



Evidence Summary

For Non-expert Clinical Searches, Google Scholar Results are Older with Higher Impact while PubMed Results Offer More Breadth

A Review of:

Nourbakhsh, E. F., Nugent, R. F., Wang, H. F., Cevik, C. F., & Nugent, K. (2012). Medical literature searches: A comparison of PubMed and Google Scholar. *Health Information and Libraries Journal*, 29(3), 214-222. doi: 10.1111/j.1471-1842.2012.00992.x

Reviewed by:

Carol Perryman
Assistant Professor
Texas Woman's University
Denton, Texas, United States of America
Email: cp1757@gmail.com

Received: 25 Nov. 2012

Accepted: 22 Jan. 2013

© 2013 Perryman. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 2.5 Canada (<http://creativecommons.org/licenses/by-nc-sa/2.5/ca/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

Abstract

Objectives – To compare PubMed and Google Scholar results for content relevance and article quality

Design – Bibliometric study.

Setting – Department of Internal Medicine at Texas Tech University Health Sciences Center.

Methods – Four clinical searches were conducted in both PubMed and Google Scholar. Search methods were described as “real world” (p. 216) behaviour, with the searchers familiar with content, though not expert at retrieval techniques. The first 20 results from each search were evaluated for

relevance to the initial question, as well as for quality.

Relevance was determined based on one author's subjective assessment of information in the title and abstract, when available, and then tested by two other authors, with discrepancies discussed and resolved. Items were assigned to one of three categories: relevant, possibly relevant, and not relevant to the question, with reviewer agreement measured using a weighted kappa statistic. The quality of items found to be ‘relevant’ and ‘possibly relevant’ was measured by impact factor ratings from Thomsen Reuters (ISI) Web of Knowledge, when available, as well as

information obtained by SCOPUS on the number of times items were cited.

Main Results – Google Scholar results were judged to be more relevant and of higher quality than results obtained from PubMed. Google Scholar results are also older on average, while PubMed retrieved items from a larger number of unique journals.

Conclusion – In agreement with earlier research, the authors recommended that searchers use both PubMed and Google Scholar to improve on the quality and relevance of results. Searches in the two resources identify unique items based upon the ranking algorithms involved.

Commentary

Comparisons and tests of the utility, quality, and relevance of searching Google Scholar and PubMed for clinical questions in previous research (e.g., Mastrangelo et al., 2010) have found that Google Scholar is a valuable adjunct to PubMed searching that may be easier for the non-expert searcher (Shultz, 2007). As well, findings have shown that Google Scholar-retrieved items tend to be older and less specific due to filters and terminology affordances not provided in Google Scholar (Anders & Evans, 2010). This too is confirmed by the present study, as each resource examined contains unique materials not indexed by the other, including the gray literature accessible via Google Scholar (Shultz, 2007). Comparative measures of quality have included ranking position in results lists, presence of terms and related terms in abstracts and titles (Tober, 2011), and measures of sensitivity and precision (Anders & Evans, 2010). Using retrieval rankings to compare PubMed with Google Scholar is questionable at best, as search algorithms and objectives are quite different. The authors compared only the first twenty results from Google Scholar to those in PubMed, yet these resources rank results very differently. Google Scholar also indexes and retrieves items from a very broad spectrum of disciplines, while PubMed coverage, though still broad, is limited to

biomedical publications. The relevance of retrieved items is assessed only through subjective examination of item titles and abstracts (though not all items had abstracts), and no further information, including the titles of items found relevant, were included.

The authors based their quality assessments on information from SCOPUS about the citedness (Tober, 2011) and overlap of results from related Cochrane reviews (Anders & Evans, 2010). However, and without explanation, the authors have chosen to use Web of Knowledge for impact factor information rather than SCOPUS, even though these databases do not provide the same publication coverage. This aspect of evaluation would have been improved by using just one of these resources. In addition, recognized problems with the use of impact factors and citation metrics to impute quality are not discussed as they relate to the present study.

This present study offers little new information to this still relatively sparse corpus. The authors conducted searches using a 'real world' level of search expertise, which is a departure from previous efforts, and of some value in that clinicians are known to employ a limited number of search terms and to examine only the first page of results. However, a lack of rigor and transparency in this study mars potential applicability.

