Will Web Search Engines Replace Bibliographic Databases in the Systematic Identification of Research?

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Will web search engines replace bibliographic databases in the systematic identification of research?

Abstract

The availability of web search engines provides opportunities in addition to those provided by bibliographic databases for identifying academic literature, but their usefulness for retrieving research is uncertain. A rigorous literature search was undertaken to investigate whether web search engines might replace bibliographic databases, using empirical research in health and social care as a case study. Eight databases and five web search engines were searched between 20 July and 6 August 2015. Sixteen unique studies that compared at least one database with at least one web search engine were examined, as well as drawing lessons from the authors' own search process. Web search engines were limited in that the searcher cannot be certain that the principles of Boolean logic will apply and they were more limited than bibliographic databases in their functions such as exporting abstracts. Recommendations are made for improving the rigour and quality of reporting studies of academic literature searching.

Keywords

academic librarianship; bibliographic databases; Google Scholar; health research; literature searching; social care research; systematic searching; web search engines

Introduction

Academic librarians have the task of guiding higher education staff and students in their search for relevant materials. A particular demand in higher education is the retrieval of empirical research such as is required as part of research grant applications and research degree theses. As the volume of available information grows, the task of retrieving relevant research (and avoiding retrieving irrelevant research) becomes ever more demanding. The availability of the World Wide Web provides possibilities for efficient searching if it can be used effectively. The aim of the study reported here was to investigate whether web search engines have the potential to replace bibliographic databases for use in the systematic identification of research.

Although the lessons of the study were expected to be generalizable across higher education subject areas, to make a manageable project the focus was restricted to searching for empirical research and also by discipline. A search within the disciplines of health and social care was selected because this spanned both health disciplines where searching is well-established and a social science discipline where there is emerging interest although the methodologies are not well developed.

Although there is a considerable literature base on systematic searching in some disciplines, examination of the implications of utilising web search engines such as Google Scholar in the systematic identification of research is a more recent

development. Niu and Hemminger (2011) found that the academic scientists in their study were equally split in their preference for starting a search on the library web page or with Google. However, those using Google were less confident that their searches were returning all of the relevant information. The purpose of this study was to undertake a review of research that directly compared searching with web search engines against bibliographic database searching so as to give a sound knowledge base for use by librarians in advising staff and students in their searching endeavours.

The project involved three distinct stages:
- a systematic search using eight bibliographic databases and five web search engines (including Google Scholar) to retrieve articles that either evaluated or used at least one bibliographic database and one commercial search engine as part of a review or empirical study;
- appraisal of the retrieved relevant documents using a quality appraisal framework; and
- synthesis of findings and conclusions.

Methodology

The systematic review followed the PRISMA guidelines (Moher *et al.*, 2009) for the identification, screening and inclusion of relevant papers, which culminated in the following nine step process (Best *et al.*, 2014). See Figure 1 below.
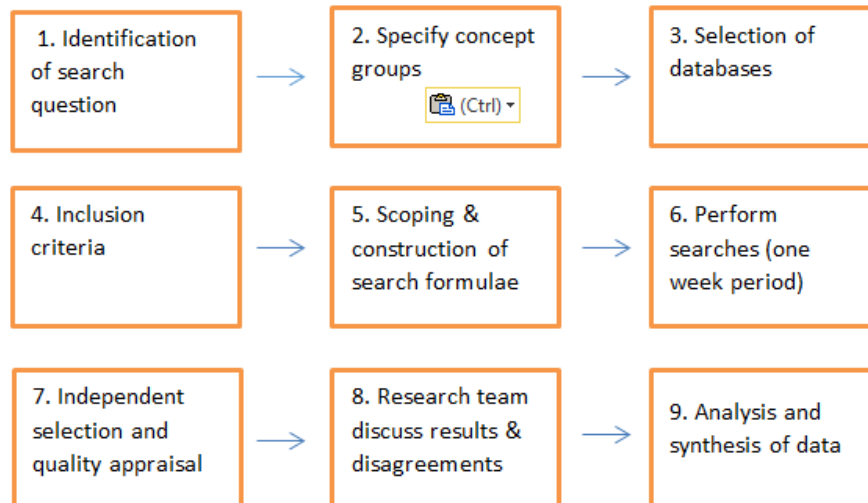
Figure 1: Nine-step process (adapted from Best et al., 2014)

The search question was defined as *'what is published in peer reviewed articles comparing the performance of bibliographic databases and web search engines for topics relating to health and social care?'* To achieve this, three main concept groups were identified (1) literature searching; (2) bibliographic databases and (3) search engines. These groupings formed the basic structure of searches across all databases (see Appendix). Health and social care was tested as a potential fourth concept group during piloting but was removed to increase sensitivity or recall and simplicity of search structure.

Selection of databases

The selection of appropriate databases is a key factor in the literature searching process (Lawrence, 2008). Eight bibliographic databases (Figure 2) were chosen ranging from inter-disciplinary databases, such as Scopus, to those focusing on health (Medline), social care (Social Services Abstracts) and information retrieval / library and information science (Library and Information Science Abstracts). The final selection of bibliographic databases was taken following a number of consultations with a subject librarian at the University and based on approaches established within previous research (Taylor *et al.*, 2007; McFadden *et al.*, 2012; Best *et al.*, 2014). Other databases considered, but removed following scoping exercises, were Communication Abstracts, PubMed and Social Care Online. Reasons for their omission included limited search sophistication, utility for topic and retrieval of articles were predominantly available through other databases.

- ASSIA (Applied Social Sciences Index and Abstracts), Platform: Proquest
- CINAHL Plus (Cumulative Index to Nursing and Allied Health Literature), Platform: EBSCOhost
- LISA (Library and Information Science Abstracts), Platform: Proquest
- Medline, Platform: OVIDSP
- PsycINFO, Platform: OVIDSP
- Scopus
- SSA (Social Services Abstracts), Platform: Proquest
- SSCI (Social Sciences Citation Index), Platform: Web of Science

Figure 2: List of databases used

Search engines were defined as websites that search public areas of the web in response to the terms that the user types in[1]. Five web search engines (Figure 3) were selected based on usage (traffic) as assessed using the following tools: Alexa, Reliablesoft and eBiz[2]. Each of these sites identified the same top four search engines: Google, Yahoo, Bing and Ask. Google Scholar was also included given its availability and use within academia.

- Ask
- Bing
- Google
- Google Scholar
- Yahoo

Figure 3: List of web search engines used

Inclusion and exclusion criteria

---

[1] A useful comparison of library databases and web search engines is provided by the University Library, University of Illinois at Urbana-Champaign (http://www.library.illinois.edu/ugl/howdoi/compare1.html, last accessed 31 October 2016).
[2] Alexa (http://www.alexa.com/topsites), Reliablesoft (https://www.reliablesoft.net/top-10-search-engines-in-the-world/), eBiz (http://www.ebizmba.com/articles/search-engines).

