Codifying Discrepancies among MEDLINE Platforms to Advance Instruction and Practice

C. Sean Burns [a], Tyler Nix [b], Robert M Shapiro, II [a] and Jeffrey T Huber [a]

[a] University of Kentucky, USA

[b] University of Michigan, USA

sean.burns@uky.edu, tnix@umich.edu, shapiro.rm@uky.edu, jeffrey.huber@uky.edu

ABSTRACT

MEDLINE is an essential bibliographic database for health professionals and is an indispensable part of clinical care. Despite that one of MEDLINE's main features is its use of MeSH, in practice, MEDLINE is offered on a number of platforms and each of these platforms takes very different approaches to using MeSH and in indexing other database fields, leading to variations in search results among the platforms. This project is based on a longitudinal study of these platforms with the goal of outlining how they differ and to provide guidelines for their use by instructors, students, and practitioners.

TOPICS

indexing; classification; database systems; information retrieval; curriculum

INTRODUCTION

MEDLINE is an essential, specialized bibliographic database for health and medical professionals. It is used as "a clinical care tool that health care professionals use to avoid adverse events, make changes to patient care, and answer clinical questions" (Dunn, Marshall, Wells, & Backus, 2017). MEDLINE and other sources are used to create products such as systematic reviews and these reviews are of higher quality when librarians and information specialists are involved in their production (Rethlefsen, Farrell, Trzasko, & Brigham, 2015). It seems that most systematic reviews are not created with the assistance of librarians or information specialists (Rethlefsen et al., 2015), which may be why systematic reviews as a research design are coming under increased criticism for being misleading, redundant, not useful, or statistically problematic (Annane, Jaeschke, & Guyatt, 2018; Chevret, Ferguson, & Bellomo, 2018; Elsner, 2018; Ioannidis, 2016). This also means that it is important to teach students in library and information science (LIS) programs how to use MEDLINE and other health, medical, and bioinformatic databases if librarians are to continue to contribute to and advance this type of research.

One of the main advantages that MEDLINE offers is the ability to search using Medical Subject Headings (MeSH). MeSH itself is organized as a tree-like thesaurus and contains sixteen branches or main categories (Chan & Salaba, 2016). MeSH headings may appear on a single branch of the tree or on multiple branches (U.S. National Library of Medicine, n.d.). This aspect of the MeSH thesaurus provides a way to increase search precision. In theory, this type of controlled vocabulary provides a complex, powerful, and predictable information retrieval device. However, in practice, it becomes complicated because MEDLINE is offered on multiple platforms including PubMed, ProQuest, EBSCOhost, Web of Science, and Ovid. Each of these platforms provides an unique interface to the MEDLINE records, and these interfaces offer different search functions and make different assumptions about how to search MeSH by default. For example, when searching a MeSH term in PubMed, the default is to automatically explode and include all narrower terms across all branches in the search results. However, on other platforms, like ProQuest, EBSCOhost, and Web of Science, terms are not automatically exploded. The implication is that queries that include only MeSH terms or MeSH terms in addition to other fields may return different records and/or different counts of records.

An additional complication is added because each platform indexes fields (author name fields, title fields, publication name fields, etc.) differently, and this may increase search result variation when comparing queries across platforms. For example, PubMed applies automatic term mapping (U.S. National Library of Medicine, n.d.), which means that uncontrolled terms are checked against multiple fields (or tables) when a query is executed. It is not always clear from the documentation how the other platforms index their fields. Another issue is that each platform offers different fields to search MEDLINE. For example, in PubMed's MEDLINE, users may select among five different Date fields, including Date-Completion, Date-Create, Date-Entrez, Date-MeSH, Date-Modification, Date-Publication, but in ProQuest's MEDLINE, there is only a single, less defined Date field.

Given the complexities introduced by the availability of multiple MEDLINE platforms, the purpose of this project is to examine why there are search result variations across the MEDLINE platforms in order to understand the differences between them and to produce clear guidelines for their use by students in LIS programs, for instructors teaching these platforms, and for practitioners using them in the field.

