Health Information and Libraries Journal

Health Libraries Group | cilip

DOI: 10.1111/j.1471-1842.2012.00992.x

# Medical literature searches: a comparison of PubMed and Google Scholar

Eva Nourbakhsh*, Rebecca Nugent†, Helen Wang*, Cihan Cevik* & Kenneth Nugent*

*Department of Internal Medicine Texas Tech University Health Sciences Center, Lubbock, TX, USA, and †Department of Statistics Carnegie Mellon University, Pittsburgh, PA, USA

## Abstract

*Background*: Medical literature searches provide critical information for clinicians. However, the best strategy for identifying relevant high-quality literature is unknown.

*Objectives*: We compared search results using PubMed and Google Scholar on four clinical questions and analysed these results with respect to article relevance and quality.

*Methods*: Abstracts from the first 20 citations for each search were classified into three relevance categories. We used the weighted kappa statistic to analyse reviewer agreement and nonparametric rank tests to compare the number of citations for each article and the corresponding journals' impact factors.

*Results*: Reviewers ranked 67.6% of PubMed articles and 80% of Google Scholar articles as at least possibly relevant ($P = 0.116$) with high agreement (all kappa $P$-values < 0.01). Google Scholar articles had a higher median number of citations (34 vs. 1.5, $P < 0.0001$) and came from higher impact factor journals (5.17 vs. 3.55, $P = 0.036$).

*Conclusions*: PubMed searches and Google Scholar searches often identify different articles. In this study, Google Scholar articles were more likely to be classified as relevant, had higher numbers of citations and were published in higher impact factor journals. The identification of frequently cited articles using Google Scholar for searches probably has value for initial literature searches.

*Keywords*: citations, Google Scholar, impact factor, PubMed, searches

### Key Messages

#### Implications for Practice

- Medical literature searches using either PubMed or Google Scholar to answer clinical questions recover unique 'relevant' articles at the top of their lists (first 20 articles in this study).
- In this study, the articles retrieved by Google Scholar searches were more frequently classified as at least possibly relevant than those retrieved from the PubMed database. They also had higher numbers of citations and came from journals with higher impact factors.
- PubMed articles also appearing in the initial Google Scholar lists often have higher numbers of citations than articles only appearing in the initial PubMed list.

#### Implications for Policy

- Searches with the Google Scholar search engine can retrieve important reference resources.
- Preliminary searches using both Google Scholar and PubMed provide article lists with recency, relevance and more frequent citations.
- Optimal searches depend on the clinical question, selection of search terms and strategy, and input from search specialists.

Correspondence: Kenneth Nugent, Department of Internal Medicine, Texas Tech University Health Sciences Center, Lubbock 79430, TX, USA. Email-kenneth.nugent@ttuhsc.edu

## Introduction

Clinicians and medical trainees can obtain medical information from multiple electronic resources by searching databases to identify and retrieve information relevant to the particular question at hand. Institutional libraries often provide access to multiple databases, including subscription journals and databases. The success using a particular database will depend on the resource coverage, the currency of the coverage and the relevancy algorithms used to identify articles. Success will also depend on the type of information needed, the selection of search terms and strategy, and subject expertise. Some questions require a broad overview of a topic, and some require very specific, potentially unique information. PubMed is the most widely used free database for medical literature and provides a search interface to MEDLINE, a repository containing approximately 5000 biomedical journals dating back to 1950. Searches with PubMed utilise a controlled vocabulary and various limiters (i.e. search constraints) to provide a list of journal articles with the most recent articles at the top of the list. In November 2004, Google launched Google Scholar's experimental or beta version. This is a specialist search engine that retrieves information from multiple electronic resources, including peer-reviewed journals, books, abstracts, technical reports and conference proceedings (http://en.wikipedia.org/wiki/Google_Scholar. Accessed 15 June 2009.).[1–3] However, its update frequency is slower and its relevancy algorithms are not publicly available.[4,5] Research studies on Google Scholar's proprietary algorithms with reverse engineering [a process used to understand the software code in software by analysing its operation (output)] have demonstrated that the number of citations is an important criterion in the initial list of articles and that the date of publication is not an important criterion. Therefore, the Google Scholar search engine likely has both advantages and disadvantages compared with the PubMed database and will not necessarily identify the same literature given identical search terms. However, most clinicians use search engines and databases only as information recovery tools and may not be aware of underlying differences in the search algorithms.[6] In this study, we employed the typical search methods of the average user, and we compared the search results from the PubMed database and Google Scholar specialist search engine on four clinical questions and analysed the results to characterise their articles' relevance and quality.