Inclusion criteria were defined as articles published in peer-reviewed journals, available in full text, in English language that evaluated or used bibliographic databases or web search engines to retrieve research. To be included, articles must have utilised and compared at least one or more bibliographic database against one or more web search engine for their literature search and focused primarily on a topic relevant to health and social care. Whether or not an article related to health and social care was assessed independently by two qualified health and social care professionals and was a major factor in bibliographic database selection. Given that Google Scholar was launched in November 2004, it was decided that articles published pre-January 2005 would not be included.

Search strategy

Where a database contained a subject index or thesaurus, it was agreed that the indexed term (if found) would be used in preference to user-generated 'free text' keywords. Where no subject index was available, user generated keywords were to be entered. The searches also employed the relevant truncation and wildcard characters for the selected databases, increasing the number of available search terms.

Most databases offered a 'sort by relevance' function. However, database suppliers rarely describe how 'relevance' is determined. Measures may include how frequently the term is present in the results, whether the term appears in a subject index, abstract, article title or author-supplied keywords or a number of other algorithms.

Despite these databases using automatic relevance sorting as standard, it was evident that some of the results appearing high up in the relevance sort, were there based on partial keywords from within a multi-word concept instead of as precise phrases. To overcome this, it was decided to include verbatim 'phrase-searching' on key terms, where necessary. In addition, the search strategy used more advanced proximity connectors to control more precisely the relationship between the key terms.

Medline and PsycInfo did not offer a 'sort by relevance' option for results. The OVID sensitivity and precision filters were not used, being primarily for clinical queries.

While most of the searches behaved in the expected fashion, PsycInfo gave rise to a particular oddity. The original search strategy for this database returned 1,611 articles. However, after a cursory review of results, it was clear that the majority of hits were returned largely as a result of the database retrieving its own name within the record. *Non-inclusion* of the term 'PsycInfo' (in preference to using the exclusion connector NOT) returned a more relevant set of results. This issue did not arise with any of the other databases searched.

The search strategies developed for the Scopus and LISA searches are included in the appendix as examples.

Constructing an effective search using the web search engines presented various challenges. Pilot searches were performed on each of the five search engines to increase familiarity and explore search options available. The non-academic types of material on the web also presented challenges for these searches. It was ascertained that academic papers more often use the term 'effectiveness' than 'effective', the latter term retrieving a plethora of self-help guides to database searching. As a result, the decision was made to use the term 'effectiveness' not 'effective' in the final search strategy

For consistency, it was decided a basic search would be performed in each web search engine. Web search engines are sensitive to long search strings (Stevenson *et al.*, 2015), for example, Google Scholar will truncate search strings after 256 characters (Boeker *et al.*, 2013). Thus attention was focused on finding the most directly relevant words with limited alternative meanings e.g. effectiveness OR sensitivity OR retrieval AND internet OR web OR search engine. From this learning a simplified, but comprehensive, search formula based on indexed terms and key words present within bibliographic searching was identified. This resulted in the development of the following search; effectiveness AND "bibliographic database" AND "search engine". As web-based search engines routinely retrieve an unmanageable number of hits, the decision was made to review and include articles present within the first 100 hits only. This approach is similar to that used in previous studies using web search engines to retrieve literature (McFadden *et al.*, 2012; Stevenson *et al.*, 2015). For searches involving web search engines the following points should also be noted.

- The searcher cannot be certain that the basic principles of Boolean logic will apply in web search engines, particularly when the order of concept groups is altered.
- None of these web search engines had a truncation facility as used on bibliographic databases (most have truncation for missing words, but not missing letters).
- Web search engines cannot export abstracts in the way that databases do.
- Google states that the AND operator may be assumed between terms; OR must be put in capitals.
- Yahoo has an Advanced Search function which is difficult to locate and was not found to be helpful in this exercise but may be explored in further studies.

Searches were performed from 20 July to 6 August 2015 by two members of the research team independently of each other. Following this process a third reviewer cross-checked selected articles against the pre-defined inclusion criteria.

Search results

Searches performed across all 13 databases/search engines returned 6,689 results. Each abstract was reviewed by two members of the research team using the inclusion criteria outlined above. This resulted in the identification of 16 unique articles (see Table 1). Scopus returned the highest number of relevant articles (n= 11), followed by

LISA (n=8), SSCI (n=8) and Medline (n= 6). Web search engines were broadly similar: Google Scholar, Bing and Yahoo returned 3 relevant hits each and Google and Ask 2 relevant hits.

Table 1: Search results

Database/Search Engine	Hits	Relevant	Sensitivity	Precision
Scopus	551	11	68.8%	2.0%
LISA	542	8	50.0%	1.5%
SSCI	656	8	50.0%	1.2%
Medline	1,536	6	37.5%	0.4%
CINAHL Plus	815	3	18.8%	0.4%
ASSIA	381	2	12.5%	0.5%
PsycINFO	1,611	2	12.5%	0.1%
SSA	97	1	6.3%	1.0%
Bing	100*	3	18.8%*	3.0%*
GoogleScholar	100*	3	18.8%*	3.0%*
Yahoo	100*	3	18.8%*	3.0%*
Ask.Com	100*	2	12.5%*	2.0%*
Google	100*	2	12.5%*	2.0%*
Total	6,689			

*16 unique hits; * = study relates to first 100 hits presented*

Sensitivity, precision and number needed to read (NNR)

Sensitivity is the fraction of the total number of relevant articles across all searches that are retrieved from one particular database/search engine (Taylor *et al*., 2015. Scopus (68.8%) recorded the highest sensitivity scores (see Table 1 above) of the bibliographic databases. The lowest sensitivity scores were found within ASSIA and PsycInfo (both 12.5%) and SSA (6.3%). Sensitivity scores for web search engines ranged from 18.8% (Bing, Google Scholar, and Yahoo) to 12.5% (Ask and Google).

Precision is defined as the proportion of relevant articles retrieved from all the results in a specific search (Montori *et al*., 2005), however in recent studies Number Needed to Read (NNR) has been additionally offered as a metric to aid understanding (Best *et al*., 2014). Within the bibliographic databases, Scopus recorded the highest precision score (2.0%) followed by LISA (1.5%) and SSCI (1.2%). The lowest precision score was found in PsycINFO (0.1%). Translating this into NNR scores, Scopus would require one to read 50 articles before finding a relevant one appear. For LISA and SSCI this would be 68 and 82 respectively. PsycINFO would require an individual to read 806 articles in order to find a relevant hit.