PROJECT BACKGROUND

We are conducting a longitudinal study (October 2018 – September 2019) of five MEDLINE platforms offered by PubMed, ProQuest, EBSCOhost, Ovid, and Web of Science. Our data includes 29 sets of search queries, five queries per set for each of the platforms, and is collected monthly. The searches are designed to be semantically and logically equivalent to each other on a per set basis. For example, one search set is designed to search each platform for the

MeSH heading *neoplasms* only without exploding the term. The search queries in the set differ only to match the syntax required by each of the platforms.

This paper will provide an overview of the longitudinal study by focusing on selected search sets in our study. One of these search sets includes a query that searches each platform for the MeSH term *neoplasms* AND the keyword *immune* AND limited by publication dates from 1950–2015. This query has produced inconsistent results across the platforms but the result counts are small enough (range of 169 to 184 records across the five platforms) to allow us to download the records and inspect them closely.

In a preliminary analysis, we have found that one reason why there are differences in search results for equivalent queries is because the platforms index their records differently. For example, for the search described above, EBSCOhost's MEDLINE returns 12 more records than PubMed's MEDLINE because EBSCOhost indexes the references lists and the term *immune* appears in the names of journal titles that have been cited by the retrieved record but not anywhere else (e.g., title, abstract, headings). It is debatable whether the appearance of a term in a journal title in a reference list makes EBSCOhost's approach to retrieve that record more or less relevant than PubMed's approach, but the implications are considerable when queries can retrieve thousands or more records and librarians, information specialists, and others need to sift through them to identify what is relevant. We will also discuss the search records for queries that only include MeSH terms on single and multiple branches. In these cases, we have found that some platforms take divergent approaches to these types of queries.

DISCUSSION

Each of these MEDLINE platforms is producing results inconsistent with the others, and it is not clear which platform is ideal for conducting systematic reviews or for providing an optimal "clinical care tool" (Dunn et al., 2017), even if such a thing as the latter exists. This research should lead to better guidelines on the selection and use of MEDLINE platforms and provide evidence-based support for instructors who teach these systems in LIS programs.

REFERENCES

Annane, D., Jaeschke, R., & Guyatt, G. (2018). Are systematic reviews and meta-analyses still useful research? Yes. *Intensive Care Medicine*, *44*(4), 512–514. https://doi.org/10.1007/s00134-018-5102-3

Chan, L. M., & Salaba, A. (2016). *Cataloging and classification an introduction*. Lanham, Md: Rowman & Littlefield Publishers, Inc.

Chevret, S., Ferguson, N. D., & Bellomo, R. (2018). Are systematic reviews and meta-analyses still useful research? No. *Intensive Care Medicine*, *44*(4), 515–517. https://doi.org/10.1007/s00134-018-5066-3

Dunn, K., Marshall, J. G., Wells, A. L., & Backus, J. E. (2017). Examining the role of MEDLINE as a patient care information resource: an analysis of data from the Value of Libraries study. *Journal of the Medical Library Association: JMLA, 105*(4), 336. https://doi.org/10.5195/jmla.2017.87

Elsner, B. (2018). Systematic reviews for informing clinical practice. *Physiotherapy Research International: The Journal for Researchers and Clinicians in Physical Therapy, 23*(1). https://doi.org/10.1002/pri.1703

Ioannidis, J. P. A. (2016). The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly, 94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210

Rethlefsen, M. L., Farrell, A. M., Trzasko, L. C. O., & Brigham, T. J. (2015). Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *Journal of Clinical Epidemiology, 68*(6), 617–626. https://doi.org/10.1016/j.jclinepi.2014.11.025

U.S. National Library of Medicine. (n.d.). Automatic Term Mapping [Training Material and Manuals]. Retrieved March 14, 2019, from https://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_040.html

U.S. National Library of Medicine. (n.d.). The MeSH Tree Structure (continued) [Training Material and Manuals]. Retrieved March 12, 2019, from https://www.nlm.nih.gov/bsd/disted/pubmedtutorial/015_025.html