## Methods

We compared searches with PubMed and Google Scholar by evaluating articles recovered for four clinical questions for relevance and quality. Relevance was determined from the abstract and classified (see below) by reviewers based on subject knowledge. Article quality was inferred from the number of citations for the article in the literature and from the impact factor of the journal in which the article was published. The authors picked four clinical questions whose answers would require four different clinical search strategies. The first question would require identification of randomised controlled trials, the second question identification of clinical studies with a particular therapeutic approach, the third question identification of clinical studies with a particular disease association, and the fourth question identification of studies that compared physician characteristics to patient outcomes. These questions were chosen because the authors (KN, CC, EN) had knowledge of the relevant medical literature, had written review articles on these topics and/or were working on manuscripts related to the questions. We use PubMed and Google Scholar frequently for literature searches but are not experts in search strategies. Hence, we think we are typical of most users of these search engines, namely good at some content areas but not trained in medical informatics.[7,8]

Both PubMed and Google Scholar were accessed at the Texas Tech University Health Sciences Center library website; one author (KN) performed all searches. The clinical questions are listed below; specific search strings and limiting terms are listed in Table 1.

Question 1: 'Do intracardiac defibrillators reduce mortality in patients with non-ischemic cardiomyopathy?'

Question 2: 'Is thrombolytic therapy effective in patients with prosthetic valve thrombosis?'

Question 3: 'Is hyperthyroidism associated with pulmonary hypertension?'

**Table 1** Search strings and search limiting terms

| | | |
|---|---|---|
| Q1 | | |
| PubMed | | |
| | Strings | Defibrillator, implantable [Mesh] AND cardiomyopathy [Mesh] |
| | Limits | Human, randomized controlled trials, English, all adults: 19+ years |
| Google Scholar | | |
| | Strings | ICDs and non-ischemic dilated cardiomyopathy and randomized controlled trial |
| | Limits | Medicine, pharmacology, and veterinary sciences, English pages only |
| Q2 | | |
| PubMed | | |
| | Strings | Heart valve prosthesis, implantation [Mesh] AND thrombosis [Mesh] AND thrombolytic therapy [Mesh] |
| | Limits | Human, English, all adults: 19+ years, case reports, clinical series, reviews |
| Google Scholar | | |
| | Strings | Prosthetic valve thrombosis and thrombolytic therapy |
| | Limits | Medicine, pharmacology, and veterinary sciences, English pages only |
| Q3 | | |
| PubMed | | |
| | Strings | Hyperthyroidism [Mesh] AND hypertension, pulmonary [Mesh] |
| | Limits | Human, English, all adults: 19+ years |
| Google Scholar | | |
| | Strings | Hyperthyroidism and pulmonary hypertension |
| | Limits | Medicine, pharmacology, and veterinary sciences, English pages only |
| Q4 | | |
| PubMed | | |
| | Strings | Recertification AND internal medicine |
| | Limits | None |
| Google Scholar | | |
| | Strings | Recertification and internal medicine |
| | Limits | Medicine, pharmacology, and veterinary sciences, English pages only |

Question 4: 'Does ABIM recertification translate into improved patient outcomes?'

We examined the abstracts from the first 20 citations for each search. This decision was based on the practical assumption that most people do not look past the first page of any search results and, if the expected result is not on the first page, may search again with different terms. This assumption is consistent with recent studies on search strategies.[8–10] Given the importance of the initial results page and its dependence on the different search algorithms employed by PubMed and Google Scholar (discussed later), this study focuses on understanding the differences in the first 20 citations rather than being a comprehensive comparison of all results produced by these two search engines. We report the most common journals, the distribution of publication dates and the number of journals appearing in both search engines (Table 2).