Precision scores for web search engines ranged from 3% (Bing, Google Scholar, and Yahoo) and 2% (Ask and Google). NNR scores for Bing, Google Scholar, and Yahoo would result in 33 articles needed to read, and Ask and Google would require 50. Note, the NNR scores for web search engines is based on the first 100 results.

Quality appraisal

In order to provide independent assessment of the quality of articles retrieved, the following framework was developed. It draws on the checklists and tools developed by Centre for Reviews and Dissemination (2009), McGowan *et al.* (2010), McGowan *et al.* (2016) Rader *et al.* (2014) and Shamseer *et al.* (2015). The components of the quality appraisal framework utilised in this study can be found in Figure 4 below.

---

1. SEARCH METHODOLOGY: Has reference been made to existing search methodology literature? Have the author(s) provided some theoretical/empirical support for their reasons for the search and search strategy?
2. SELECTION OF DATABASES: Have the author(s) given a reasonable explanation for the selected databases and web search engine(s)? Are the choices justified?
3. SEARCH STRATEGY AND FORMULAE: Are details of the complete search strategy for at least one of the searches provided? Is there evidence of tailoring to meet the search requirements of individual databases?
4. SEARCH PERFORMANCE: Are details provided on the performance of database and search strategy i.e. precision, sensitivity or other appropriate calculation(s)?
5. INCLUSION AND EXCLUSION: Are there clearly defined inclusion / exclusion criteria?
6. IDENTIFICATION OF RESULTS: Were retrieved items reviewed independently by a minimum of two researchers?
7. TIMEFRAME: Have the authors stated when they conducted their searches?
8. EXPERTISE: Was an information specialist(s) amongst the authors?

Figure 4: Quality appraisal framework for electronic literature searching (QAFELS)

The results of the quality appraisal are presented below.

1. SEARCH METHODOLOGY
Almost all of the articles made reference to existing search methodology literature and provided theoretical or empirical support for their search strategy. This was not evident however in the article by Carinci (2012), and was covered in a limited or partial manner in Jones (2005), and Tober (2011).

2. SELECTION OF DATABASES
The databases utilised in the studies included: Academic Search Elite; AgeLine; Applied Social Sciences Index and Abstracts (ASSIA); ArticleFirst; Basic BIOSIS; BIOSIS Previews; CAB Abstracts; Cochrane Library; Cumulative Index to Nursing and Allied Health Literature (CINAHL Plus); ECO; EconLit; EMBASE; GEOBASE; HighWire; Index to Theses; MEDLINE; PAIS International; POPLINE; ProQuest; PsycINFO; PubMed, ScienceDirect; SciFinder Scholar; Scopus; Social Care Online; Social Sciences Abstracts; Social Sciences Citation Index (SSCI); Social Services Abstracts; SocINDEX; SUMSearch; Web of Knowledge; Web of Science; WilsonWeb; ZETOC. The web search engines were: Google; Google Scholar; and Scirus (which no longer exists). With such a wide variety of databases available, it is important that authors justify their selection. In a small number of the articles, little or no effort was made to justify or explain the choice of databases (for example, Carinci, 2012). In

addition to specifying and explaining which databases were used and why, authors should also adhere to best practice guidelines and include information about the host server through which each database was accessed. According to the Centre for Reviews and Dissemination (2009: 249) the following elements should be included when reporting on systematic searches:

- the name of the database searched;
- the name of the database provider/system used;
- the date when the search was run;
- the years covered by the search; and
- the full strategy used (this can go in an appendix or supplementary file).

While it is understandable that Google Scholar dominated the choice of search engines (used in all sixteen studies), it is somewhat surprising that other additional search engines were not included. To enable more effective comparison of web searching versus use of bibliographic databases, interrogation of a wider range of popular web search engines would be valuable. It would also be useful to compare Google Scholar to other web search engines which only happened in three of the reviewed studies and in each case the second web search engine was Google.

Table 2: Databases and search engines used in the sixteen studies

Authors and date	Databases searched
Anders and Evans (2010)	PubMed; Google Scholar
Badgett *et al.* (2015)	PubMed; SUMSearch; Google; Google Scholar
Bramer *et al.* (2013)	PubMed; Google Scholar
Carinci (2012)	PubMed; Google; Google Scholar
Doldi and Bratengeyer (2005)	BIOSIS Previews; CAB Abstracts; Google; Google Scholar; Scirus
Freeman *et al.* (2009)	PubMed; Google Scholar
Jones (2005)	ArticleFirst; BasicBIOSIS; BIOSIS Previews; ECO; HighWire; MEDLINE; ProQuest; WilsonWeb; SciFinder Scholar; Google Scholar
Linder *et al.* (2015)	PubMed; Scopus; Web of Science; Google Scholar
McFadden *et al.* (2012)	Applied Social Sciences Index and Abstracts (ASSIA); Cochrane Library; Cumulative Index to Nursing and Allied Health Literature (CINAHL); Index to Theses; MEDLINE; PsycINFO; Social Care Online (SCO); Social Sciences Citation Index (SSCI); Social Services Abstracts (SSA); ZETOC; Google Scholar
Nourbakhsh *et al.* (2012)	PubMed; Google Scholar
Shariff *et al.* (2013)	PubMed; Google Scholar
Shariff *et al.* (2012)	EMBASE; MEDLINE; PubMed; Web of Knowledge; Google Scholar; Scirus

Tober (2011)	PubMed; ScienceDirect; Scopus; Google Scholar
Walters (2009)	Academic Search Elite; AgeLine; ArticleFirst; EconLit; GEOBASE; MEDLINE; PAIS International; POPLINE; Social Sciences Abstracts; Social Sciences Citation Index (SSCI); SocINDEX; Google Scholar
Walters (2007)	Academic Search Elite; AgeLine; ArticleFirst; GEOBASE; POPLINE; Social Sciences Abstracts; Social Science Citation Index; Google Scholar
Wright *et al.* (2014)	MEDLINE; Scopus; Web of Science; Google Scholar

3. SEARCH STRATEGY AND FORMULAE

There was variance across the studies in terms of the level of detail provided in relation to the search strategy. As the table below shows, all of the articles provided some level of explanation of the search strategy or the database(s) included in their respective studies. It should be noted that not all of the articles were reporting on advanced searches that involved the development of a complex search formula. In some cases the search comprised of a single keyword search, and others were known-title searches. In several cases detailed supplementary online files are included with the article, which show all of the search formulae that were used.

