Can PubMed Alone Capture the Evidence Base for Targeted Literature Reviews? A Case Study

MSR237

Priscilla WittKopf[1], Rajat Goel[2], Kassandra Schaible[3], Caroline von Wilamowitz-Moellendorff[1]

[1]Thermo Fisher Scientific, London, UK; [2]Thermo Fisher Scientific, Mumbai, India; [3]Thermo Fisher Scientific, Waltham, MA, USA

Background

- Targeted literature reviews (TLRs) are a cornerstone of evidence synthesis, particularly in health economics and outcomes research. Standard practice involves searching MEDLINE and Embase to enhance recall through complementary coverage. It is also important to distinguish databases from interfaces: Ovid MEDLINE and Ovid Embase are accessed via Ovid, while PubMed is a free interface primarily for MEDLINE. Key similarities and differences in coverage, indexing, and retrieval are summarized in **Table 1**.

Table 1. Comparison of Ovid Embase, Ovid MEDLINE, and PubMed

Feature	Ovid MEDLINE	Ovid Embase	PubMed
Access	Subscription	Subscription	Free
Underlying Database	MEDLINE	Embase	MEDLINE + additional records (e.g., in-process, ahead-of-print, PubMed-not-MEDLINE)
Indexing Vocabulary	MeSH	Emtree	MeSH (automatic term mapping)
Search Functionality	Advanced (controlled vocabulary, adjacency operators, field-specific searching)	Advanced (Emtree + proximity operators, detailed drug/ device indexing)	Simpler interface; automatic term mapping; limited proximity searching
Conference Abstracts	Limited	Extensive	Limited
Content Coverage	Core biomedical journals	Broader international and pharmacological coverage	MEDLINE + additional publisher-supplied and non-indexed records
Key Strength	Structured, reproducible search strategies	Comprehensive coverage	Accessibility

Abbreviation: MeSH = medical subject heading

- Recent advancements in artificial intelligence (AI) have transformed literature screening workflows, enabling semi-automated or fully assisted study selection processes. However, the integration of AI tools introduces new operational considerations, including database accessibility and licensing constraints. PubMed is unique in offering unrestricted programmatic access and compatibility with external AI models, making it an attractive option for AI-assisted systematic workflows.[1,2]
- Despite these advantages, the implications of relying solely on PubMed for TLRs remain unclear. Differences in indexing, database scope, and filter behavior may affect study retrieval and downstream evidence synthesis. This is particularly relevant when pragmatic filters—designed for real-world applicability rather than strict methodological sensitivity—are applied.
- This study aimed to compare PubMed-only search strategies with combined MEDLINE plus Embase searches in the context of AI-assisted TLRs, focusing on retrieval overlap, database-specific contributions, and the impact of filter design.

Methods

Study Design

- This methodological comparative study evaluated whether PubMed alone could capture the evidence base for TLRs compared with a combined database approach. Two parallel search strategies were conducted:
  - MEDLINE and Embase via Ovid
  - PubMed using adapted search syntax
- All database searches in MEDLINE and Embase were conducted before November 2024.
- The primary outcomes were:
  - The proportion of included studies identified by both approaches (overlap)
  - The number and proportion of studies uniquely identified by each approach

Targeted Literature Reviews

- Two TLRs were conducted using protocol-driven methodologies with predefined research questions, inclusion and exclusion criteria, and search strategies. TLR 1 focused on cost-effectiveness studies and TLR 2 focused on disease burden and observational studies.
- For each TLR, citations retrieved from database searches were supplemented by targeted keyword searches and citation-chasing of relevant systematic reviews.
- Search Strategy Development and Translation: Search strategies were initially developed for MEDLINE and Embase (Ovid) using a combination of controlled vocabulary (medical subject heading [MeSH] and Emtree terms) and free-text keywords. Strategies were translated for PubMed through manual adaptation of syntax, field tags, and controlled vocabulary mappings (e.g., Emtree to MeSH and text-word equivalents).