Three authors (Q1, Q2: CC, KN; Q3, Q4: KN, EN) independently reviewed the quality of information in each abstract to determine whether the abstract clearly was relevant to the question [category 1], possibly was relevant to the question and could lead [via articles citing the article and references) to other more relevant articles (category 2), or was not relevant (category 3)]. These decisions were based on our knowledge of the topics and represent a 'real-world' approach to information retrieval. When comparing PubMed and Google Scholar results, we were primarily interested in 'relevant' articles classified as categories 1, 2. If the reviewers did not agree on whether or not an article was at least possibly relevant [category 1, 2] vs. not relevant [category 3], the abstract was discussed prior to assigning a final classification. The reviewer who gave an initial assignment of category 3 then had an opportunity to change the rating to category 1 or 2, or the other reviewer could downgrade the abstract to category 3. The reviewers, however, were not required to change their assignments. This discussion was necessary on seven abstracts (Q1: 2, Q2: 0, Q3: 2, Q4: 3) from the PubMed searches and twelve abstracts (Q1: 2, Q2: 3, Q3: 3, Q4: 4) from the Google Scholar searches. Post-discussion, reviewers agreed on whether or not the abstracts for Questions 1, 2, 3 were category 1, 2 (although not necessarily on the final category). On Question 4, the two reviewers could not agree about the corresponding classification of one Google Scholar abstract (category 2 vs. 3) and five PubMed abstracts [category 2 vs. 3 (2); category 3 vs. 2 (2); category 1 vs. 3 (1)]. In subsequent analysis of category 1, 2 articles, only Question 4 articles classified as at least possibly relevant by both authors were included. The agreement among the reviewers on their individual classifications both before and after these discussions was determined with Cohen's weighted kappa sta-

**Table 2** Characterising the search engine results for PubMed and Google Scholar

| | No. of items returned | No. of items in both* | Publication date* (median; IQR) | No. of journals* | Journals with > 1 Item* (No. of Items) |
|---|---|---|---|---|---|
| Q1 | | | | | |
| PubMed | 14 | 4 | 2002.5 [1999, 2004] | 12 | J Am Coll Cardiol (2), Am J Cardiol (2) |
| Google Scholar | 2790 | | 2003.5 [2002, 2005] | 10 | J Am Coll Cardiol (7), N Engl J Med (2), Am Heart J (2), Pacing Clin Electrophysiol (2), Circulation (2) |
| Q2 | | | | | |
| PubMed | 23 | 0 | 2003† [2001, 2006] | 17 | J Heart Valve Dis (3), Ann Thorac Surg (2) |
| Google Scholar | 4460 | | 1998 [1994, 2002] | 8 | J Am Coll Cardiol (7), Chest (4), Am Heart J (2), Circulation (2), J Heart Valve Dis (2) |
| Q3 | | | | | |
| PubMed | 23 | 9 | 2005.5† [2004, 2007] | 19 | Am J Med (2) |
| Google Scholar | 10 500 | | 2001 [1997, 2004] | 15 | Chest (4), Ann Intern Med (2), Am J Med (2) |
| Q4 | | | | | |
| PubMed | 82 | 0 | 2006.5† [2004, 2007] | 15 | Ann Intern Med (4), J Cont Educ Health Prof (2), Am J Gastroenterol (2) |
| Google Scholar | 13 700 | | 1995 [1993, 2004] | 5 | Ann Intern Med (14), J Gen Int Med (3) |

*For the first 20 items.
†Publication dates significantly different (Wilcoxon $P < 0.001$).

tistic (Table 3).[12] This inter-rater measure is used with ordered rankings and weights any disagreement according to its degree. Statistically significant agreement is indicated by $P$-values $\leq 0.05$. Our subsequent analyses on article quality focused on the articles identified in post-reviewer discussions as categories 1, 2 (i.e. at least possible relevant).

