2018-08-01

Design and implementation of a tool for conversion of search strategies between PubMed and Ovid MEDLINE

Wanner, AA

http://hdl.handle.net/10026.1/11989

10.1002/jrsm.1314
Research Synthesis Methods

Design and implementation of a tool for conversion of search strategies between PubMed and Ovid MEDLINE

Amanda Wanner, Primary and Community Care Research Group, Plymouth University, UK
Niki Baumann, College of Physicians and Surgeons of British Columbia, Canada

Abstract

Background

Both PubMed and Ovid MEDLINE contain records from the MEDLINE database. However, there are subtle differences in content, functionality, and search syntax between the two. There are many instances in which researchers may wish to search both interfaces, such as when conducting supplementary searching for a systematic review or using a previously published search strategy in a different interface, but little guidance on how to best conduct these searches.

The aim of this project is to describe differences in search functionality between Ovid MEDLINE and PubMed, provide guidance for converting search strategies between the two, and develop an easy to use, freely-available web-based tool to automate search syntax translations.

Case presentation

In this paper, we present a custom-built freely available online tool, Medline Transpose, to streamline the process of converting search strategies between Ovid MEDLINE and PubMed. With this tool, users can paste a strategy formatted for one interface into the search box, and immediately retrieve an output formatted for use in the other interface, with recommendations for changes that users can make to the strategy where an exact translation does not exist.

Conclusion

This novel approach has the potential to reduce time and errors that database users spend translating search strategies.

Keywords: systematic review; literature searching; software tool; MEDLINE; PubMed; databases, bibliographic; information science; information storage and retrieval

Introduction

Well-designed systematic reviews use comprehensive searching strategies and multiple database searches [1]. Systematic searches must balance sensitivity and precision such that the number of records to screen is manageable, but researchers can be sure that relevant articles are captured. MEDLINE is by far the most frequently used database for biomedical systematic review searches.

One recent study found that 98% of systematic reviews used the MEDLINE database in 2014, with the next most frequently occurring database (Embase) at 63% [2]. The Cochrane Handbook of Systematic Reviews of Interventions also requires use of the MEDLINE database for systematic review searches [3]. PubMed and Ovid are the two most popular interfaces for accessing the MEDLINE database, and most researchers utilise one of the two on a regular basis [4]. Several recent publications have evaluated the potential benefits and methods of searching both Ovid MEDLINE and PubMed for systematic reviews [5-10].

Both PubMed and Ovid MEDLINE search the MEDLINE database; however, there are differences in syntax and retrieval between the two [11]. Variances between use of Ovid MEDLINE and PubMed in systematic reviews have caused differences in citation retrieval [12], errors in search translation [13], and changed the number of included articles in systematic reviews [7]. Despite confusion in the literature about how to accurately convert a search strategy between the two interfaces properly, and the differences in records contained in each, there is currently no resource that provides this information in one place.

There have been several previous attempts to standardise and automate search translation between medical databases[14-17]; however, these tools all rely on a "find and replace" style of translation and only support a limited subset of database syntax, which provides a less accurate search translation. In this paper we present a custom-built tool that aims to capture the full range of syntax options within these two interfaces.

Medline Transpose tool

The newly developed tool, Medline Transpose, is a web-based application that provides a more accurate translation of searches than previously built tools. It aims to provide the most accurate search translations between Ovid Medline and PubMed, while supporting learning by providing further information where an exact translation is not possible, or where the translation may be unexpected. This allows users to make informed decisions about how to best translate a search when exact options are not available.

The Medline Transpose interface has three main components: a text-input box, radio selection buttons, and the results pane. The text input box can receive input pre-formatted for either Ovid MEDLINE, PubMed, or neither (i.e. with no field codes). Users can input either a single or multi-line search strategy. There are two radio button choices below the search box. The first allows the user to select whether the search translation should be from Ovid MEDLINE to PubMed or from PubMed to Ovid MEDLINE. The second allows the user to decide whether they prefer a more "accurate" or a more "efficient" translation. In the few cases where an "exact" translation would be awkward (as in the case of [tiab] into .ti,ab,cl,oa,kw,kf), this second set of options allows the user to choose whether they prefer a more exact translation or if an approximation would suffice. The results pane contains the translated search strategy, with warnings or suggestions where an exact translation is not possible.