Table 3: Reporting of search strategy in the sixteen studies

Authors and date	Details of search strategy
Anders and Evans (2010)	Full details are provided in the article
Badgett *et al.* (2015)	Search strategy is discussed, with examples only provided
Bramer *et al.* (2013)	Full details are provided in a supplementary file
Carinci (2012)	Full details are provided in the article
Doldi and Bratengeyer (2005)	The search process is described, however there is insufficient detail of the search formulae used
Freeman *et al.* (2009)	The search process is described in detail, however the search formulae that were used are not included
Jones (2005)	A single search term was used for the searches in each of the databases
Linder *et al.* (2015)	Full details are provided in the article
McFadden *et al.* (2012)	The search process is described, including the concept groups used in the searches. Complete details are provided for one database
Nourbakhsh *et al.* (2012)	Full details are provided in the article
Shariff *et al.* (2013)	The search process is described in detail (with further information provided in supplementary files), however the search formulae that were used are not included
Shariff *et al.* (2012)	The search process is described. A title/author search was undertaken
Tober (2011)	The same two word search term was used for all databases
Walters (2009)	Full details are provided in an appendix
Walters (2007)	The search process is described. Title keyword searches were used (full details are not provided)
Wright *et al.* (2014)	This study focused on citation searching

In those studies that used more than one database we also looked to see if the searches were adapted and tailored to each individual database, where appropriate. This was evident in the studies by Doldi and Bratengeyer (2005); McFadden *et al.* (2012); and Walters (2009).

4. SEARCH PERFORMANCE

All of the articles provided some level of evaluation of database performance. Most often these involved calculations of sensitivity (recall) and precision, summarised in tabular form. However, in the process of appraising the sixteen studies it became clear that there was some variance in terms of how the authors defined or applied relevance measurements.

Table 4: Reporting of database performance / relevance measurements in the sixteen studies

Authors and date	Calculations regarding the performance of the database(s)
Anders and Evans (2010)	Recall; Full-text recall; Precision; Full-text precision
Badgett *et al.* (2015)	Mean precision
Bramer *et al.* (2013)	Coverage; Recall; Precision
Carinci (2012)	Relevance calculation based on criteria identified by the author
Doldi and Bratengeyer (2005)	Number of documents retrieved; Relevance (although not clear how this was calculated); Length of time search took
Freeman *et al.* (2009)	Number of citations located and web pages yielded; Characteristics of the first 100 citations; Target primary literature articles located by year within the first 100 citations
Jones (2005)	Number of articles retrieved; Number of articles retrieved within specified date ranges
Linder *et al.* (2015)	Sensitivity; Precision
McFadden *et al.* (2012)	Sensitivity; Precision
Nourbakhsh *et al.* (2012)	Yes, tables comparing the results and relevance calculations, as well as number of citations and impact factors
Shariff *et al.* (2013)	Coverage; Recall; Precision
Shariff *et al.* (2012)	Coverage; Proportion of articles not contained in PubMed that were contained in the other bibliographic databases
Tober (2011)	Recall; Precision; Importance
Walters (2009)	Recall; Precision
Walters (2007)	Citation coverage tables
Wright *et al.* (2014)	Sensitivity; Precision

Sensitivity (or recall) is the ratio of relevant items retrieved to all relevant items available on the topic. In practice it is it is usually presented as the "percentage of relevant articles relative to the total number of unique relevant articles when all search results were combined" (Linder *et al.*, 2015: 414). According to McFadden *et al.* (2012: 628), "sensitivity might be regarded as the 'completeness' of the database in terms of the fraction of the total number of relevant items potentially available that are retrieved by the search on that database" and as Wright *et al.* (2014: 4) state: "for a systematic review, a high level of sensitivity is required to ensure as few potentially relevant records as possible are missed." The majority of studies that included sensitivity or recall measurements adhered to this understanding. While Tober (2011) technically used the term correctly, recall was used as a synonym for number of 'hits' or results in each of the searches. Tober's usage reflects the situation where retrieval of an

article containing that word is considered a hit rather than where the article is reviewed to establish if it is relevant to the search topic.

Precision is the ratio of relevant items retrieved relative to all retrieved items. According to Anders and Evans (2010: 580) "A higher precision is desirable because it reflects a more specific, efficient search." The issue for the precision measurements in the studies was what they counted as the number of retrieved items, i.e. whether it was all retrieved items or the first 10, 50, 100 etc. It is worth examining the issue of precision in Google Scholar searching in more detail. There was considerable variance in the articles in relation to their findings regarding precision. A key reason for this is that different approaches to measuring precision were adopted. Bramer *et al.* (2013: 2) argue that it is important that "actual precision in GS [Google Scholar] should be calculated as the number of relevant references found in those first 1,000 references, divided by the number of hits that can actually be viewed, which is 1,000 at most." In contrast other studies (as noted below) measured the precision of web search engines in relation to a specified selected denominator such as 'the first 100 hits' which are, of course, sorted by relevance by the web search engine's (unknown) algorithm. The following methods and reported outcomes can be found in the articles identified in this study:

- Bramer *et al.* (2013) used practical precision which was calculated by the number of relevant hits in 1,000 results (with 1,000 as denominator), and the result was a precision value of 1.9%;
- McFadden *et al.* (2012) calculated precision for the first 100, 200, and 300 search results in Google Scholar and found that for the first 100 search results in Google Scholar the precision value was 10%, which was the same as the two databases, CINAHL and PsycINFO, which had the highest precision amongst the databases;
- Walters (2009) calculated the precision of Google Scholar and eleven other databases for the first 10, 20, 30, 40, 50, 75, and 100 hits, and for all search results. Google Scholar ranked 4th overall for the first 10 hits, but moved up to first place when the first 100 hits were examined in each of the search systems.
- Badgett *et al.* (2015) calculated precision for the first 50 results (Google and Google Scholar had the lowest precision across the four databases: 54% of searches performed by Google retrieved no high quality, relevant citations giving a median precision for Google of 0%);
- Shariff *et al.* (2013) calculated precision for both the first 40 results and for all the retrieved citations and found that the precision in Google Scholar for the first 40 results was 7.6% which was higher than PubMed (4.7%), but that when all citations were examined, the precision of Google Scholar decreased to 0.8%, compared to PubMed which was 6.0%.
- Tober (2011) considered only the first 10 search results in his precision calculations and had a unique formula for calculating precision based on the prevalence of the search terms in the title and abstract of the citation which gave a points-based score. Google Scholar had a mean precision value of 25 'points', behind ScienceDirect (35 'points') and Scopus (45 'points');
- The focus of the Wright *et al.* (2014) study was citation searching and they found that Google Scholar had a precision calculation of 0.48% (the average across all

the databases in their study was 0.5% and the highest precision was found in CENTRAL which was 1%).