The number of citations for articles in categories 1 and 2 was determined from Scopus. The impact factor of the journals identified for articles in categories 1 and 2 was determined from the ISI Web of Knowledge. We report the median number of citations with an intraquartile range and the median journal impact factor with an intraquartile range for all category 1, 2 articles retrieved by PubMed and Google Scholar (across all questions) in Table 4. We used Wilcoxon's nonparametric rank-sum test to determine statistically significant distributional differences between PubMed and Google Scholar.[13] We also used a multiple comparison Kruskal–Wallis nonparametric rank-sum test to determine whether or not there were distributional differences across the four questions for each search engine.[14] (The Wilcoxon is the special case of a two sample Kruskal–Wallis test.) We also compared the articles that appeared in the first 20 in both search engines with those that appeared only in one search engine (Table 5). Two-sided $P$-values $\leq 0.05$ were considered statistically significant.

## Results

The time required for searches for each question ranged from 6 to 16 min; most of the time involved the user reviewing MeSH terms for relevance to the question. The total number of retrieved articles ranged from 14 to 82 in the PubMed database and from 2790 to 13 700 in the Google Scholar searches (Table 2). Publication dates and other characteristics of the search results are also included in Table 2. As previously discussed, we limited our analysis to the first 20 articles except for Question 1 on which PubMed returned only 14 articles. In our PubMed searches, abstracts were included for the majority of articles (Q1: 100%, Q2: 70%, Q3: 90%, Q4: 60%). Google Scholar searches almost uniformly returned articles with abstracts (Q1: 100%, Q2: 100%, Q3: 100%,

**Table 3** Characterising the relevance of content for the first 20 items before/after reviewer discussion

| | Before reviewer discussion relevance categories 1, 2, 3 | After reviewer discussion relevance categories 1, 2, 3 |
|---|---|---|
| Q1 | | |
| PubMed | R1: 6, 2, 6 | R1: 6, 2, 6 |
| | R2: 6, 4, 4 | R2: 6, 2, 6 |
| | **0.702 (0.0014)*** | **0.854 (<0.001)** |
| Google Scholar | R1: 10, 8, 2 | R1: 10, 8, 2 |
| | R2: 10, 10, 0 | R2: 9, 9, 2 |
| | *0.333 (0.074)* | **0.632 (<0.001)** |
| Q2 | | |
| PubMed | R1: 10, 3, 7 | R1: 10, 3, 7 |
| | R2: 10, 3, 7 | R2: 10, 3, 7 |
| | **1.000 (<0.001)** | **1.000 (<0.001)** |
| Google Scholar | R1: 14, 5, 1 | R1: 14, 6, 0 |
| | R2: 17, 1, 2 | R2: 16, 4, 0 |
| | **0.400 (0.020)** | **0.737 (<0.001)** |
| Q3 | | |
| PubMed | R2: 19, 1, 0 | R2: 19, 1, 0 |
| | R3: 18, 0, 2 | R3: 18, 2, 0 |
| | **0.375 (0.0021)** | **0.643 (0.0021)** |
| Google Scholar | R2: 17, 2, 1 | R2: 17, 2, 1 |
| | R3: 15, 1, 4 | R3: 16, 3, 1 |
| | **0.369 (0.029)** | **0.610 (0.0011)** |
| Q4 | | |
| PubMed | R2: 1, 8, 11 | R2: 1, 8, 11 |
| | R3: 6, 3, 11 | R3: 5, 5, 10 |
| | 0.202 (0.193) | **0.419 (0.009)** |
| Google Scholar | R2: 1, 6, 13 | R2: 1, 6, 13 |
| | R3: 3, 4, 13 | R3: 3, 5, 12 |
| | **0.531 (0.0027)** | **0.771 (<0.001)** |

*Weighted kappa for reviewer agreement (*P*-value < 0.05 in bold; 0.05 ≤ *P*-value < 0.10 in italics).