Methods (continued)

Data Analysis

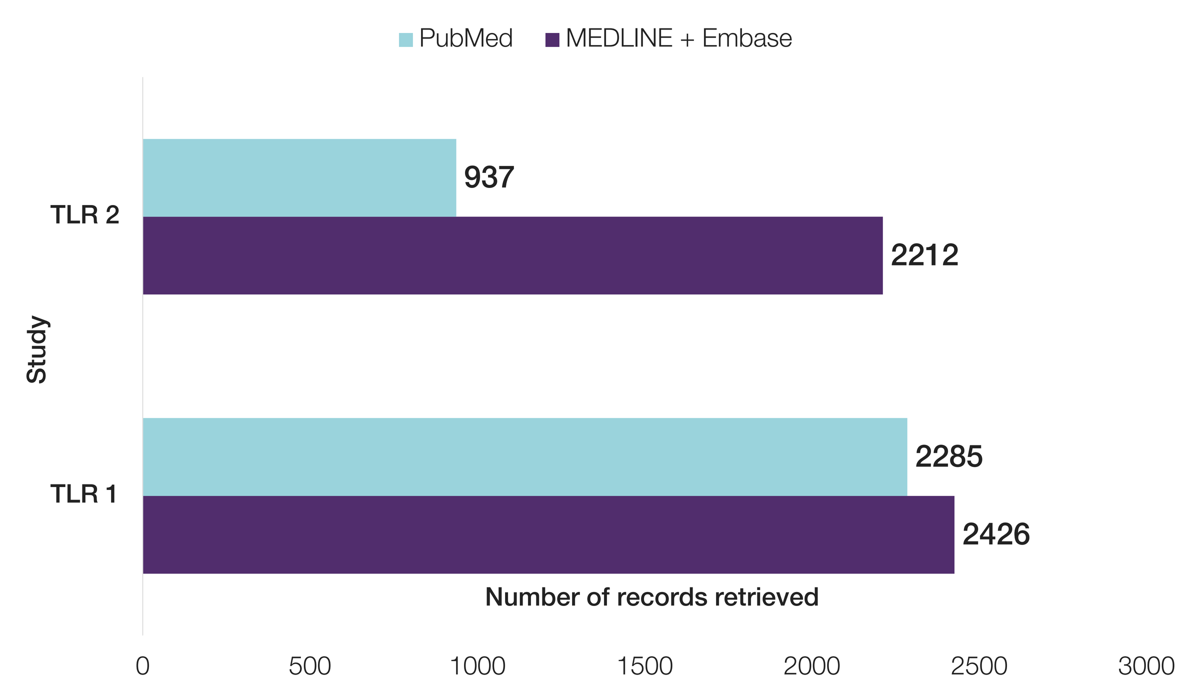
- For each TLR, the following were calculated:
  - Number of included studies identified by each search approach
  - Number and proportion of overlapping studies
  - Number and proportion of studies unique to each approach
- These outcomes were used to assess the extent to which PubMed alone could replicate the evidence base identified through the combined MEDLINE and Embase approach.

Results

Search Results and Study Selection

- Across both TLRs, the combined MEDLINE and Embase (Ovid) searches consistently retrieved a greater number of records compared with PubMed alone (**Figure 1**).
- For TLR 1 (cost-effectiveness; validated economic filter), PubMed retrieved 2,285 records compared with 2,426 from the combined MEDLINE and Embase search, with 26 studies meeting the inclusion criteria.
- For TLR 2 (observational studies; pragmatic design filter), PubMed retrieved 937 records compared with 2,212 from the combined MEDLINE and Embase approach, with 25 studies meeting the inclusion criteria.
- Following screening and application of eligibility criteria, final included study counts demonstrated substantial but incomplete overlap between search approaches.

Figure 1. Study Retrieval Comparison



Abbreviation: TLR = targeted literature review

AI-assisted Screening Performance

- The AI-assisted screening model demonstrated consistent performance across both database approaches despite differences in record pools (**Figure 2**).