Even bearing in mind the different methodologies for precision in the studies, the results from McFadden *et al.* (2012) and Walters (2009) may indicate that Google Scholar could offer better precision for social science searching than for the biomedical sciences, although this would need further investigation to test the argument. As studies such as those reviewed here accumulate, the knowledge base available to librarians and other information scientists will provide an increasingly robust evidence base for their work.

5.   INCLUSION AND EXCLUSION
Are there clearly defined inclusion / exclusion criteria (where appropriate)? Inclusion and exclusion criteria were provided in seven of the articles: Badgett *et al.* (2015), Bramer *et al.* (2013), Carinci (2012), Freeman *et al.* (2009), McFadden *et al.* (2012), Tober (2011), and Wright *et al.* (2014).

6. IDENTIFICATION OF RESULTS
Ensuring that multiple authors are independently reviewing the search results for relevance contributes to the rigour of the work (Centre for Reviews and Dissemination, 2009). In the majority of articles there was no explicit reporting of independent reviewing of search strategies, whether inclusion / exclusion criteria were met, or of the relevance of search results. This is an issue for subject searching moreso than known-item searching. There was evidence of a review process in Freeman *et al.* (2009), McFadden *et al.* (2012), Nourbakhsh *et al.* (2012), Shariff *et al.* (2013), Shariff *et al.* (2012), and Wright *et al.* (2014). According to the Cochrane Handbook (Higgins and Green, 2011, section 7.2.4): "Decisions about which studies to include in a review are among the most influential decisions that are made in the review process. However, they involve judgement. To help ensure that these judgements are reproducible, it is desirable for more than one author to repeat parts of the process. In practice, the exact approach may vary from review to review, depending in part on the experience and expertise of the review authors. … Using at least two authors may reduce the possibility that relevant reports will be discarded (Edwards 2002). It is most important that the final selection of studies into the review is undertaken by more than one author".

7. TIMEFRAME
Analysis of the articles for whether the date that searches were undertaken was included reveals an important weakness in reporting systematic searches in the literature as nine out of the sixteen articles did not provide information about when exactly the searches were executed. For the remaining seven that did provide information about when they conducted their searches two were published within one year of the search, two within two years, and one within three years. The article by Badgett *et al.* (2015) was reporting on searches undertaken in 2009, and in the article by Bramer *et al.* (2013) the searches were carried out over a six year period (2007-2013).

Eleven (69%) of the articles were published between 2010 and 2015, with the remaining five (31%) published between 2005 and 2009. Date of publication (and as discussed above, the date of the search) is important as the availability of different search tools has changed over time, as has their scope and size. Several authors (for example, Beller *et al.*, 2013) have discussed the issue of up-to-date systematic reviews for clinical practice, however this is also an issue when comparing the results of systematic searches for different databases and at different points in time. This is reflected in the PRISMA (Preferred Reporting of Items for Systematic reviews and Meta-Anaylses) statement of good practice for systematic reviews in healthcare: "At a minimum, for each database searched, authors should report the database, platform, or provider (such as Ovid, Dialog, PubMed) and the start and end dates for the search of each database. This information lets readers assess the currency of the review, which is important because the publication time-lag outdates the results of some reviews" (Liberati *et al.*, 2009). The Doldi and Bratengeyer (2005) article exemplifies the limitations of examining older articles on systematic searching in databases to inform current search behaviour as this study was carried out in 2004 prior to the launch of Google Scholar. Furthermore, the Scirus database, which was included in the Doldi and Bratengeyer (2005) study and the Shariff *et al.* (2012) study, was retired in 2014.

8. EXPERTISE

It was possible to identify the institutional and departmental affiliation for fourteen of the sixteen articles and whether the authors included one or more information specialists (librarian, information specialist or library/information academic). Eight of the articles had one or more authors that were information specialists. The value of including librarians and/or other information professionals in the systematic review process is highlighted in Yoshii *et al.* (2009), Hausner *et* al. (2012), Delaney and Bates (2015), Foster (2015), Koffel (2015) amongst others. The concept of embedded librarianship reflects the move towards including information specialists / librarians as full collaborators within the systematic review research team because of their knowledge of search systems and information retrieval expertise.

Findings

Five of the articles (Anders and Evans, 2010; Bramer *et al.*, 2013; Freeman *et al.*, 2009; Nourbakhsh *et al.*, 2012; and Shariff *et al.*, 2013) compared searches in Google Scholar and PubMed. There was consensus that Google Scholar could be used in conjunction with PubMed and other databases (Bramer *et al.*, 2013; Nourbakhsh *et al.*, 2012) or as an "adjunct resource" (Anders and Evans, 2010: 581), but that Google Scholar should not be viewed as "the definitive search engine for published medical literature" (Freeman *et al.*, 2009: 483). This was also borne out in the other studies that involved searches in Google Scholar and multiple databases, with the clear message that, "A range of databases is required for even modestly comprehensive searching" (McFadden *et al.*, 2012: 626). No other web search engines apart from Google, Google Scholar, and in two cases Scirus (a science-specific search engine that was in existence from 2001-2014) were used in any of the studies.

The limitations of Google Scholar that were identified in the articles are categorised and listed below.

Limitations of Google Scholar

Lack of advanced search features:
- Lack of filtering options, e.g. "it lacks a filter similar to PubMed's Clinical Queries to search for rigorously designed studies" (Anders and Evans, 2010: 581)
- Search terms cannot be mapped against the MEDLINE Medical Subject Headings (Anders and Evans, 2010; Badgett *et al.*, 2015).
- There are limited advanced search options for developing complex search strings or expressions and lack of a controlled vocabulary (Freeman *et al.*, 2009).

Lack of stability over time:
- By sorting results "by a mix of estimated importance and relevance" it can be difficult to trace developments over time in Google Scholar, for example, where the findings published in an article in a major journal have been challenged or disputed in a subsequent article in another journal (Badgett *et al.*, 2015: 2).
- It is difficult to replicate existing searches for systematic reviews in Google Scholar as it is not possible to "be certain of GS' [Google Scholar's] coverage at any specific point in time … Bibliographic databases such as PubMed offer additional database management dates next to publication dates. These databases keep track of content, and note when changes are made. In PubMed using the field mesh date ([mhda]), one can be rather certain what a search result would have been on a given date. GS [Google Scholar] lacks these management dates, and only offers publication dates" (Bramer *et al.*, 2013: 6).
- Google can change the functionality of Google Scholar, features can disappear (Bramer *et al.*, 2013).
- Lack of stability in Google Scholar search results, which are "too variable to ensure stable and reproducible results over time" (Carinci, 2012: 220).