Search translations used by Medline Transpose were developed by a thorough review of documentation of the field codes available in Ovid MEDLINE and the tags available in PubMed, using a chart to match up equivalent terms. The chart was organised into sub-sections for readability and is available on the Documentation page of the Medline Transpose website

(medlinetranspose.github.io). Where there were inconsistencies or ambiguities, the authors contacted database support for clarification and tested search terms between the two interfaces until the most accurate translation could be identified.

The Medline Transpose program was hand-coded by the authors using JavaScript. The project code is freely available on the open-source repository GitHub and the website is hosted by GitHub Pages [18]. The code is available to be used or modified under the terms of the MIT license [19]. The website design uses a modular and customisable framework package that is freely available from W3Schools [20].

When a user conducts a search in the interface, the program uses regular expressions to match patterns in the input to known syntax. Regular expressions allow the program to recognise a range of complex inputs that may be subtly different, and account for user input mistakes in some cases, for example if there are extra spaces in the search strategy or if the user accidentally inputs an en-dash or em-dash in place of a hyphen.

Stakeholder feedback was used to guide development at all stages of the project. During early development, a small group of health information professionals were contacted via snowball sampling to test the program. An online survey was used to gauge satisfaction with the website's functionality and interface design. Free text boxes were also used to elicit suggestions for feedback or improvement. This feedback was used to improve the program before its official launch in 2017. Further feedback from a conference presentation and Twitter launch were also used to guide development in the later stages.

Field codes between the two databases do not always neatly match up. In some cases, the naming is different (such as "keyword" in Ovid vs "other term" in PubMed), and in others there is no exact logical equivalent (such as the commonly used "ti,ab" search in Ovid). In another case, the documentation of field codes did not match real world usage, such as the "Create Date" and "Entrez Date" fields, where the fields have been swapped between the PubMed and Ovid interfaces (an unresolved issue as of the writing of this publication). Finally, there are several instances in Ovid in which commonly used syntax by users do not appear in the help documentation, such as the popular "or/1-2" syntax style of writing out Boolean operators in shorthand. These idiosyncrasies have been carefully collated and incorporated into the program where possible.

Case uses

There are many anticipated uses of the Medline Transpose tool, both for the casual searcher and within formal literature review projects. The following are suggested uses for different audiences.

Case use: finding unique PubMed content

Ovid MEDLINE and PubMed both search the MEDLINE database. However, there is additional content in PubMed that is not contained in the former, such as book chapters, PubMed Central journals that are not indexed in MEDLINE, and author manuscripts [21]. In one case study, a search strategy was originally created for Ovid MEDLINE to take advantage of its more sophisticated search capabilities (e.g. adjacency searching) and later run in PubMed to find any missing supplementary content. In 6 out of 7 systematic reviews that they tested, at least one additional study was found that met the review's inclusion criteria. These results, however, will need to be replicated in light of

Ovid MEDLINE's 2016 update to include Epub ahead of print citations [7]. Medline Transpose can be used to take a search strategy originally created for use in Ovid MEDLINE, and allow the researcher to find additional content in PubMed.

Case use: search filter translation

Search filters are pre-determined search strategies on a specific topic that have been tested based on their sensitivity and specificity in capturing articles on the desired topic [22]. Traditionally, search filters have been methodological [23], but topical filters, such as finding citations on a particular population group, are increasingly popular as well. Because search filters must be developed and tested within the context of a particular database, they are published using the corresponding database syntax. Although there are examples in which search filters have been published with both a PubMed and Ovid MEDLINE translation [24], many are not. Previous research has demonstrated the potential of user error when search strategies are translated by hand [13], so it is anticipated that use of Medline Transpose may reduce this user error by providing a more accurate search filter translation for the desired database interface.

Case use: mid-search interface switch

Occasionally, it becomes apparent midway through a search that features of the other interface would enhance the search strategy. For instance, if the searcher is searching PubMed for "minimal clinically important difference"[tiab] and then discovers that some researchers use similar phrases such as "minimum clinical important difference", they may wish to switch to the Ovid interface to make use of simultaneous phrase and truncation options, which cannot be combined in PubMed. A searcher may also wish to switch interfaces to take advantage of other tools such as the pharmacological action tag in PubMed, PubMed's "results by year" graph, PubMed's E-utilities API, or the adjacency operator in Ovid MEDLINE. Because most of the search has already been entered into one interface, the searcher can copy and paste into Medline Transpose, and then copy and paste the resulting strategies into the other interface. Quick interface switches are also useful when a temporary glitch occurs in the interface being used, which avoids having to wait for the glitch to be resolved.

