

DISCUSSION PAPER/COMMENTARY

Supplement 2: The content of PubMed, MEDLINE, and PMC

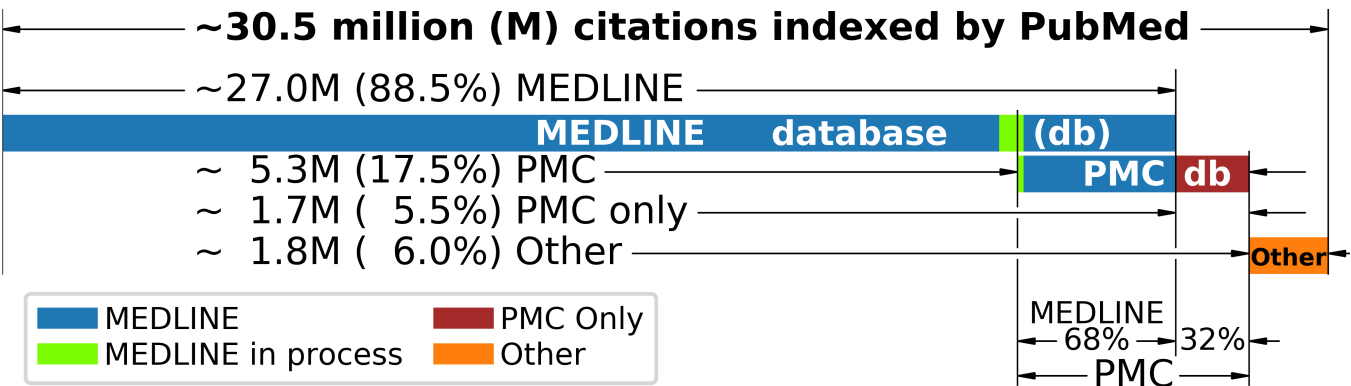
D. V. Klopfenstein<sup>1</sup> | Will Dampier\*<sup>2</sup>

<sup>1</sup>School of Biomedical Engineering, Science, and Health Systems, Drexel University, PA, USA  
<sup>2</sup>Department of Microbiology and immunology, Drexel University College of Medicine, PA, USA

**Correspondence**  
\*Will Dampier. Email: wnd22@drexel.edu

**Summary**  
PubMed is the free public interface to biomedical publication databases with the two largest databases being MEDLINE and PubMed Central (PMC) as shown in Fig 1, which is copied from the letter to the editor for easier reading in this supporting information. More information about the content of PubMed can be found on their *About PubMed* page which summarizes the content of PubMed as: “Citations in PubMed primarily stem from the biomedicine and health fields, and related disciplines such as life sciences, behavioral sciences, chemical sciences, and bioengineering.”

1 | THE CONTENT OF PUBMED, MEDLINE, AND PMC



**FIGURE 1 Content available through PubMed.** About 88.5% of the ~30.5 million citations accessible through PubMed are in MEDLINE (top dark blue bars) or are about to be added to (top light blue bars) the MEDLINE database (top dark and light blue bars). The MEDLINE papers that are free to access and are also indexed in the PMC database (middle light and blue bars) and comprise over 68% of papers indexed in PMC. About 5.5% of PubMed papers are only available in PMC (middle brown bar). Almost all of the remaining 6% of papers (bottom orange bar) are behind a paywall. To create this figure, we wrote a Python script, located in the *pmidcite* proof-of-concept project, that uses NCBI’s Entrez programming utilities<sup>1</sup> to query, download and use PubMed count data to create the figure.

<sup>0</sup>Abbreviations: ANA, anti-nuclear antibodies; APC, antigen-presenting cells; IRF, interferon regulatory factor

## 2 | MEDLINE

MEDLINE, shown as the top blue bars in Fig 1, is a highly selective database that was started in the 1960s and has journal articles dating back to 1946<sup>2</sup>. Journals that wish to have their articles indexed by MEDLINE must apply and be chosen for inclusion based on the high quality and originality of their articles' scientific content and the value of the science to the worldwide MEDLINE audience. Inclusion in MEDLINE gives researchers the ability to filter their query for peer-reviewed articles using *Medical Subject Headings* (MeSH), which include but are not limited to attributes related to funding sources, and genetic and chemical data, and clinical query filters for categories like diagnosis and clinical prediction guides.

## 3 | PMC

PubMed Central (PMC), shown as the middle bars in Fig 1, is not to be confused with PubMed, despite the similar names. PMC, begun in 2000, is an open-access archive for free full-text journals whose scope is biomedical and life sciences articles. To be considered for inclusion in PMC, an article must be free to access, accepted into a journal, and peer-reviewed.

## 4 | PMC AND MEDLINE

There is an overlap of the content of the MEDLINE and PMC databases with over 3.6 million or 68% (shown as the middle blue bars in Fig 1) of PMC articles also present or in process to be in MEDLINE. Articles often occur in both MEDLINE and PMC because the author opted to make their MEDLINE manuscript free to the public through the open-access model by paying an extra fee, rather than leaving it behind the journal's paywall.

## 5 | PMC, NOT MEDLINE

The 32% of PMC entries that are not also in MEDLINE (middle brown bar in Fig 1) are a small subset (about 5.5%) of the entirety of PubMed. McKeever et. al found that of the journals not indexed by MEDLINE, 23.3% are about to become indexed in MEDLINE, 15.3% have applied to be included and are being evaluated for consideration, 51.7% will remain not indexed <https://www.overleaf.com/project/5ea4985a15b938000162ebc9y> MEDLINE because they are either in non-MEDLINE journals or they are out of scope, and 9.7% are from an older version of MEDLINE covering papers from 1946 through 1965<sup>3</sup>.

If the journal in the non-MEDLINE group is on PMC's list of accepted open-access journals, the paper will either be automatically deposited to PMC or deposited if the journal is pay-walled, but has an option for an author to make their work open-access for a fee and the fee has been paid.

## 6 | PMC AND THE NIH PUBLIC ACCESS POLICY

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<sup>4</sup>. Rather, the author or their publisher must submit their manuscript to the NIH Manuscript Submission system (NIHMS), which will then 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.

## 6.1 | Other PubMed Content

Articles not indexed in either MEDLINE or PMC make up about 6% (bottom orange bar in Fig 1) of all content accessible through the PubMed interface. The most common reason (93%) articles fall into this category is that they reside behind a pay-wall. The remaining free articles include those not found in PMC-participating journals and articles like editorials or historical perspectives.

## References

1. Sayers EW, Agarwala R, Bolton EE, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 2018; 47(D1): D23–D28. doi: 10.1093/nar/gky1069
2. (NIH) USNLoM. MEDLINE, PubMed, and PMC (PubMed Central): How are they different?. <https://www.nlm.nih.gov/bsd/difference.html>; 2020. Accessed: 2020-01-04.
3. McKeever L, Nguyen V, Peterson SJ, Gomez-Perez S, Braunschweig C. Demystifying the Search Button. *Journal of Parenteral and Enteral Nutrition* 2015; 39(6): 622–635. doi: 10.1177/0148607115593791
4. (NIH) USNLoM. 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?. <https://support.nlm.nih.gov/knowledgebase/article/KA-03415/en-us>; 2020. Accessed: 2020-01-04.

**How to cite this article:** D. V. Klopfenstein and W. Dampier (2020), Commentary to Gusenbauer and Haddaway 2020: Evaluating retrieval qualities of Google Scholar and PubMed, *Research Synthesis Methods*, 2020;00:N–N.