Impact of Google cookies on searching:
- As reported by Badgett *et al.* (2015: 5), Google and Google Scholar searches should be undertaken on a computer or device that has been cleared of all Google / Google Scholar cookies, as these might alter the search results, for example, by putting those that are geographically closer further up the search results.

Lack of information about how Google ranks search results:
- The algorithms used by Google Scholar to sort the results are unavailable for scrutiny and can be revised by the company (Badgett *et al*, 2015).

Limitations in terms of size of search query, number of results, and detail provided in results:
- Search queries are limited to 256 characters (Bramer *et al.*, 2013).
- Only the first 1,000 results from a Google Scholar search are viewable, regardless of the stated number of search results (Bramer *et al.*, 2013).
- "Google Scholar provides relatively few abstracts" (Walters, 2007: 1130).
- "GS [Google Scholar] results pages list no more than four authors for any one article and no more than 99 characters of any article title" (Walters, 2007: 1125-6).
- "Google Scholar does not delete double hits from different databases" (Tober, 2011: 143).

Google Scholar coverage:

- Google Scholar may have less coverage of most recent research – according to Anders and Evans (2010: 582) "Google Scholar apparently receives delayed updates from MEDLINE, important, recent studies could be missed, further reducing sensitivity" and bias towards older literature, "Since Google Scholar places more relevance on articles that are cited most often, the citations located are reportedly biased toward older literature" (Freeman *et al.*, 2009: 482).
- "Lack of transparency of the database content" (Freeman *et al.*, 2009: 482).

This is not to say that the articles do not also highlight perceived limitations in database searching. Badgett *et al.* (2015: 2) point out that important articles in PubMed can be difficult to distinguish when results are sorted by date of publication. However this is now less of an issue, as in October 2013 PubMed introduced a sort by relevance option[3] (although it may the case that users of PubMed are unaware of this feature).

Different approaches were also apparent with regard to establishing full-text access to search results in Google Scholar, for example, Anders and Evans (2010) went through a university library link which provided full-text access where there was an institutional subscription to that journal, whereas Freeman *et al.* (2009) deselected the library link for their Google Scholar search to prevent the full-text of subscription-based items from being retrieved and likewise Shariff *et al.* (2013) only documented full-text articles that could be accessed without subscription or payment.

Discussion

In evaluating the sixteen studies the focus was not on evaluating the effectiveness of the search term choices, but on the search structure, complexity, reporting of the search strategy, and findings.

A number of studies were using a 'gold standard' set of articles to see if these could be retrieved from web search engines. While this approach is arguably justifiable when web search engines such as Google Scholar were in their infancy, they lack meaning and value when every article has a web presence.

Web search engines are mediocre on sensitivity when compared to databases. There is more work to do on precision. Precision scores were artificially high for the web search engines because only the first 100 results were considered. Further research into the precision of web search engines is required.

'Sort by relevance' functions open up new vistas for research on the topic, such as comparing these functions across databases and web search engines for example. It also opens up new possibilities for defining 'good enough rigour' for practical purposes. Other than for systematic reviews to the standard of the Cochrane and

---

[3] "The retrieved by relevance sort order for search results is based on an algorithm that analyzes each PubMed citation that includes the search terms. For each search query, "weight" is calculated for citations depending on how many search terms are found and in which fields they are found. In addition, recently-published articles are given a somewhat higher weight for sorting." (PubMed Help, http://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Displaying_and_Sorting_Your_S)

Campbell Collaborations, for reasons of time constraint and skill some measure of 'reasonableness' is used for such purposes as most PhD theses and research grant applications. This is normally considered in terms of how many (and which) databases to search, and the amount of time and specialist expertise used in constructing the search formulae. If all relevant databases (and, where appropriate, web search engines) have a sort by relevance function, then 'reasonable rigour' may also include some reference to the number of hits sorted by relevance that will be considered. One implication of this is that searchers could make a judgment call as to whether to sort by relevance and include only the first x number of results (as is the current practice with web search engines). So this is an additional decision for the search team – in addition to selecting appropriate databases and search engines they can also now determine how many results to include. This will involve a trade-off between precision and sensitivity.

Conclusion

The sixteen studies reviewed here each compared searching at least one bibliographic database with searching using at least one web search engine. It is apparent as a major finding that web search engines are not at a stage of development where they could replace bibliographic databases. Although their precision (when using only the first hundred hits sorted by relevance) was on a par with the best databases, their sensitivity in retrieving material relevant to an identified scholarly topic was only at the level of the weaker bibliographic databases for the search strategy reported here. Further research is required on using larger numbers of relevant hits retrieved by web search engines. The limited research to date suggests that this will substantially reduce precision for only modest gains in sensitivity (McFadden *et al.*, 2012).

In some instances, the studies reporting comparisons between web search engine and database searching were disappointing in terms of search details reported, presenting problems for replication. It is important for the development of knowledge in the field that future studies provide sufficient detail for replication.

In addition to drawing lessons from the findings in the sixteen retrieved studies, we can also learn from the search process undertaken in this study which included searching using five web search engines. In particular it is noteworthy that the searcher cannot be confident that the basic principles of Boolean algebra apply when using web search engines.

This paper has been written to assist scholars, academic librarians and information specialists in making best use of their time and resources when searching for literature. This study was limited to searching for health and social care research. However the methodology used here and the main lessons are generalizable across topic domains. Rapid reviews and other literature syntheses may require different approaches. As the volume of published material continues to grow exponentially it is more imperative than ever that approaches are developed – such as the example

outlined here – to search effectively and efficiently, and to learn from our endeavours in ways that will add greater rigour in future.

References

Anders, M.E. and Evans, D.P. (2010). Comparison of PubMed and Google Scholar literature searches. *Respiratory Care*, 55(5), 578-583.

Badgett, R.G., Dylla, D.P., Megison, S.D. and Harmon, E.G. (2015). An experimental search strategy retrieves more precise results than PubMed and Google for questions about medical interventions. *PeerJ*, 3: e913. https://dx.doi.org/10.7717/peerj.913. Last accessed 31 October 2016.

Beller, E., Chen, J., Wang, U. and Glasziou, P. (2013). Are systematic reviews up-to-date at the time of publication? *Systematic Reviews*, 2 (36). https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/2046-4053-2-36. Last accessed 31 October 2016.

Best, P., Taylor, B., Manktelow, R. and McQuilkin, J. (2014). Systematically retrieving research in the digital age: case study on the topic of social networking sites and young people's mental health. *Journal of Information Science*, 40(3), 346–356.

Boeker, M., Vach, W. and Motschall, E. (2013). Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. *BMC Medical Research Methodology*, 13(131). http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-13-131. Last accessed 31 October 2016.

