# Google Scholar is not enough to be used alone for systematic reviews

**Dean Giustini[1], Maged N. Kamel Boulos[2]**

**1** UBC Library, UBC iSchool, Canada, **2** Plymouth University, UK

## Abstract

Background: Google Scholar (GS) has been noted for its ability to search broadly for important references in the literature. Gehanno et al. recently examined GS in their study: *'Is Google scholar enough to be used alone for systematic reviews?'* In this paper, we revisit this important question, and some of Gehanno et al.'s other findings in evaluating the academic search engine.

Methods: The authors searched for a recent systematic review (SR) of comparable size to run search tests similar to those in Gehanno et al. We selected Chou et al. (2013) contacting the authors for a list of publications they found in their SR on social media in health. We queried GS for each of those 506 titles (in quotes ""), one by one. When GS failed to retrieve a paper, or produced too many results, we used the allintitle: command to find papers with the same title.

Results: Google Scholar produced records for ~95% of the papers cited by Chou et al. (n=476/506). A few of the 30 papers that were not in GS were later retrieved via PubMed and even regular Google Search. But due to its different structure, we could not run searches in GS that were originally performed by Chou et al. in PubMed, Web of Science, Scopus and PsycINFO®. Identifying 506 papers in GS was an inefficient process, especially for papers using similar search terms.

Conclusions: Has Google Scholar improved enough to be used alone in searching for systematic reviews? No. GS' constantly-changing content, algorithms and database structure make it a poor choice for systematic reviews. Looking for papers when you know their titles is a far different issue from discovering them initially. Further research is needed to determine when and how (and for what purposes) GS can be used alone. Google should provide details about GS' database coverage and improve its interface (e.g., with semantic search filters, stored searching, etc.). Perhaps then it will be an appropriate choice for systematic reviews.

MeSH Keywords: Google Scholar, information retrieval, PubMed, searching, systematic reviews

Correspondence: mnkamelboulos@plymouth.ac.uk

## Introduction

Since its debut in 2004, Google Scholar (GS) has been viewed in the field of biomedical research as a flawed but useful tool in searching the scientific literature [1,2]. GS is widely-recognized as an excellent source of grey literature in biomedicine [3-5]. Despite its broad coverage, GS is considered ill-designed for expert searching [6]. One librarian said that *"… plug-in-the-keyword-and-hope-for-the-best tools like Google Scholar are poor choices for serious search questions such as clinical queries, bibliographic reviews, comprehensive literature searches, or other questions that require a more sophisticated approach"* [7]. Expert searchers were admonished to use trusted databases such as the Cochrane Library, PubMed and Embase when literature reviews were required (i.e., for grants, clinical trials and systematic reviews) [8]. The early buzz of GS eventually ebbed and was replaced by detailed comparisons against other tools such as PubMed and Scirus [9]. A consensus seemed to emerge that GS was not as current as PubMed and some expert searchers placed it a year behind or more [7]. Searchers also noticed that PubMed and Google Scholar fulfilled different purposes [10]. In head-to-head comparisons with curated databases, GS was deemed inadequate for subject searching and did not offer what expert searchers wanted to see in a literature database.

MEDLINE, produced by the US National of Medicine in Bethesda, Maryland, has been the gold standard for structured searching (especially via Ovid's Interface) for decades. While its place in biomedical searching seems secure, some researchers have argued that GS is a better choice for some retrieval queries, especially in browsing for articles and locating highly-cited papers [11]. In recognition of its speed and familiar interface, one editorial asked Google to think about creating a subset of GS for evidence-based medicine. But that would require transparency from Google about GS, and they were not about to produce a list of journal suppliers and grey literature that were crawled to create the database. Searchers were left to surmise its scope and make guesses as to what was in it [12].

Google Scholar is a useful tool to help researchers locate in seconds relevant papers from billions of pages across the Web [13] (and in many cases directly retrieve the full text of those papers). For that, it is highly-valued and useful, and every expert searcher should use it for that purpose. Allied to its easy-to-use interface, GS is a time-saver for quick searches especially compared to similar searches on PubMed, which can be unwieldy. In any case, knowing the strengths and weaknesses of GS will help researchers decide when and how to use it. Google has created a useful tool with links to articles and grey literature. But GS was already deemed unsuitable for literature reviews due to its limited search (filtering and qualifiers) functionality; its inability to draw on the power of the MeSH vocabulary (used in MEDLINE/PubMed) was cited as a critical flaw [14,15].