Case use: use of online citation analysis tools that use PubMed's API

There are many websites that make use of the National Library of Medline's API, which allows users to interface with publicly available PubMed citation data. Some of the most popular sites include PubReminer[25], which allows users to analyse the metadata of their search results, or visualisations that show results pictorially[26]. These online programmes require a search in PubMed syntax, so Ovid users may find Transpose useful for analysing their search results, for example, to see which journals most often publish articles on their topic, to analyse yearly trends of article publication on their topic, or to see the MeSH terms most commonly associated with the articles they have found so far.

Discussion

One limitation of Medline Transpose is that many Ovid field codes or PubMed tags do not have an equivalent search in the other interface. If a similar search option is available, the user is provided with a warning or suggestion. If no equivalent is available, a warning is provided, so that the user can decide how they wish to adjust their search or strategy. Investigation is ongoing into the less-documented aspects of the Ovid and PubMed interfaces, with the goal of including as many

functionalities as possible. Next steps will include further testing and refinements of the conversion strategies as additional exceptions are identified.

Medline Transpose is a unique tool that can accurately and efficiently translate a search strategy between the Ovid MEDLINE and PubMed interfaces. It is freely available online at medlinetranspose.github.io. Its simple interface design allows users to insert a strategy from either of the two interfaces and immediately retrieve a translation, with additional information where an exact translation may not be possible. The tool also provides information about and translations for concepts that are less well known, such as subset filters. Initial feedback from usability testing with information professionals has been positive. It is our hope that regular use of the Medline Transpose tool can help information specialists to better understand the intricacies of the two interfaces and lead to more accurate and efficient searching behaviours, and contribute to higher quality searches in the published literature.

Acknowledgements

Works Cited

1. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 2009;6(7):e1000100.
2. Lam MT, McDiarmid M. Increasing number of databases searched in systematic reviews and meta-analyses between 1994 and 2014. *J Med Libr Assoc.* 2016;104(4):284.
3. Higgins JP, Green S. *Cochrane handbook for systematic reviews of interventions.* Version 5.1.0 [updated March 2011] ed: The Cochrane Collaboration; 2011.
4. De Groote SL, Shultz M, Blecic DD. Information-seeking behavior and the use of online resources: a snapshot of current health sciences faculty. *J Med Libr Assoc.* 2014;102(3):169.
5. Damarell RA, Tieman JJ, Sladek RM. OvidSP Medline-to-PubMed search filter translation: a methodology for extending search filter range to include PubMed's unique content. *BMC Med Res Methodol.* 2013;13:86.
6. Ruth Wong, Carlos Sharpin, Joanna Ashe, Claire Wallnutt, Gulhane L. Ahead-of-print publications: testing to find the best approach for guideline development at the National Clinical Guideline Centre (NCGC). NICE Joint Information Day 2016; 2016; London, UK.
7. Duffy S, de Kock S, Misso K, Noake C, Ross J, Stirk L. Supplementary searches of PubMed to improve currency of MEDLINE and MEDLINE In-Process searches via Ovid. *J Med Libr Assoc.* 2016;104(4):309-312.
8. Kaunelis D, Farrah K, Severn M. The Missing 2%: PubMed NOT MEDLINE. Poster presented at Canadian Health Libraries Association Conference; 2010; Kingston, ON.
9. Sampson M, de Bruijn B, Urquhart C, Shojania K. Complementary approaches to searching MEDLINE may be sufficient for updating systematic reviews. *J Clin Epidemiol.* 2016;78:108-115.
10. Thompson JC, Quigley JM, Halfpenny NJ, Scott DA, Hawkins NS. Importance and methods of searching for E-publications ahead of print in systematic reviews. *Evid Based Med.* 2016;21(2):55-59.