Figure 2. Model Parameters

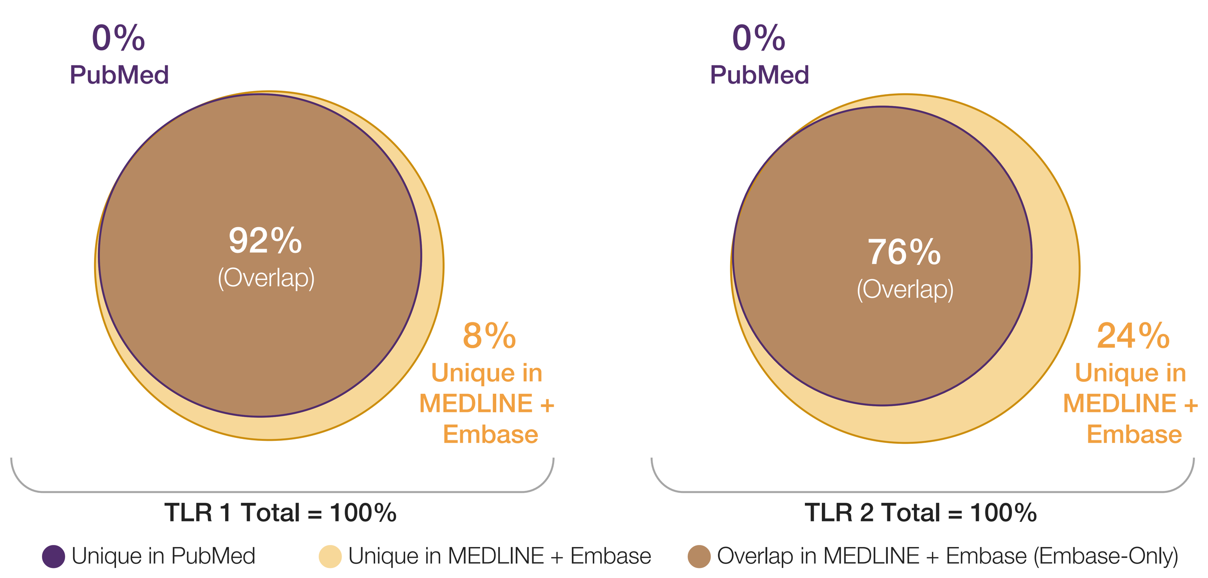
Source - von Wilamowitz-Moellendorff C et al. *Value Health*. 2025 Dec 1;28(12):S651-2.[9]
Abbreviation: TLR = targeted literature review

Results (continued)

Overlap between Database Approaches

- The distribution of studies identified through PubMed, Embase, and their overlap for both TLRs is presented in **Figure 3**.
- Across both TLRs, studies not captured by PubMed-only searches were identified exclusively through Embase. This pattern is consistent with known differences in database coverage and indexing practices, including broader journal inclusion and conference abstract coverage within Embase.
- The magnitude of these differences was more pronounced in TLR 2, suggesting that database-specific indexing plays a greater role when less standardized or pragmatic search filters are applied.

Figure 3. % of Records Overlap and Unique



Abbreviation: TLR = targeted literature review

Discussion

- In this methodological case study of two AI-assisted TLRs, PubMed-only and combined MEDLINE plus Embase searches yielded largely overlapping evidence bases. However, important differences emerged, particularly when pragmatic observational study design filters were applied. Under these conditions, the combined database approach consistently identified additional studies not captured by PubMed alone, indicating that search strategy design and database coverage jointly influence evidence retrieval.
- Despite differences in record pools, the AI-assisted screening workflow demonstrated consistent performance across both approaches, with comparable recall and prioritization patterns. These findings suggest that while AI-assisted screening can improve efficiency, it does not mitigate limitations introduced at the search stage, reinforcing the importance of robust and well-validated search strategies.

Comparison with Existing Literature

Database Coverage and Overlap

- Our findings align with previous research demonstrating substantial but incomplete overlap between PubMed/MEDLINE and Embase. Embase has been shown to provide broader coverage of biomedical literature, including conference abstracts and European journals, which are not fully indexed in MEDLINE or PubMed.[3,4]
- Systematic review guidance, including Cochrane and Preferred Reporting Items for Systematic reviews and Meta-Analyses literature search extension recommendations, consistently advises searching multiple databases to maximize sensitivity and reduce retrieval bias.[5] The present study extends this evidence by demonstrating that the magnitude of missed studies is not constant, but instead varies depending on how searches are operationalized.