In 2013, French researchers, Gehanno et al., published a study that asked a simple question to which most expert searchers thought they knew the answer: *'Is Google scholar enough to be used alone for systematic reviews?'* [16] The authors state that GS' coverage has improved and ask whether its "coverage is high enough to be used alone in systematic reviews". In other words, the authors ask whether GS might replace MEDLINE and other bibliographic databases to perform costly, time-intensive searches for systematic reviews. The clearly-stated question and

conclusions of Gehanno et al. are examined in this paper; we ask whether Google Scholar has improved enough over the years to be used alone in systematic reviews.

## Methods

The authors searched for a systematic review that was comparable in size to Gehanno et al. We selected a recent study in our area of expertise (health/public health informatics), Chou et al. (2013), and contacted the authors for a list of the 506 publications they found in their SR on social media in health[1]. To test Google Scholar's ability to locate articles from an existing systematic review, we searched for all of the publications found by Chou et al. [17].

We tested whether the 500+ articles that formed the basis of Chou et al.'s SR were indexed by GS. Since we knew what we were looking for, and were not testing GS' ability to produce relevant documents, our searches were straightforward title searches. Chou et al. provided us with an Excel spreadsheet of the titles of papers (n=514) that comprised their systematic review. After correcting for minor errors, we looked for 506 unique items occurring either as simple citations or full-text links to papers within GS. We checked for the presence of these 506 publications by querying GS for the title of each study (in quotes ""), one by one. When a search failed to retrieve the required article, or produced too many results to browse, we opted to use Google's allintitle: command to increase our precision and search accuracy by limiting our search to the titles of articles. Some papers that were not found in GS were later searched and found in regular Google Search. Our results were double-checked title-by-title for completeness and accuracy against those listed by Chou et al.

Secondly, we tried to replicate Chou et al.'s search strategy and keywords (as detailed in [17]) in GS. We queried GS for: *health\* AND ("social media" OR "new media" OR "participatory media" OR "user-generated content" OR Facebook OR MySpace OR Twitter OR YouTube OR "Second Life" OR LinkedIn OR wiki\* OR blog\* OR "Web 2.0" OR "online social network" OR "social networking")*. We set query conditions as follows: *year range as 2004-2011; include citations*. It should be noted that Google uses stemming technology instead of asterisks, so those asterisks in the above query are ignored/not needed in GS.

Due to the different database structure and search syntax used in GS, our searches for these 506 papers using Chou et al.'s original search strategy and keywords yielded unmanageable results of approximately >750,000 items (as at 5 February 2013). Using Google's allintitle: command reduced our search results considerably to a collection of <450 items, but this was not a full subset of Chou et al.'s 506 items. Multiple attempts and combinations of keywords (and syntax) are needed if one were to find (discover) in GS the 476 (out of 506) papers cited in Chou et al.'s systematic review without already knowing their full titles.

---

[1] Other reasons Chou et al (2013) was chosen: we were able to contact the authors and obtain a full list of publications they found in their review. The list was a good representative size (about 500), and made a good test case and head-to-head challenge. Most importantly, it was comparable (in size) to the set of 700+ papers used in the Gehanno et al.'s study, which they pooled from 29 systematic reviews. We expect other researchers to try and replicate Gehanno et al.'s approach in their own fields, since GS coverage may vary by discipline.

## Results

Even though GS produced records for ~95% of those papers as cited by Chou et al. (n=476/506), numerous iterative searches were required to find all of them. In GS, we could not build search sets effectively or transfer results to a spreadsheet or reference manager. GS made our work more difficult as citations had to be managed one at a time. Due to its rudimentary structure, we could not run the search strings as used by Chou et al. in PubMed, Web of Science, Scopus and PsycINFO®. GS did not understand these expert search strings and was unable to translate them in any coherent way using its auto-correct feature. A few of the 30 papers that we could not find in GS (see Table I) were found in PubMed and even regular Google Search. Identifying each paper was inefficient, especially where two papers used similar keywords or metadata.[2] GS' ability to search into the full-text of papers combined with PageRank's algorithm is useful, and helps with browsing. These features on their own do not compensate for GS' obvious problems with searchability (discoverability) and database quality.