11.   Katchamart W, Faulkner A, Feldman B, Tomlinson G, Bombardier C. PubMed had a higher sensitivity than Ovid-MEDLINE in the search for systematic reviews. *J Clin Epidemiol.* 2011;64(7):805-807.
12.   Boeker M, Vach W, Motschall E. Semantically equivalent PubMed and Ovid-MEDLINE queries: different retrieval results because of database subset inclusion. *J Clin Epidemiol.* 2012;65(8):915-916.
13.   Falck-Ytter Y, Motschall E. New search filter for diagnostic studies: Ovid and PubMed versions not the same. *BMJ.* 2004;328:1040.
14.   Kamdar BB, Shah PA, Sakamuri S, Kamdar BS, Oh J. A novel search builder to expidite search strategies for systematic reviews. *Int J Technol Assess Health Care.* 2015;31(1-2):51-53.
15.   The University of Texas School of Public Health Library. Convert an Ovid Medline search to a PubMed search. 2014; https://sph.uth.edu/content/uploads/2014/09/convert_medline_to_pubmed.pdf.
16.   Bramer W. Improving efficiency and confidence in systematic literature searching. 2015; https://clinicallibrarian.wordpress.com/2015/06/29/eahil2015-improving-efficiency-and-confidence-in-systematic-literature-searching/.
17.   Centre for Research in Evidence-Based Practice at Bond University. Systematic Review Accelerator. http://crebp-sra.com/.
18.   GitHub. https://github.com/.
19.   Open Source Initiative. The MIT License. 2006; https://opensource.org/licenses/mit-license.php.
20.   W3Schools. W3.css Tutorial. https://www.w3schools.com/w3css/.
21.   National Library of Medicine. MEDLINE, PubMed, and PMC (PubMed Central): How are they different? [factsheet]. 2002; https://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html.
22.   Jenkins M. Evaluation of methodological search filters—a review. *Health Info Libr J.* 2004;21(3):148-163.
23.   McKibbon KA, Wilczynski NL, Haynes RB. Retrieving randomized controlled trials from medline: a comparison of 38 published search filters. *Health Info Libr J.* 2009;26(3):187-202.
24.   Montori VM, Wilczynski NL, Morgan D, Haynes RB. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ.* 2005;330(7482):68.
25.   Koster J. PubMed PubReMiner. 2004; http://hgserver2.amc.nl/cgi-bin/miner/miner2.cgi.
26.   Sperr E. Visualizing PubMed. https://esperr.github.io/visualizingpubmed/.

APPENDIX

Worked example a search translation from Ovid MEDLINE to PubMed

In this example, we demonstrate a possible use of Medline Transpose in a systematic search. We use a modified search from a previously conducted scoping review by one of the authors to demonstrate.

1. The initial search in Ovid MEDLINE retrieves 295 hits:

#	Searches	Results
1	(post-tuberculosis or posttuberculosis or posttubercular or post-tubercular or post-TB or treated tubercul* or treated pulmonary TB or treated pulmonary tuberculosis or treated tb).ti,ab,kw.	707
2	exp Bronchial Diseases/ or exp Granuloma, Respiratory Tract/ or exp Hemoptysis/ or exp Hypertension, Pulmonary/ or exp Lung Diseases, Fungal/ or exp Lung Diseases, Obstructive/ or exp Pulmonary Atelectasis/ or exp Pulmonary Fibrosis/ or exp Solitary Pulmonary Nodule/ or exp Cough/ or exp Dyspnea/ or exp Respiratory Insufficiency/ or exp Respiratory Hypersensitivity/ or Hemoptysis/	443412
3	(bronchitis or bronchiectasis or bronchospasm or bronchial or asthma* or aspergillosis or copd or coad or cpa or atelectasis or cough* or dyspnea or breathlessness or "shortness of breath" or wheez* or h?emoptysis or lung or respirat* or chest or pulmon* or breath* or airway or airflow or laryn*).ti,ab,kw.	1742449
4	1 and (2 or 3)	396
5	limit 4 to yr="1990 -Current"	318
6	remove duplicates from 5	295

2. The search (text from the middle column only) is then pasted into Medline Transpose, which returns the following results:

("post-tuberculosis"[tiab] OR "posttuberculosis"[tiab] OR "posttubercular"[tiab] OR "post-tubercular"[tiab] OR "post-TB"[tiab] OR treated tubercul*[tiab] OR "treated pulmonary TB"[tiab] OR "treated pulmonary tuberculosis"[tiab] OR "treated tb"[tiab])

"Bronchial Diseases"[mesh] OR "Granuloma, Respiratory Tract"[mesh] OR "Hemoptysis"[mesh] OR "Hypertension, Pulmonary"[mesh] OR "Lung Diseases, Fungal"[mesh] OR "Lung Diseases, Obstructive"[mesh] OR "Pulmonary Atelectasis"[mesh] OR "Pulmonary Fibrosis"[mesh] OR "Solitary Pulmonary Nodule"[mesh] OR "Cough"[mesh] OR "Dyspnea"[mesh] OR "Respiratory Insufficiency"[mesh] OR "Respiratory Hypersensitivity"[mesh] OR "Hemoptysis"[mesh:noexp]

