

Google Scholar and Author Manuscripts.

D. V. Klopfenstein¹, William Dampier^{2*},

1 School of Biomedical Engineering, Science, and Health Systems, Drexel University, Philadelphia, PA, USA

2 Department of Microbiology and immunology, Drexel University College of Medicine, Philadelphia, PA, USA

* wnd22@drexel.edu

Abstract

Google Scholar (GS) is the number one academic search engine. PubMed is popular in the fields of biology and chemistry as a source of full-text documents and for its search engine. It has a plethora of features recommended for academic search interfaces that GS lacks. We wrote this un-reviewed *author manuscript* to test if GS's web crawlers would find and add it to the GS search results.

Introduction

A comprehensive review of the literature is the foundation for all research. Modern literature reviews are primarily performed using on-line search engines [1] [2]. The two most popular free academic search tools are PubMed and Google Scholar [3].

Researchers worldwide are drawn to GS as their top choice for academic research [1] [3] [4] [5] because of its intuitive and familiar search interface [1] [6], a *Cited by N* link under every document result, a *Cite* link to download a document's citation to bibliographic management software such as EndNote for every document result, high citation counts, large literature coverage, and the researcher profile pages.

But the GS search interface has severe deficiencies that make literature searches laborious and, most importantly, unreproducible. However, many researchers are unaware of the drawbacks of GS [7]. For example, search results for a given query are dropped from one month to another [3] [8] [9] [10], with no documentation as to what has been dropped. Additionally there is no way to download full search results in bulk [7] [8], resulting in the need to click and click and click to page through possibly hundreds of search results, choosing either 10 or 20 results per page. In GS, a maximum of 1,000 search results can be seen in total for a given search query. Worrisome, the GS sorting algorithm is largely undocumented, so if the search results return over 1,000 searches, there is no way to know which results over the 1,000 maximum were excluded.

This un-reviewed author manuscript will be supplemental information for a different peer-reviewed paper submitted to a respected journal. In the peer-reviewed paper, we compare GS and PubMed.

Materials and methods

Using this paper as the materials, we followed the methods advocated in Google Scholar's *Inclusion* help page, which instructs individual authors who wish to have their un-reviewed, un-published work cited by Google Scholar to do the following:

1. Upload your paper in pdf format to your website.
2. Add a link to the paper from your publication page, publications.html

The requirements for the paper are:

- The paper must be a file with a *.pdf* extension.
- The title of the paper must be at the top of the first page using a large font.
- The authors must be listed immediately below the title on a separate line.
- The paper must have a bibliography section at the end called either *Bibliography* or *References*.

Results and Discussion

The GS help page states that their search robots normally find papers and include them within GS within several weeks. We will not know the results until then.

Author manuscripts in PubMed Central (PMC) have gotten a bad rap [11]. The belief is that *author manuscripts* are of low quality and have a role in propagating fake science [12]. The implication being that by providing a submission process for author manuscripts, PMC is creating a backdoor into PubMed that allows authors to skip quality checks like peer-review or evaluation for acceptance by a legitimate journal.

There is confusion regarding the difference between PubMed, MEDLINE, and PubMed Central (PMC) [12] [13], which unfortunately has resulted in the quality of the entire body of work covered by PubMed to be questioned by some. When critics claim that “PubMed is lacking in quality”, they usually mean to say, “PubMed Central (PMC) is lacking in quality.” PMC has standards and a review process that is overlooked by the critique. Klopfenstein and Dampier describe the PMC submission review process as:

If the journal which accepted the paper is not on PMC’s list of accepted open-access journals, but the work has been funded by the NIH or another approved funder, the papers fall under the NIH Public Access Policy open access mandate, which was drafted in 2004 and mandated in 2008. The public access policy mandates that any NIH-funded manuscript that is peer-reviewed and accepted for publication into a journal must be made publicly available in electronic format through the PMC.

A paper submitted through this path is called an *author manuscript*. Although it is called an *author manuscript*, individual authors are not allowed to directly submit their paper to PMC [14]. Rather, the author or thier publisher must submit their manuscript to the NIH Manuscript Submission system (NIHMS), which will process and deposit the manuscript into PMC. The numbers of *author manuscripts* submitted to PMC per year is small and have been declining each year since 2015. When published, the author manuscript is clearly marked in PMC with the text, *Author Manuscript*, running down the left margin and a banner at the top clearly stating the primary funding organization of the research.