From an initial set of four clinical questions, authors employed search strings in PubMed and Google Scholar, using different limiters in two of the four searches for both databases. The limiters for Q1 and Q2 in PubMed are *reports*, *clinical series*, and *reviews* (Q2), but in Google Scholar, only a single limiter, *randomized controlled trials*, was used for Q1. In both instances for Google Scholar searches, the authors set search limits to *English language* and to the then-available disciplinary set of *Medicine, pharmacology, and veterinary sciences* (At the time of this review, disciplinary set limiters were no longer available in Google Scholar). As the authors limited their relevance and quality assessments to the first twenty results retrieved, the different search strings

and filters may have radically altered findings, retrieval rankings, and evaluations of quality.

Discrepancies between the initial question and the search strings used in the two search engines are not clarified or explained. While Q2 and Q4 each include a facet about outcomes, the search strings listed include no mention of this concept. The result is that readers cannot discern whether assessments of relevance were based on the complete initial questions using the information provided. For assessments of quality, the authors used Web of Knowledge to check journal title impact factor rankings, paired with statistics about how often the retrieved items were cited in ensuing literature. Several problems are apparent in the use of this methodology. First, the authors state that not all journal titles for retrieved items are indexed in Web of Knowledge, while unlisted titles are not provided. Second, while Google Scholar also retrieved non-article items, these are not likely to be indexed in Web of Knowledge. In both cases, this gap has undoubtedly affected quality assessments. The lack of data for impact factors or citedness is not addressed except as a brief footnote (p. 218). Moreover, citedness has been disputed as a measure of quality, but the authors do not address this. While the authors employed solid and appropriate descriptive statistics to describe inter-rater reliability and correlations between impact factors and citedness, failure to address this issue affects the research rigour. Finally, problems with Google Scholar reliability are recognized to limit a more rigorous and supported comparison between it and other, more conventional bibliographic databases, including PubMed. As Jacsó (2012) has concluded, Google Scholar metadata is “substandard, neither reliable nor reproducible and it distorts the metric indicators at the individual, corporate and journal levels” (p. 462). Considering his remarks, this reviewer can only speculate that the present research is one example of exactly what Jacsó warned against when he stated:

It is hoped that the wailing sound of air-raid sirens in this paper will act as an early warning for the tempting

siren song in current papers about using Google Scholar to compute bibliometric data (publication and citation counts, the h-index and its variants) for ranking journals on a nationwide scale as part of assessing the scholarly productivity and impact of universities and colleges. (p. 463)

Ultimately, the value of the study is limited by lack of transparency, making it difficult to evaluate or replicate the work. Readers are asked to accept assessments of relevance without seeing the relevant/non-relevant citations, or even inclusion and exclusion criteria with which to deepen understanding and enable replication.

The perspectives of non-expert searchers in Google Scholar and PubMed comprise a valuable contribution to a scarce body of literature. Awareness of a more naïve searcher’s perspective is needed to inform information professionals working with clinicians who have more advanced knowledge of subjects, but who are limited in their searching expertise. In addition, the research provides a basis for further study that may lead to improvement of retrieval mechanisms and techniques for both PubMed and Google Scholar.

This reviewer used a bibliometric tool (Perryman, 2009) while evaluating this study, as no currently available tool would work to evaluate this research methodology. The question set is based upon existing published tools, with questions specific to bibliometric studies

References

- Anders, M. E., & Evans, D. P. (2010). Comparison of PubMed and Google Scholar literature searches. *Respiratory Care*, 55(5), 578-583.
- Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. *Information Processing & Management*, 36(4), 533-550.

- Jacsó, P. (2012). Using Google Scholar for journal impact factors and the h-index in nationwide publishing assessments in academia—siren songs and air-raid sirens. *Online Information Review*, 36(3), 462-478.
- Perryman, C. (2009). *Critical appraisal tool for bibliometric studies*. Retrieved from <http://evidence-based-librarian.blogspot.com/2013/03/another-update-bibliometric-study.html>
- Shultz, M. (2007). Comparing test searches in PubMed and Google Scholar. *Journal of the Medical Library Association*, 95(4), 442.
- Tober, M. (2011). PubMed, ScienceDirect, Scopus or Google Scholar: Which is the best search engine for an effective literature research in laser medicine? *Medical Laser Application*, 26(3), 139-144.
- Mastrangelo, G., Fadda, E., Rossi, C. R., Zamprogno, E., Buja, A., & Cegolon, L. (2010). Literature search on risk factors for sarcoma: PubMed and Google Scholar may be complementary sources. *BMC Research Notes*, 3(1), 131.