("bronchitis"[tiab] OR "bronchiectasis"[tiab] OR "bronchospasm"[tiab] OR "bronchial"[tiab] OR asthma*[tiab] OR "aspergillosis"[tiab] OR "copd"[tiab] OR "coad"[tiab] OR "cpa"[tiab] OR "atelectasis"[tiab] OR cough*[tiab] OR "dyspnea"[tiab] OR "breathlessness"[tiab] OR "shortness of

breath"[tiab] OR wheez*[tiab] OR h?emoptysis[tiab] OR "lung"[tiab] OR respirat*[tiab] OR "chest"[tiab] OR pulmon*[tiab] OR breath*[tiab] OR "airway"[tiab] OR "airflow"[tiab] OR laryn*[tiab])

#1 AND (#2 OR #3)

#4 AND ("1990"[PDAT] : "2018"[PDAT])

In PubMed, [tiab] searches for title, collection title, abstract, other abstract, and author-supplied keywords. Therefore, the 'kw' field tag has been removed as it is redundant.

PubMed does not support wildcards in the middle of a term. For the most accurate translation, write out all possible word combinations.

3. There are two warnings. The first tells the author that there is no need to use a kw tag in PubMed, as the [tiab] command already searches this field. The second tells the author that the wildcard symbol used in the strategy needs to be replaced, as it is not supported in PubMed.
4. The author then manually replaces the term h?emoptysis[tiab] with all word variations: hemoptysis[tiab] OR haemoptysis[tiab].
5. Each line of the strategy is then manually inputted into the PubMed Advanced Search page. The search returns 295 results (on line #5):

#7	Search (#5 AND #9)	48
#6	Search ((pubstatusaheadofprint OR publisher[sb] OR pubmednotmedline[sb]))	2614790
#5	Search (#4 AND ("1990"[PDAT] : "2018"[PDAT]))	295
#4	Search (#1 AND (#2 OR #3))	372
#3	Search (("bronchitis"[tiab] OR "bronchiectasis"[tiab] OR "bronchospasm"[tiab] OR "bronchial"[tiab] OR asthma*[tiab] OR "aspergillosis"[tiab] OR "copd"[tiab] OR "coad"[tiab] OR "cpa"[tiab] OR "atelectasis"[tiab] OR cough*[tiab] OR "dyspnea"[tiab] OR "breathlessness"[tiab] OR "shortness of breath"[tiab] OR wheez*[tiab] OR hemoptysis[tiab] OR haemoptysis[tiab] OR "lung"[tiab] OR respirat*[tiab] OR "chest"[tiab] OR pulmon*[tiab] OR breath*[tiab] OR "airway"[tiab] OR "airflow"[tiab] OR laryn*[tiab]))	1639815
#2	Search ("Bronchial Diseases"[mesh] OR "Granuloma, Respiratory Tract"[mesh] OR "Hemoptysis"[mesh] OR "Hypertension, Pulmonary"[mesh] OR "Lung Diseases, Fungal"[mesh] OR "Lung Diseases, Obstructive"[mesh] OR "Pulmonary Atelectasis"[mesh] OR "Pulmonary Fibrosis"[mesh] OR "Solitary Pulmonary Nodule"[mesh] OR "Cough"[mesh] OR "Dyspnea"[mesh] OR "Respiratory	402563

	Insufficiency"[mesh] OR "Respiratory Hypersensitivity"[mesh] OR "Hemoptysis"[mesh:noexp])	
#1	Search (("post-tuberculosis"[tiab] OR "posttuberculosis"[tiab] OR "posttubercular"[tiab] OR "post-tubercular"[tiab] OR "post-TB"[tiab] OR treated tubercul*[tiab] OR "treated pulmonary TB"[tiab] OR "treated pulmonary tuberculosis"[tiab] OR "treated tb"[tiab]))	603

6. In this example we use Duffy et al's strategy[7] for finding additional unique content in PubMed (lines #6 and #7 above). This retrieves 48 potentially unique references.

7. We add the results from our Ovid MEDLINE and PubMed searches to EndNote X8, and de-duplicate to find unique references.
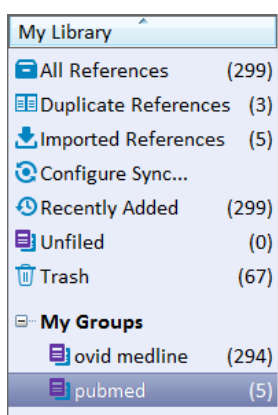
8. In this case we find 5 references unique to PubMed.