PubMed is clear that its database is built on a foundation of medical subject headings or MeSH terms, and each field in its 23 million citations is searchable. GS builds its structure on a simple interface design, vast interdisciplinary content and link popularity (which papers are cited most often). On the positive side, GS achieved a high percentage (95%) of "known-items" from Chou et al. but not all. Papers not found in GS were unique items from the four curated databases mentioned by Chou et al., PubMed, Web of Science, Scopus and PsycINFO® (Table I). GS is not flexible, precise or indexed (enough) to be used alone for systematic reviews. Its 'keyword search' capability, allied to Google's PageRank, is a poor replacement for controlled vocabulary searching and its interface does not provide enough flexibility to accommodate search filters, wildcards and expert search hedges, all of which are required for systematic reviews. We particularly noted the lack of a GS search filtering option to limit the scope of search results 'by discipline' such as 'health and medicine', since GS is catering for, and indexing articles from a very wide range of disciplines, and the same keywords can sometimes retrieve irrelevant, non-health-and-medicine-related articles.

In this modest study, identifying 506 papers among results and multiple screens was akin to searching for a needle in a haystack – painful, prickly and a time waster. Gehanno et al.'s search for 738 papers from 29 systematic reviews was similarly onerous but they, like us, knew what they were looking for [16]. This is a critical point in both studies: searching for known items is a much simpler exercise than trying to locate (or discover) those papers in the first place. GS' broad, undocumented corpus produces a lot of noise (and irrelevant hits) in its results, making it an unsuitable exclusive choice for systematic review searching.
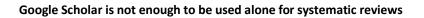
---

[2] Searching for papers in Google Scholar was not an easy task. After about 200 queries from the same machine, Google Scholar decided that our searches indicated we were bots and blocked our IP address. Clearing cookies only partially solved the problem: GS used captcha to solve each submitted query, so another IP address was needed to continue checking the remaining publications.
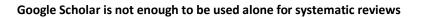
**Table I.** Articles missing in Google Scholar (as at 5 February 2013) and their original database sources where Chou et al. found them. The last two articles in the table were only (indirectly) retrievable when we dug deep in search results, but not as direct/first search result hits or via GS allintitle: command.

| Authors | Title | Source title | Year | Database where Chou et al. found article |
|---|---|---|---|---|
| White, J. | Everything you always wanted to know about stress (but were afraid to ask) or trying to reach the 'hard to reach' | Clinical Psychology Forum | 2011 | Scopus |
| Gannon, KE; Moreno, MA | Display of risk and protective health behaviors on incoming freshmen's Facebook profiles | Pediatric Research | 2009 | Web of Science |
| Horrigan, BJ | NIH and Wikimedia Foundation Collaborate to Improve Online Health Information | Explore-The Journal of Science and Healing | 2009 | Web of Science |
| Hwang, K; Etchegaray, J; Bernstam, E; Thomas, E | Predictors of intention to share educational health information via online social network ties | Journal of General Internal Medicine | 2010 | Web of Science |
| Pemu, PE; Quarshie, AQ; Josiah-Willock, R; Ojutalayo, FO; Alema-Mensah, E; Ofili, EO | Socio-demographic Psychosocial and Clinical Characteristics of Participants in e-HealthyStrides (c): An Interactive ehealth Program to Improve Diabetes Self-Management Skills | Journal of Health Care for the Poor and Underserved | 2011 | Web of Science |
| [No authors listed] | Why blog on about mental health? | Mental Health Today | 2006 | MEDLINE |
| Gronstedt, A. | Second Life produces real training results | T+D (Training + Development) | 2007 | Scopus |
| Hawn, C. | Report from the field: Take two aspirin and tweet me in the morning: How twitter, Facebook, and other social media are reshaping health care | Health Affairs | 2009 | Scopus |
| Malvey, D., Alderman, B., Todd, A.D. | Blogging and the health care manager | Health Care Manager | 2009 | Scopus |
| Russell, J. | Web 2.0 technology: How is it impacting your employer brand? | Nursing Economics | 2009 | Scopus |

