

The Missing 2%: PubMed NOT MEDLINE

David Kaunelis, MLIS; Kelly Farrah, MLIS; Melissa Severn, MSt, Canadian Agency for Drugs and Technologies in Health, Ottawa, Ontario, Canada

BACKGROUND

All but approximately 2% of PubMed citations are found in Ovid MEDLINE. Although PubMed contains more citations, MEDLINE offers more sophisticated search capabilities. Thus, for a complete and complex search, both databases are often searched.

PURPOSE

To establish a search filter for retrieving PubMed citations not found in Ovid's version of MEDLINE. The use of such a filter will avoid duplication when searching both sources.

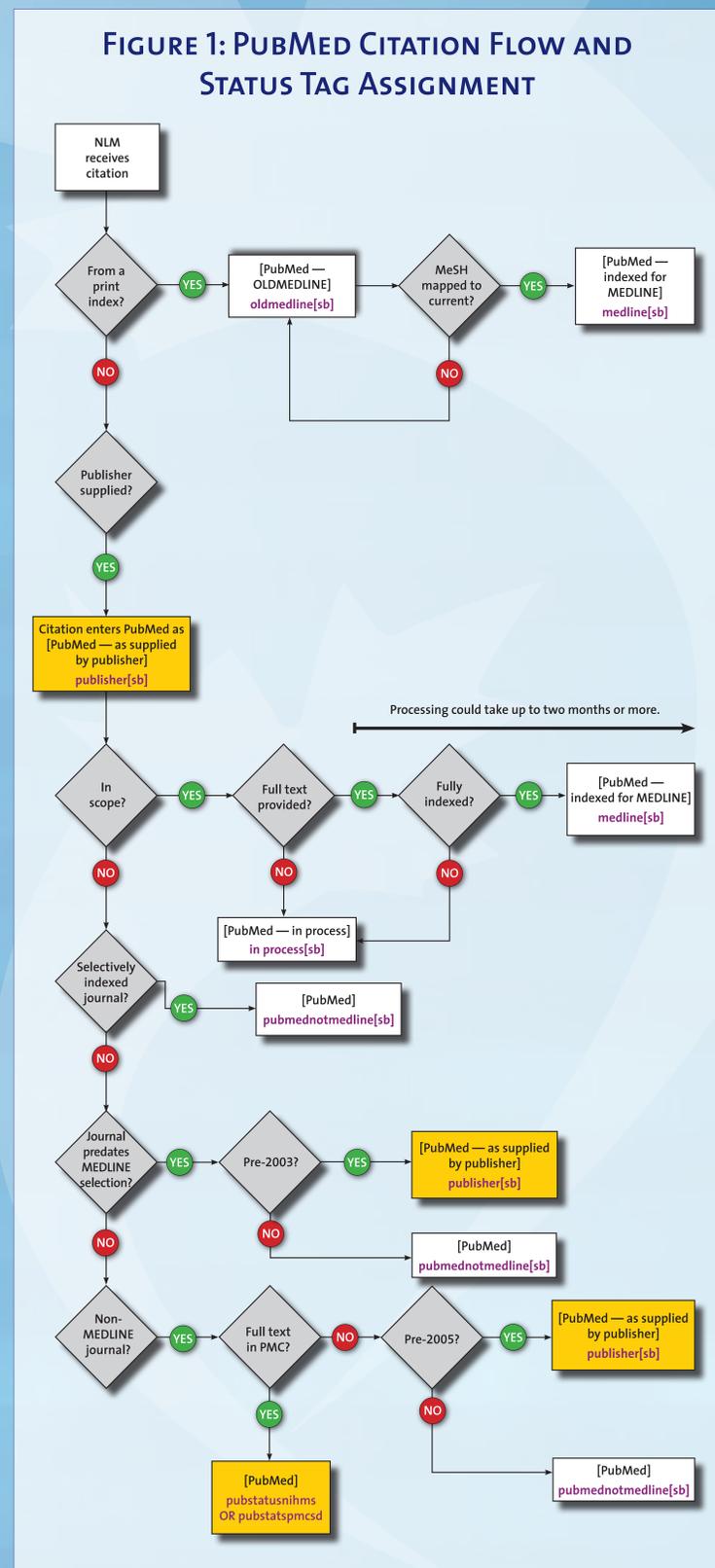
METHODS

Representatives from Ovid and the National Library of Medicine (NLM) were contacted to provide background information on the content found in MEDLINE and PubMed. Further information was gathered from web-based sources. Two search filters were developed. The first one was based on PubMed's status subsets, while the other filter was based on the PubMed entry date. Both filters were tested by a team of CADTH Information Specialists by comparing the results of identical searches in the two databases. Table 1 describes the various status tags.

PubMed Status Tag	How to Search in PubMed	Number of PubMed Citations (as of May 14, 2010)	Search Results in Ovid's MEDLINE
PubMed – as supplied by Publisher	publisher[sb]	534,303	Not available in MEDLINE. Most, but not all, citations with publisher status will eventually be available in MEDLINE
PubMed	pubstatusnihms*	7,119	Not available in MEDLINE
	pubstatuspmcsd*	770	Not available in MEDLINE
	pubmednotmedline[sb]	389,335	Available in PreMEDLINE
PubMed – in process	in process[sb]	345,581	Available in PreMEDLINE
PubMed – OLDMEDLINE	oldmedline[sb]	471,316	All citations available in the OLDMEDLINE database
PubMed – indexed for MEDLINE	medline[sb]	18,071,284	Almost all available in MEDLINE (except the current day or so)

NLM = National Library of Medicine; PreMEDLINE = MEDLINE® In-Process & Other Non-Indexed Citations.
*Included in publisher[sb] subset.

Figure 1 shows the many possible routes a citation can take once received by NLM. Each route ends with the citation receiving one of the status tags described in Table 1. However, a citation's status tag can change as it progresses through different stages of NLM's processing.



Yellow boxes represent missing 2%.
PMC = PubMed Central.

RESULTS

Two search filters are recommended to capture PubMed citations that are not found in Ovid's MEDLINE databases (including PreMEDLINE and OLDMEDLINE).

- A filter using the subset "publisher" was found to be effective in identifying the vast majority of unique PubMed citations.

Status subset filter	publisher [sb]
----------------------	----------------

- An additional search filter limiting results to those very recently added to PubMed was also found to be useful. This filter retrieves citations from the "in process" and "medline" subsets that may not have yet been uploaded into MEDLINE.

Date filter	(in process [sb] OR medline [sb]) AND "2010/05/31"[Entrez Date] : "3000"[Entrez Date]
-------------	---

While Ovid maintains that there are daily uploads from PubMed, testing indicated that it is advisable to search back two or three days to ensure that everything that is not yet available in Ovid's MEDLINE is retrieved. This, however, may result in a slight overlap between the MEDLINE and the PubMed search.

To limit the search to the past few days, use the [Entrez Date] tag, or go to the "Advanced Search" page and search by "Entrez Date."

Advanced Search

Search Builder

Entrez Date: 2010/05/31 to present Index

Format: YYYY/MM/DD. Month and days are optional

CONCLUSIONS

- The filters discussed in this poster were found to work well in identifying unique PubMed citations.
- However, there is no foolproof way to limit PubMed searches solely to citations not found in MEDLINE.
- The status tag for an individual citation is updated regularly, as it progresses through different stages of processing, making precise searching difficult.