**Table 4** Comparing citations and journal impact factors† for all category 1, 2 articles

| | No. of [1, 2] articles (%) | No. of citations* (*P*-value)** | Impact factor* (*P*-value)** |
|---|---|---|---|
| All category 1, 2 articles | | | |
| PubMed‡ | 50 (67.6%) | 1.5 [0, 11.75] | 3.55 [1.849, 5.351] |
| Google Scholar | 64 (80.0%) | 34.0 [14.00, 57.00] (**<0.0001**) | 5.17 [3.519, 13.54] **(0.036)** |
| Comparing question pairs*** | | | |
| PubMed | | Q1 ≠ Q2 | None |
| Google Scholar | | Q2 ≠ Q3 | Q2 ≠ Q4; Q3 ≠ Q4 |

*Median [IQR].
**We used a Wilcoxon/Kruskal–Wallis test to test for difference between the search engines.
***We used a multiple comparison Kruskal–Wallis test to look for significantly different question pairs for each search engine ($P < 0.05$).
†ISI did not have impact factors available for all journals.
‡Recall PubMed only returned 14 articles for Q1.
Bold values in the table represent *P*-values less than or equal to 0.05.

**Table 5** Characterising the category 1, 2 articles retrieved by both search engines

| | *n* | No. of citations*(*P*-value) | Impact factor*(*P*-value) |
|---|---|---|---|
| Articles retrieved by | | | |
| Both engines | 13 | 22 [8, 48] | 5.23 [3.358, 8.195] |
| Only one engine | 88 | 14.5 [1, 38.25] (0.1514) | 5.10 [1.914, 10.92] (0.7315) |
| Only PM | 37 | 1 [0, 4] (<**0.0001**) | 3.30 [1.624, 5.709] (0.1948) |
| Only GS | 51 | 35 [19, 57] (0.3247) | 5.10 [3.530, 13.54] (0.7625) |

*Median [IQR]; *P*-value comparing each group to the 'both' group: $P \leq 0.05$ in bold.

Q4: 95%). Abstract lengths ranged from approximately 30 words for case reports to 400 words for structured scientific abstracts. We used titles to determine relevance for articles that did not have abstracts. These articles were commentaries, letters or case reports retrieved by PubMed searches on Questions 2 and 4 and a commentary retrieved by Google Scholar on Question 4. Four articles occurred in both the PubMed and Google Scholar initial lists for Question 1 and nine articles for Question 3. There were no overlapping citations in Questions 2,4. For all questions except Question 1, the median publication date for Google Scholar searches was older than that of the PubMed database (all *P*-values < 0.001). However, in all cases, the PubMed database had more unique journals than Google Scholar searches.

There was good agreement among the reviewers about the relevance of the articles prior to discussion ($\kappa = 0.202–1.000$; Table 3). After reviewer discussion of the abstracts, the kappa statistics increased slightly in magnitude and significance (0.419–1.00; Table 3). There was no 'dominant' reviewer, and all three reviewers made only slight changes to their initial rankings. After reviewer discussion and possible reclassification, the number of relevant and possibly relevant articles for each question ranged from 8 (57%) to 20 (100%) in the PubMed searches and from 7 (35%) to 20 (100%) in the Google Scholar searches. Overall, reviewers ranked 67.6% of the PubMed articles and 80% of the Google Scholar articles as at least possibly relevant ($P = 0.116$)

As we had no *a priori* hypothesis that search result differences might be dependent on the question, category 1 and 2 articles initially were combined across all questions for comparison (Tables 4 and 5). The median number of citations was 34 [14.00, 57.00] for Google Scholar articles and 1.5 [0, 11.75] for PubMed articles ($P < 0.0001$). The median journal impact factor for the Google Scholar articles was 5.17 [3.52, 13.54]; for the PubMed articles, it was 3.55 [1.85, 5.35] ($P = 0.036$) (Table 4). For each search engine, we also compared the category 1, 2 articles by question to potentially identify any significant differences among the questions using a Kruskal–Wallis nonparametric rank-sum test correcting for multiple comparisons ($P < 0.05$, Table 4). Among the PubMed articles, Question 1 differed in number of citations from Question 2; no differences were found among the questions' impact factors. Among the Google Scholar articles, Question 2 differed in citations from Question 3 and in impact factors from Question 4. Question 3 also differed in impact factors from Question 4. However, no one question was clearly different from the other three questions, and so the articles remain combined.