| | | | | |
|---|---|---|---|---|
| Strongin, R. | Health reform in 140 characters | Medical Device and Diagnostic Industry | 2010 | Scopus |
| Tan, L. | Psychotherapy 2.0: MySpace® blogging as self-therapy | American Journal of Psychotherapy | 2008 | Scopus |
| [Anonymous] | Web 2.0, Health and Informatics | Methods of Information in Medicine | 2009 | Web of Science |
| Arikan, Y; Benker, T | Internet and Social Media Impacts on Turkish Healthcare Professionals' Reaching Health and Drug Side Effect-Related Information | Drug Safety | 2011 | Web of Science |
| Benker, T; Arikan, Y | Turkish Patients' Use of Internet and Social Media for Healthcare and Drug Side Effect Information | Drug Safety | 2011 | Web of Science |
| Botelho, R | Motivate healthy habits (part II): using web 2.0 & 3.0 technologies to generate social movements | Swiss Medical Weekly | 2009 | Web of Science |
| Evans, WD; McLeod, C; Thomas, SL | Social Media Marketing and Health Behaviours: Industry Strategies, Consumer Behaviours, and Public Health Responses | Annals of Behavioral Medicine | 2011 | Web of Science |
| Grinfeld, MJ; Hensel, BK; Cassidy, JT; Walker, SE; Parker, JC | A new media solution to coordination of care for juvenile arthritis: The JAHelp.org advocacy-oriented health care access project | Arthritis and Rheumatism | 2006 | Web of Science |
| Hamm, KM; Simeonov, IM; Heard, SE | Using Technology To Harness and Organize Expertise in the Development of Health Education Materials: How a Wiki Can Help You Collaborate | Clinical Toxicology | 2009 | Web of Science |
| Hartland, D; Duffton, R; Home, J; D'Aguilar, C; Berktay, L; Tomkinson, A; et al. | Health promotion (HP) and health outcomes: impacts of old and new media campaigns on referral patterns for HIV testing: implications for the National HIV Saving Lives Campaign | HIV Medicine | 2011 | Web of Science |
| Hartoonian, N; Ormseth, S; Bantum, EO; Owen, J | Process and outcome evaluation of a social-networking website for health promotion | Annals of Behavioral Medicine | 2008 | Web of Science |

| Kane, I; Walkosz, B; Giese, B | DOSOMETHINGONTHE.NET: Health Marketing for New Media | Annals of Behavioral Medicine | 2010 | Web of Science |
|---|---|---|---|---|
| Kondro, W | Health and environment blog | Canadian Medical Association Journal | 2011 | Web of Science |
| Nocker, G; Schachinger, A | Trends of future health communication and promotion via-Web 2.0 /Social Media | European Journal of Public Health | 2010 | Web of Science |
| Ojcius, D | Tracking public health via Twitter | Nature Reviews Microbiology | 2011 | Web of Science |
| Paek, HJ; Hove, T; Jeong, HJ; Kim, M | Peer or expert? The persuasive impact of YouTube public service announcement producers | International Journal of Advertising | 2011 | Web of Science |
| Toth-Cohen, S | The garden of healthy aging: collaborative project development in the virtual world of Second Life | Gerontologist | 2009 | Web of Science |
| Wapner, J | The healthy type - The therapeutic value of blogging becomes a focus of study | Scientific American | 2008 | Web of Science |
| Truccolo, I.; Bufalino, R.; Annunziata, M.A.; Caruso, A.; Costantini, A.; Cognetti, G.; et al. | National Cancer Information Service in Italy: An information points network as a new model for providing information for cancer patients | Tumori | 2011 | Scopus |
| Bastida, R | Use of collaborative web-based technology in mental health - Wiki use in practice | International Journal of Mental Health Nursing | 2008 | Web of Science |

## Conclusions

Is Google Scholar enough to be used for systematic review searching? No. Contrary to Gehanno et al.'s conclusions that GS "could even be used alone" [16], we found that GS was not up to the required search standard for systematic reviews. Despite its high sensitivity and vast coverage, GS was unable to locate all known-items cited in a previously-completed systematic review. We were able to retrieve most (but not all) of the papers used by Chou et al. [17] in their systematic review, because we already knew their titles and were searching for them one by one. But would we have been able to discover them as easily if we did not already know their exact titles? Based on our results, the answer was 'no' (when we tried to replicate Chou et al.'s search strategy in GS and queried GS for the topics of those papers [instead of their titles]). GS can sometimes be less precise than PubMed and similar bibliographic databases, returning hundreds or thousands of results, many of them irrelevant, thus requiring extensive human filtering of the results [5, 18].

Furthermore, GS' changing content, unknown updating practices and poor reliability make it an inappropriate sole choice for systematic reviewers. As searchers, we were often uncertain that results found one day in GS had not changed a day later and trying to replicate searches with date delimiters in GS did not help. Papers found today in GS did not mean they would be there tomorrow. In summary, GS could not be viewed on par with tools such as MEDLINE, Embase and the Web of Science. Gray et al. said it best that *"Google scholar's value to the sciences"* may be that it can be used *"for initial & supplemental information gathering"* [11].