Bramer, W.M., Giustini, D., Kramer, B.M. and Anderson, P. (2013). The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. *Systematic Reviews*, 2(115). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3882110/. Last accessed 31 October 2016.

Carinci, F. (2012). Implementation of the Tallinn Charter in the WHO European region: where is the evidence? a systematic comparison of online information before and after the end of the year of signature. *Health Policy*, 104(3), 215-221.

Centre for Reviews and Dissemination. (2009). *Systematic Reviews: CRD's guidance for undertaking reviews in health care.* https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf. Last accessed 31 October 2016.

Delaney, G. and Bates, J. (2015). Envisioning the academic library: a reflection on roles, relevancy and relationships. *New Review of Academic Librarianship*, 21(1), 30-51.

Doldi, L.M. and Bratengeyer, E. (2005). The Web as a free source for scientific information: a comparison with fee-based databases. *Online Information Review*, 29(4), 400-411.

Foster, M. (2015). An overview of the role of librarians in systematic reviews: from expert search to project manager. *Journal of EAHIL*, 11(3), 3-7.

Freeman, M.K., Lauderdale, S.A., Kendrach, M.G. and Woolley, T.W. (2009). Google Scholar versus PubMed in locating primary literature to answer drug-related questions. *Annals of Pharmacotherapy*, 43(3), 478-484.

Hausner, E., Wafffenschmidt, S., Kaiser, T. and Simon, M. (2012). Routine development of objectively derived search strategies. *Systematic Reviews*, 1(19). https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/2046-4053-1-19. Last accessed 31 October 2016.

Hawker, S.; Payne, S.; Kerr, C.; Hardey, M. and Powell, J. (2002). Appraising the evidence: reviewing disparate data systematically. *Qualitative Health Research*, 12 (9), 1284-1299.

Higgins, J. and Green, S. (Eds.) (2011). *Cochrane Handbook for Systematic Reviews of Interventions.* http://handbook.cochrane.org/. Last accessed 31 October 2016.

Jones, D.Y. (2005). Biology article retrieval from various databases: making good choices with limited resources. *Issues in Science and Technology Librarianship*, Fall 2005. http://www.istl.org/05-fall/refereed.html. Last accessed 31 October 2016.

Koffel J. (2015). Use of recommended search strategies in systematic reviews and the impact of librarian involvement: a cross-sectional survey of recent authors. *PLoS ONE*, 10(5). http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0125931. Last accessed 31 October 2016.

Lawrence D. W. (2008). What is lost when searching only one literature database for articles relevant to injury prevention and safety promotion? *Injury Prevention*, 14(6), 401-404.

Liberati, A., Altman, D., Tetzlaff, J., Mulrow, C., Gøtzsche, P., Ioannidis, J., Clarke, M., Devereaux, P., Kleijnen, J. and Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration, *BMJ*, 339, b2700. http://www.bmj.com/content/339/bmj.b2700. Last accessed 31 October 2016.

Linder, S.K., Kamath, G.R., Pratt, G.F., Saraykar, S.S. and Volk, R.J. (2015). Citation searches are more sensitive than keyword searches to identify studies using specific measurement instruments. *Journal of Clinical Epidemiology*, 68(4), 412-417.

McFadden, P., Taylor, B., Campbell, A. and McQuilkin, J. (2012). Systematically identifying relevant research: case study on child protection social workers' resilience. *Research on Social Work Practice*, 22(6), 626-636.

McGowan, J. and Sampson, M. (2005). Systematic reviews need systematic searchers, *Journal of the Medical Library Association*, 93(1), 74-80.

McGowan, J., Sampson, N. and Lefebvre, C. (2010). An evidence based checklist for the peer review of electronic search strategies (PRESS EBC), *Evidence based Library and Information Practice*, 5(1). https://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/7402/6436. Last accessed 31 October 2016.

McGowan, J., Sampson, M., Salzwedel, D., Cogo, E., Foerster, V., Lefebvre, C. (2016). PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *Journal of Clinical Epidemiology*, 75, 40-46.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. and The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement, *PLoS Medicine*, 6(7). http://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1000097. Last accessed 31 October 2016.

Montori, V.M., Wilczynsk, N.L., Morgan, D. and Haynes, R.B. (2005). Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ,* 330, 68. http://www.bmj.com/content/330/7482/68. Last accessed 31 October 2016.

Niu, X. and Hemminger, B. M. (2012). A study of factors that affect the information-seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*, 63(2), 336-353.

Nourbakhsh, E., Nugent, R., Wang, H., Cevik, C. and Nugent, K. (2012). Medical literature searches: a comparison of PubMed and Google Scholar. *Health Information and Libraries Journal*, 29(3), 214-222.

Rader, T., Mann, M., Stansfield, C., Cooper, C. and Sampson, M. (2014). Methods for documenting systematic review searches: a discussion of common issues. *Research Synthesis Methods*, 5(2), 98-115.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P. and Stewart, L. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*, 350, g7647. http://www.bmj.com/content/349/bmj.g7647. Last accessed 31 October 2016.

Shariff, S.Z., Bejaimal, S.A., Sontrop, J.M., Iansavichus, A.V., Haynes, R.B., Weir M.A. and Garg, A.X. (2013). Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. *Journal of Medical Internet Research*, 15(8), e164. https://www.jmir.org/2013/8/e164/. Last accessed 31 October 2016.

Shariff, S.Z., Sontrop, J.M., Iansavichus, A.V., Haynes, R.B., Weir, M.A., Gandhi, S., Cuerden, M.S. and Garg, A.X. (2012). Availability of renal literature in six bibliographic databases. *Clinical Kidney Journal*, 5(6), 610-617.

Stevenson, M., Taylor, B. and Knox, J. (2015). Risk in dementia care: searching for the evidence. H*ealth, Risk and Society,* 18(1-2), 4-20.

Taylor, B., Campbell, K. and McGlade, A. (2015) *Understanding and Using Research in Social Work.* London: Sage.

Taylor, B.J., Wylie, E., Dempster, M. and Donnelly, M. (2007). Systematically retrieving research: a case study evaluating seven databases, *Research on Social Work Practice*, 17 (6), 697-706.

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.

Walters, W.H. (2009). Google Scholar search performance: comparative recall and precision. *Libraries and the Academy*, 9(1), 5-24.

Walters, W.H. (2007). Google Scholar coverage of a multidisciplinary field. *Information Processing and Management* , 43(4), 1121-1132.

Wright, K., Golder, S. and Rodriguez-Lopez, R. (2014). Citation searching: a systematic review case study of multiple risk behaviour interventions. *BMC Medical Research Methodology*, 14(73). http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-14-73. Last accessed 31 October 2016.