Thirteen articles appeared in the first 20 citations for both PubMed and Google Scholar, and all were ranked as relevant by the reviewers (i.e. category 1). We compared these articles with articles retrieved by only one search engine ($n = 88$) in Table 5. There was no difference in the number of citations or journal impact factor ($P = 0.1514$, 0.7315). However, restricting the comparison to one search result showed slightly different results. When looking at the PubMed articles only, those that also appeared in the Google Scholar top 20 had a higher median number of citations than those that appeared only in the top 20 of the PubMed database [22 (8, 48) vs. 1 (0, 4), $P < 0.0001$]. This difference was not seen in the Google Scholar articles when comparing the articles that also appeared on PubMed list with those that only appeared on the Google Scholar list. No differences were seen in journal impact factors for either search engine.

## Discussion

This study demonstrates that both the PubMed database and the Google Scholar search engine retrieve useful but different initial sets of journal articles when used to answer relatively uncomplicated clinical questions. The Google Scholar specialist search engine retrieved more total articles than PubMed-based searches, which were restricted by the various limiters used in this study. The number of common articles in the first 20 listed by both search engines ranged from 0 to 9 articles for the four individual questions, but these overlap articles were not different in the number of citations or in journal impact factors when compared with articles recovered by only one search engine. However, all articles retrieved by both search engines were assigned the highest relevance score by our reviewers, suggesting that articles listed in the initial results of both the PubMed database and Google Scholar search engine may have higher relevance based on content assessment. Bernstam has stated that 'relevance' refers to the question of whether a search result deals with the same concepts as a query or whether, alternatively, the article satisfies the information needs of the user who developed the initial query.[9] Cosijn and Ingwersen[11] have analysed the dimensions of relevance in information retrieval using categories of relevance and attributes of relevance. The simplest approach in their analysis involves algorithms, which connect queries to information objects. All other approaches require the review of the information objects for content. We used a 'real-world' approach to relevance in this study, which probably operated at the cognitive/pertinence level in the Cosijn classification.

The results in our study likely reflect the different relevancy algorithms used by these two search engines. PubMed uses algorithms based on MeSH terms and reports the most recent articles at the top of the list (which likely have not had adequate time to be appropriately cited). In contrast, research studies on Google Scholar's proprietary algorithms have demonstrated that the number of citations is an important criterion in the initial list of articles and that the date of publication is not an important criterion.[15,16] In our study, the median publication dates for articles retrieved by Google Scholar searches were earlier than those retrieved from the PubMed database. Therefore, it seems reasonable to expect that the article lists will differ when using these two search engines for the same question. Shultz compared 10 searches with PubMed and 10 searches with Google Scholar and demonstrated that that both searches retrieved unique articles and that there was a wide range in overlap when the lists were compared.[17] Her study did not assess relevance, article quality (number of citations) or journal quality. Other authors have compared PubMed search results with Google Scholar search results on clinical topics using a bibliographic standard [a database of bibliographic records (articles)] developed by expert panels on a particular topic. This approach allows the calculation of precision (which represents the probability that a retrieved article is relevant) and recall (which represents the probability that a relevant article is retrieved). Mastrangelo and co-workers conducted literature searches on risk factors for sarcoma using this approach and determined that Google Scholar searches had a higher recall (or sensitivity) and that PubMed searches had a higher precision.[18] Anders and Evans[19] did a comparative study on the retrieval of information on respiratory care topics, using a reference list from Cochrane collaboration-based systematic reviews as a benchmark for the search results. They found that the PubMed database and Google Scholar search engine have similar recall in these searches, but that PubMed had better precision (likely improved by their use of the clinical queries filter in PubMed). These three studies suggest that Google Scholar searches compare favourably with PubMed searches but have both advantages and disadvantages. Our study focused on content relevance and article quality and suggests that the Google Scholar search engine retrieves more relevant, higher quality articles.