Google Scholar's shortcomings, while not insignificant, should not exclude it from being used in systematic reviews [18]. On the contrary, we argue that further investigation is needed to determine when and how (and for what subjects, or disciplines) GS can be used for systematic review searching. Until then, its engineers should provide full details about its database coverage and aim to improve its interface search capabilities (e.g., indexing, semantic search filters, stored searching, etc.). Only then will it be equal to the demands of thorough, replicable searches as required by systematic reviews.

## Corresponding Author

Maged N. Kamel Boulos
Associate Professor of Health Informatics
Plymouth University, United Kingdom
Email: mnkamelboulos@plymouth.ac.uk

## References

1.  Henderson J. 2005. Google Scholar: a source for clinicians? *CMAJ*. 172(12), 1549-50. PubMed http://dx.doi.org/10.1503/cmaj.050404

2.  Giustini D. 2005. How Google is changing medicine. *BMJ*. 331(7531), 1487-88. PubMed http://dx.doi.org/10.1136/bmj.331.7531.1487

3.  Banks MA. 2005. The excitement of Google Scholar, the worry of Google Print. *Biomed Digit Libr*. 2(1), 2. PubMed http://dx.doi.org/10.1186/1742-5581-2-2

4.  Kousha K, Thelwall M. 2008. Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. *Scientometrics*. 74(2), 273-94. http://dx.doi.org/10.1007/s11192-008-0217-x

5.  Anders ME, Evans DP. 2010. Comparison of PubMed and Google Scholar literature searches. *Respir Care*. 55(5), 578-83. PubMed

6.  Shultz M. 2007. Comparing test searches in PubMed and Google Scholar. *J Med Libr Assoc*. 95(4), 442-45.  and. PubMed http://dx.doi.org/10.3163/1536-5050.95.4.442

7.  Vine R. 2006. Google Scholar. *J Med Libr Assoc*. 94(1), 97-99.

8.  Giustini D, Barsky E. 2005. A look at Google Scholar, PubMed, and Scirus: comparisons and recommendations. *Journal of the Canadian Health Libraries*. 26(3), 85-89. http://dx.doi.org/10.5596/c05-030

9.  Vine R. 2006. Google Scholar. *J Med Libr Assoc*. 94(1), 97-99.

10. Nourbakhsh E, Nugent R, Wang H. 2012. Medical literature searches: a comparison of PubMed and Google Scholar. *Health Info Libr J*. 29, 214-22. PubMed http://dx.doi.org/10.1111/j.1471-1842.2012.00992.x

11. Gray JE, Hamilton MC, Hauser A. Scholarish: Google Scholar and its value to the sciences. Iss Sci Tech Librarianship. 2012;70.

12. Walters WH. 2011. Comparative recall and precision of simple and expert searches in Google Scholar and eight other databases. Portal. *Libraries and The Academy*. 11(4), 971-1006.

13. Hightower C. 2010. Shifting sands: science researchers on Google Scholar, Web of Science, and PubMed, with implications for library collections budgets. *Issues in Science and Technology Librarianship*. 63, 76-94.

14. Neuhaus E, Asher A. 2006. The depth and breadth of Google Scholar: an empirical study. *Libraries and the Academy*. 6(2), 127-141. http://dx.doi.org/10.1353/pla.2006.0026

15. Jacsó P. 2005. Google Scholar: the pros and the cons. *Online Inf Rev*. 29(2), 208-14. http://dx.doi.org/10.1108/14684520510598066

16. Gehanno JF, Rollin L, Darmoni S. 2013. Is the coverage of Google Scholar enough to be used alone for systematic reviews. *BMC Med Inform Decis Mak*. 13(1), 7. PubMed http://dx.doi.org/10.1186/1472-6947-13-7

17. Chou WY, Prestin A, Lyons C, Wen KY. 2013. Web 2.0 for health promotion: reviewing the current evidence. *Am J Public Health*. 103(1), e9-18. PubMed http://dx.doi.org/10.2105/AJPH.2012.301071

18. Mastrangelo G, Fadda E, Rossi CR, Zamprogno E, Buja A, et al. 2010. Literature search on risk factors for sarcoma: PubMed and Google Scholar may be complementary sources. *BMC Res Notes*. 3, 131. PubMed http://dx.doi.org/10.1186/1756-0500-3-131