[unpublished as of Jan 11, 2019, but will be submitted]

Conclusion

Despite the haters, PubMed is good. The **confusion** being sown by bloggers and editorial letter writers is being followed by expressions of doubt of PubMed's quality that is now seeping into peer-reviewed literature.

Confusing PubMed Central (PMC), the database, with PubMed, the tool set and search interface, due to their similar sounding names conflates two sets of citations that have vastly different sizes. PubMed, the tool set and search interface, indexes over 30.5 million articles. Only 2.5% of those are *author manuscripts* found in PubMed Central, the database. And those *author manuscripts* are vetted.

Confusion between GS's and PMC's standards and process for accepting *author manuscripts* is reaping havoc. It is GS, not PMC, that allows individual authors to publish *author manuscripts* of unquestioned quality. To be considered for submission into PMC, *author manuscripts* must be peer-reviewed, accepted into a journal, and funded by an approved entity. Additionally, PMC does not accept author papers, NIHMS does. NIHMS has a multi-step process where rejection of the paper is one possible outcome.

PubMed is not just good, it's great. Try it. You'll like it. Abandon all that clicking and clicking and paging and paging in GS. They are only using you to feed their data-gormandizing machine-learning algorithms. It is not worth the cheap empty sugar-high of juiced up citation counts, and the twisted pleasure of lurking your fellow scholars.

References

1. Hemminger BM, Lu D, Vaughan KTL, Adams SJ. Information seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*. 2007;58(14):2205–2225. doi:10.1002/asi.20686.
2. Duke L. *College libraries and student culture : what we now know*. Chicago: American Library Association; 2011.
3. Nicholas D, Boukacem-Zeghmouri C, Rodríguez-Bravo B, Xu J, Watkinson A, Abrizah A, et al. Where and how early career researchers find scholarly information. *Learned Publishing*. 2017;30(1):19–29. doi:10.1002/leap.1087.
4. Athukorala K, Hoggan E, Lehtiö A, Ruotsalo T, Jacucci G. Information-seeking behaviors of computer scientists: Challenges for electronic literature search tools. *Proceedings of the American Society for Information Science and Technology*. 2013;50(1):1–11. doi:10.1002/meet.14505001041.
5. Niu X, Hemminger BM. A study of factors that affect the information-seeking behavior of academic scientists. *Journal of the American Society for Information Science and Technology*. 2011;63(2):336–353. doi:10.1002/asi.21669.
6. Georgas H. Google vs. the Library: Student Preferences and Perceptions When Doing Research Using Google and a Federated Search Tool. *portal: Libraries and the Academy*. 2013;13(2):165–185. doi:10.1353/pla.2013.0011.
7. Boeker M, Vach W, Motschall E. Google Scholar as replacement for systematic literature searches: good relative recall and precision are not enough. *BMC Medical Research Methodology*. 2013;13(1). doi:10.1186/1471-2288-13-131.
8. Gusenbauer M, Haddaway NR. Which Academic Search Systems are Suitable for Systematic Reviews or Meta-Analyses? Evaluating Retrieval Qualities of Google

- Scholar, PubMed and 26 other Resources. *Research Synthesis Methods*. 2019;doi:10.1002/jrsm.1378.
9. Wichor Matthijs Bramer B. Variation in number of hits for complex searches in Google Scholar. *Journal of the Medical Library Association*. 2016;104(2). doi:10.5195/jmla.2016.61.
 10. Jamali HR, Asadi S. Google and the scholar: the role of Google in scientists' information-seeking behaviour. *Online Information Review*. 2010;34(2):282–294. doi:10.1108/14684521011036990.
 11. Anderson K. A Confusion of Journals — What Is PubMed Now?; 2017. [r~lhttps://scholarlykitchen.sspnet.org/2017/09/07/confusion-journals-pubmed-now/](https://scholarlykitchen.sspnet.org/2017/09/07/confusion-journals-pubmed-now/).
 12. Williamson PO, Minter CIJ. Exploring PubMed as a reliable resource for scholarly communications services. *Journal of the Medical Library Association*. 2019;107(1). doi:10.5195/jmla.2019.433.
 13. Hjørland B. Classical databases and knowledge organization: A case for boolean retrieval and human decision-making during searches. *Journal of the Association for Information Science and Technology*. 2014;66(8):1559–1575. doi:10.1002/asi.23250.
 14. I have published in a journal that does not appear in PubMed. Can I submit my manuscript to PubMed and/or PubMed Central to have it listed on your site?; 2020. <https://support.nlm.nih.gov/knowledgebase/article/KA-03415/en-us>.