In our study, Google Scholar searches retrieved category 1, 2 articles with a higher median number of citations. As discussed above, this result is explained at least partially by Google Scholar's ranking strategies to identify articles with higher number of citations.[15,16] Kalkarni *et al.*[20] compared Google Scholar to Web of Science and Scopus using searches to determine the number of citation counts and the types of citations (such as original articles, editorials, reviews, etc). They found that the citations identified by the three services differed in both number and type. Bakkalbasi *et al.*[21] compared these three search engines to determine citation counts in two quite different disciplines (oncology and condensed matter physics) and concluded that the search tool providing the most complete set of citing literature probably depends on both the subject and the publication year of the given article. These studies (including ours) suggest that the Google Scholar search engine is an important adjunct to other resources for tracking citations.

Several groups of information scientists have studied methods to provide relevance ranking for articles recovered from MEDLINE by using information on citations. Bernstam[9] demonstrated that citation algorithms were more effective than non-citation-based algorithms in identifying important articles that were present in a pre-existing bibliography. Lin and co-workers showed that the citation count per year, the total citation count and the journal impact factor ranked important articles recovered from MEDLINE. In their study, the citation count per year was more useful.[22] Tanaka and co-workers developed a method for identifying articles likely to be highly cited, using information available at the initial listing in MEDLINE.[23] The final method in Tanaka's study depended on MeSH terms, journal impact factor and the number of authors. All these strategies require the technical expertise to link MEDLINE databases with other databases that can provide information about the number of citations and/or journal impact factor. These strategies all move articles that are more relevant and/or higher quality to a higher rank in the list of articles and mimic the Google Scholar search engine, which uses citations as a factor in the final list of articles.

Steinbrook[6] suggests that Google Scholar searches provide a broad set of results depending on the search terms, that the PubMed database provides a more focused set of results and that in-depth searching may require the use of several search engines. PubMed searches can limit the total search effort if the clinician needs to identify all possible articles pertinent to a clinical question; Google Scholar searches often retrieve a much larger number of articles, which would be a distinct disadvantage in this situation. We would suggest that both the PubMed database and the Google Scholar search engine should be used during the initial search effort. The PubMed database provides a list of the most recent articles entered into MEDLINE by MeSH terms, which may facilitate revision of subsequent search queries. These searches also provide a list of related articles. Google Scholar searches provide a list of articles that is ranked, at least in part, by the number of citations or links with these articles. This likely identifies better-quality articles as other authors have used the information in these articles. The Google Scholar specialist search engine also allows forward searching using these 'cited by' links, which will identify more recent articles. Finally, Whipple has suggested that collaboration between content experts and expert searchers is optimal for searches.[24] We would agree, especially as these search engines have periodic updates with enhanced capabilities, that some input from expert searchers may be essential.[17] Ultimately, the best search strategy will likely depend on the question and may require the use of several databases as the major databases have different characteristics.[25,26]

## Conclusion

In summary, we have demonstrated that searches with Google Scholar identify articles relevant to clinical questions and that the initial set may differ significantly from the initial set retrieved from PubMed searches. Both search engines identify unique articles in the initial list of articles, and arguably both are useful for identifying an initial set of references. Google Scholar articles appear to have a higher number of citations and to come from higher impact journals when compared with PubMed articles. In addition, Google Scholar searches can identify information from multiple sources. Therefore, Google Scholar searches can identify important information on clinical questions and should be used in conjunction with PubMed searches.

## Disclosures

None.

## References

1 Assisi, F. C. Acharya Helped Google's ScholarlyAnurag Acharya Helped Google's Scholarly Leap. January 2005. Accessible at: http://www.indolink.com/SciTech/fr010305-075445.php. Accessed 26 September 2009.

2 Quint, B. Changes at Google Scholar: a conversation with Anurag Acharya information. Today, Inc, August 27, 2007. Accessible at: http://newsbreaks.infotoday.com/nbReader.asp?ArticleId=37309. Accessed 15 June 2009.