Yoshii, A., Plaut, D., McGraw, K., Anderson, M. and Wellik, K. (2009). Analysis of the reporting of search strategies in Cochrane systematic reviews. *Journal of the Medical Library Association*, 97(1), 21-29.

Appendix

Generic search formula (boxes linked with "AND" operator)

(computer* NEAR/3 search*) OR (database* NEAR/2 search*) OR "database search methods" OR (literature NEAR/3 "information retrieval") OR "search strategies" OR "search strategy" OR "search term" OR "search terms" OR (systematic NEAR/3 search*) OR (systematic* NEAR/2 retriev*) OR "search effectiveness" OR "search efficiency" OR precision OR sensitivity OR NNR OR "number needed to retrieve" OR "number needed to read"

"Ask.com" OR Bing OR Google OR "Google Scholar" OR internet OR (search NEAR/2 engine*) OR "web-based" OR Yahoo

bibliographic database* OR (online NEAR/3 database*) OR (subscription NEAR/3 database*) OR ASSIA OR "Campbell Collaboration" OR Cochrane OR "Communication Abstracts" OR Embase OR Medline OR PsycInfo OR PubMed OR "Social Care Online" OR "Social Services Abstracts" OR "Web of Science"

Search strategy used in Scopus (20 July 2015)

Concept 1:
(computer* W/3 search*) OR (database* W/2 search*) OR "database search methods" OR (literature W/3 "information retrieval") OR "search strategies" OR "search strategy" OR "search term" OR "search terms" OR (systematic W/3 search*) OR (systematic* W/2 retriev*) OR "search effectiveness" OR "search efficiency" OR precision OR sensitivity OR NNR OR "number needed to retrieve" OR "number needed to read"

AND

Concept 2:
bibliographic database* OR (online W/3 database*) OR (subscription W/3 database*) OR ASSIA OR "Campbell Collaboration" OR Cochrane OR "Communication Abstracts" OR Embase OR Medline OR PsycInfo OR PubMed OR "Social Care Online" OR "Social Services Abstracts" OR "Web of Science"

AND

Concept 3:
"Ask.com" OR Bing OR Google OR "Google Scholar" OR internet OR (search W/2 engine*) OR "web-based" OR Yahoo

Search strategy used in LISA (20 July 2015)

Concept 1:
(computer* NEAR/3 search*) OR (database* NEAR/2 search*) OR "database search methods" OR (literature NEAR/3 "information retrieval") OR "search strategies" OR "search strategy" OR "search term" OR "search terms" OR (systematic NEAR/3 search*) OR (systematic* NEAR/2 retriev*) OR "search effectiveness" OR "search efficiency" OR precision OR sensitivity OR NNR OR "number needed to retrieve" OR "number needed to read"

AND

Concept 2:
bibliographic database* OR (online NEAR/3 database*) OR (subscription NEAR/3 database*) OR ASSIA OR "Campbell Collaboration" OR Cochrane OR "Communication Abstracts" OR Embase OR Medline OR PsycInfo OR PubMed OR "Social Care Online" OR "Social Services Abstracts" OR "Web of Science"

AND

Concept 3:
"Ask.com" OR Bing OR Google OR "Google Scholar" OR internet OR (search NEAR/2 engine*) OR "web-based" OR Yahoo

---

List of included studies

Anders, M.E. and Evans, D.P. (2010). Comparison of PubMed and Google Scholar literature searches. *Respiratory Care, 55(5),* 578-583.

Badgett, R.G., Dylla, D.P., Megison, S.D. and Harmon, E.G. (2015). An experimental search strategy retrieves more precise results than PubMed and Google for questions about medical interventions, *PeerJ 3*: e913. https://dx.doi.org/10.7717/peerj.913. Last accessed 31 October 2016.

Bramer, W.M., Giustini, D., Kramer, B.M. and Anderson, P. (2013). The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. *Systematic Reviews, 2*: 115. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3882110/. Last accessed 31 October 2016.

Carinci, F. (2012). Implementation of the Tallinn Charter in the WHO European region: where is the evidence? a systematic comparison of online information before and after the end of the year of signature. *Health Policy, 104(3),* 215-221.

Doldi, L.M. and Bratengeyer, E. (2005). The Web as a free source for scientific information: a comparison with fee-based databases. *Online Information Review, 29(4)*, 400-411.

Freeman, M.K., Lauderdale, S.A., Kendrach, M.G. and Woolley, T.W. (2009). Google Scholar versus PubMed in locating primary literature to answer drug-related questions. *Annals of Pharmacotherapy, 43(3)*, 478-484.

Jones, D.Y. (2005). Biology article retrieval from various databases: making good choices with limited resources. *Issues in Science and Technology Librarianship*, Fall 2005. http://www.istl.org/05-fall/refereed.html. Last accessed 31 October 2016.

Linder, S.K., Kamath, G.R., Pratt, G.F., Saraykar, S.S. and Volk, R.J. (2015). Citation searches are more sensitive than keyword searches to identify studies using specific measurement instruments. *Journal of Clinical Epidemiology, 68(4)*, 412-417.

McFadden, P., Taylor, B.J., Campbell, A. and McQuilkin, J. (2012). Systematically identifying relevant research: case study on child protection social workers' resilience. *Research on Social Work Practice, 22(6),* 626-636.

Nourbakhsh, E., Nugent, R., Wang, H., Cevik, C. and Nugent, K. (2012). Medical literature searches: a comparison of PubMed and Google Scholar. *Health Information and Libraries Journal, 29(3)*, 214-222.

Shariff, S.Z., Bejaimal, S.A., Sontrop, J.M., Iansavichus, A.V., Haynes, R.B., Weir, M.A. and Garg, A.X. (2013). Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. *Journal of Medical Internet Research, 15(8)*, e164. https://www.jmir.org/2013/8/e164/. Last accessed 31 October 2016.

Shariff, S.Z., Sontrop, J.M., Iansavichus, A.V., Haynes, R.B., Weir, M.A., Gandhi, S., Cuerden, M.S. and Garg, A.X. (2012). Availability of renal literature in six bibliographic databases. *Clinical Kidney Journal, 5(6)*, 610-617.

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.

Walters W.H. (2009). Google Scholar search performance: comparative recall and precision. *Portal: Libraries and the Academy, 9(1),* 5-24.

Walters, W.H. (2007). Google Scholar coverage of a multidisciplinary field. *Information Processing and Management, 43(4)*, 1121-1132.

Wright, K., Golder, S. and Rodriguez-Lopez, R. (2014). Citation searching: a systematic review case study of multiple risk behaviour interventions. *BMC Medical Research Methodology. 14:* 73. http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-14-73. Last accessed 31 October 2016.