSSION

A critical review of the information we were able to collect regarding the four sources of scientific informa-


tion that we focused on suggests the following conclusions. PubMed is a very handy, quick, and easy to use database. Its practicality in use, the fact that it is free, and the authority it has gained through the years have made it the most frequently used resource for information in the biomedical field.

Scopus includes a more expanded spectrum of journals than PubMed and Web of Science, and its citation analysis is faster and includes more articles than the citation analysis of Web of Science. On the other hand, the citation analysis that Web of Science presents provides better graphics and is more detailed than the citation analysis of Scopus, probably because Web of Science has been designed with the intention of satisfying users in citation analysis, a field discussed and debated by scientists for decades.

There is a debate in the scientific community on Google Scholar is a database that should be used by clinicians (10, 11), because of its inadequacies (12, 13) and the fact that much information about its content coverage remains unknown. Results with Google Scholar are displayed in relation to times of visits from users, not in relation to another index of quality of the publication.

Google Scholar presents all the benefits and drawbacks of the WWW. It sometimes offers unique options in the scientific field (10, 14): in our example, using its Web search option, a free full text of the article could be retrieved from various Web sites, whereas other databases and the journal itself did not offer free access at the moment. The access may possibly be illegal, but this is a characteristic of the WWW: information is ample, but access is often uncontrolled. The need for a systematic reconstitution of the pros and cons of each database and the development of a formula for free access to such a powerful database apart from PubMed seems warranted.

In conclusion, scientific databases of biomedical information are frequently used by both clinicians and researchers. In this article, we compared the content and various practical aspects in the utility of the main

databases of biomedical scientific information. We found that PubMed remains an important resource for clinicians and researchers, Scopus covers a wider journal range and offer the capability for citation analysis [currently limited to recent articles (published after 1995) compared with Web of Science], and Google Scholar can help in the retrieval of even the most oblique information, but is marred by inadequate, less often updated, citation information 

M.E.F. had the idea for this article. M.E.F. and G.P. developed the methodology used. E.I.P., G.A.M., and G.P. identified the relevant data. M.E.F. and E.I.P. wrote the first draft of the manuscript that was revised extensively by G.P. All authors participated in subsequent revisions of the manuscript and approved its final version. M.E.F. is the guarantor.

REFERENCES

1. http://www.nlm.nih.gov/databases/databases_oldmedline.html
2. ISI Web of Knowledge. <http://scientific.thomson.com/isi/>
3. Tang, H., and Ng, J. H. (2006) Googling for a diagnosis—use of Google as a diagnostic aid: Internet based study. *BMJ* 333, 1143–1145
4. Henzinger M, and Lawrence, S. (2004) Extracting knowledge from the World Wide Web. *Proc. Natl. Acad. Sci. U. S. A.* 101 (Suppl 1), 5186–5191
5. Eysenbach, G. (2003) The impact of the Internet on cancer outcomes. *CA Cancer J. Clin.* 53, 356–371
6. Pappas, G., Papadimitriou, P., and Falagas, M. E. (2007) World Wide Web hepatitis B virus resources. *J. Clin. Virol.* 38, 161–164
7. Falagas, M. E., and Karveli, E. A. (2006) World Wide Web resources on antimicrobial resistance. *Clin. Infect. Dis.* 43, 630–633
8. The impact factor game: it is time to find a better way to assess the scientific literature. (2006) *PLoS Med.* 3, e291
9. Bakkalbasi, N., Bauer, K., Glover, J., and Wang, L. (2006) Three options for citation tracking: Google Scholar, Scopus, and Web of Science. *Biomed. Digit. Libr.* 3, 7
10. Henderson, J. (2005) Google Scholar: A source for clinicians? *CMAJ* 172, 1549–1550
11. Vine R. (2006) Google Scholar. *J. Med. Libr. Assoc.* 94, 97–99
12. <http://www.gale.com/reference/archive/200506/google.html>
13. <http://www.charlestonco.com/review.cfm?id=225>
14. Banks, M. A. The excitement of Google Scholar, the worry of Google Print. *Biomed. Digit. Libr.* 2005 Mar 22;2:2