3 Butler, D. Science searches shift up a gear as Google starts Scholar engine. *Nature* 2004, **432**, 423.

4 Vine, R. Electronic resources reviews- Google Scholar. *JMIA* 2006, **94**, 97–99.

5 Henderson, J. Google Scholar: a source for clinicians? *CMAJ* 2005, **172**, 1549–1550.

6 Steinbrook, R. Searching for the right search – reaching the medical literature. *New England Journal of Medicine* 2006, **354**, 4–6.

7 Hoogendam, A., Stalenhoef, A. F., Robbe, P. F. & Overbeke, A. J. Analysis of queries sent to PubMed at the point of care: observation of search behavior in a medical teaching hospital. *BMC Medical Informatics and Decision Making* 2008, **8**, 42.

8 Dogan, R. I., Murray, G. C., Névéol, A. & Lu, Z. Understanding PubMed user search behavior through log analysis. *Database* 2009; article ID bap018, doi: 10,1093/database/bap018.

9 Bernstam, E. V., Herskovic, J. R., Aphinyanaphongs, Y., Aliferis, C. F., Sriram, M. G. & Hersh, W. R. Using citation data to improve retrieval from MEDLINE. *Journal of the American Medical Informatics Association* 2006, **3**, 96–105.

10 Herskovic, J. R., Tanaka, L. Y., Hersh, W. & Bernstam, E. V. A day in the life of PubMed: analysis of a typical day's query log. *JAMIA* 2007, **14**, 212–220.

11 Cosijn, E. & Ingwersen, P. Dimension of relevance. *Information Processing and Management* 2000, **36**, 533–550.

12 Cohen, J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 1968, **70**, 213–220.

13 Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bulletin* 1945, **1**, 80–83.

14 Hollander, M. & Wolfe, D. *Nonparametric Statistical Methods*. New York: John Wiley & Sons, 1973: 115–120.

15 Beel, J. & Gipp, B. Google Scholar's ranking algorithm: the impact of citation counts (an empirical study). Proceedings of the 3rd International Conference on Research Challenges in Information Science 2009a, 439–446.

16 Beel, J. & Gipp, B. Google Scholar's Ranking Algorithm: The impact of articles' age (an empirical study). Proceedings of the 6th International Conference on Information Technology: New Generations 2009b, 160–164.

17 Shultz, M. Comparing test searches in PubMed and Google Scholar. *JMIA* 2007, **95**, 442–445.

18 Mastrangelo, G., Fadda, E., Rossi, C., Zamprogno, E., Buja, A. & Cegolon, L. Literature search on risk factors for sarcoma: PubMed and Google Scholar may be complementary sources. *BMC Research* 2010, **3**, 131–134.

19 Anders, M. E. & Evans, D. P. Comparison of PubMed and Google Scholar literature searches. *Respiratory Care* 2010, **55**, 578–583.

20 Kulkarni, A. V., Aziz, B., Shams, I. & Busse, J. W. Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *JAMA* 2009, **302**, 1092–1096.

21 Bakkalbasi, N., Bauer, K., Glover, J. & Wang, L. Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomedical Digital Libraries* 2006, **3**, 7.

22 Lin, Y., Li, W., Chen, K. & Liu, Y. A document clustering and ranking system for exploring MEDLINE citations. *Journal of the American Medical Informatics Association* 2007, **14**, 651–661.

23 Tanaka, L. Y., Herskovic, J. R., Lyengar, M. S. & Bernstam, E. V. Sequential result refinement for searching the biomedical literature. *Journal of Biomedical Informatics* 2009, **42**, 678–684.

24 Whipple, E. C., McGowan, J. J., Dixon, B. E. & Zafar, A. The selection of high-impact health informatics literature: a comparison of results between the content expert and the expert searcher. *JMIA*. 2009, **97**, 212–218.

25 Falagas, M. E., Pitsouni, E. I., Malietzis, G. A. & Pappas, G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB Journal* 2008, **22**, 338–342.

26 Hölscher, C. & Strube, G. Web search behavior of Internet experts and newbies. *Computer Networks* 2000, **33**, 337–346.

Received *8 December 2010*; Accepted *3 May 2012*