Impact of Search Filters

- A key contribution of this study is the observation that pragmatic observational study design filters amplified differences between database approaches. This is consistent with prior evaluations showing that methodological filters—particularly for observational studies—can vary substantially in sensitivity and specificity across databases.[6]
- Unlike randomized controlled trial filters, which are relatively well-validated, observational study filters are often less standardized and may perform inconsistently depending on indexing practices and database structure. PubMed, which relies heavily on MeSH indexing and text-word searching, may be more sensitive to suboptimal filter translation compared with Embase, where Emtree indexing can enhance retrieval of study designs.[7,8]
- These findings underscore that filter selection and validation are critical determinants of evidence capture, particularly in TLRs focused on real-world or observational evidence.

AI-assisted Screening in Evidence Synthesis

- The consistent performance of the AI-assisted screening model across the MEDLINE and Embase records sets supports prior evidence that machine learning–based prioritization can improve screening efficiency without substantially compromising recall.
- Importantly, our results demonstrate that AI performance remained stable despite differences in database source and record composition. This suggests that AI-assisted screening is robust to variation in input datasets, provided sufficient training data are available..

Strengths and Limitations

Strengths

- In this study, we investigated a parallel comparison of database strategies across two real-world TLRs, thereby improving the applicability of the findings to a real-word evidence synthesis context. The use of identical eligibility criteria and screening workflows minimizes methodological variability.
- The integration of AI-assisted screening reflects contemporary review practices and allows evaluation of its performance in a realistic setting.

Limitations

- The study is based on two TLRs within a specific clinical domain (chronic kidney disease), which may limit generalizability to other therapeutic areas.
- While efforts were made to ensure conceptual equivalence of search strategies, differences in database indexing and syntax may have influenced retrieval.
- While the application of pragmatic observational filters reflects real-world practice, it may introduce variability that is context-dependent and may not be generalizable to all filter designs.
- The combined MEDLINE and Embase approach was used as a pragmatic benchmark and not a comprehensive reference standard. Additionally, studies through supplementary methods highlight the need for complementary approaches, including grey literature searches and citation tracking, to improve retrieval completeness.

Conclusions

- Across two AI-assisted TLRs, PubMed-only and combined MEDLINE plus Embase searches yielded overlapping evidence bases; however, gaps in retrieval were larger when pragmatic observational study design filters were applied. AI-assisted screening demonstrated consistent and efficient performance across differing record pools, but the final included evidence remained dependent on database coverage and search strategy design. These findings highlight the importance of understanding database-specific coverage (where searches are run) and methodological choices such as interface and filter selection ("how searches are run") when designing AI-assisted TLRs.

References

1. Ge L, et al. *Syst Rev*. 2024;13(1):269.
2. Sanghera R, et al. *J Am Med Inform Assoc*. 2025;32(5):893-904.
3. von Wilamowitz-Moellendorff C, et al. *Value Health*. 2025;28(12)(1):S651-S652.
4. Cichewicz A, Schaible K. Use of Artificial Intelligence for Rapid Epidemiology Reviews with Pooled Prevalence and Incidence Estimates. presented at: ISPOR—The Professional Society for Health Economics and Outcomes Research; May 13-16, 2025; Montreal, Quebec, Canada.
5. Bramer WM, et al. *J Med Libr Assoc*. 2018;106(4):531-541.
6. Lam MT, McDiarmid M. *J Med Libr Assoc*. 2016;104(4):284-289.
7. Rethlefsen ML, et al. *Syst Rev*. 2021;10(1):39.
8. Li L, et al. *Cochrane Database Syst Rev*. 2019;3(3):MR000041.
9. Higgins J, et al. Cochrane Handbook for Systematic Reviews of Interventions: Section 4.4.7 Search filters. Cochrane 2024. Accessed April 24, 2026. https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-04#section-4-4-7
10. King's College London. Searching for Systematic Reviews & Evidence Synthesis: Using Filters. Updated April 15, 2026. Accessed April 24, 2026. https://libguides.kcl.ac.uk/systematicreview/SRFilters

Disclosures

Funding provided by Thermo Fisher Scientific. All authors are employees of PPD™ Evidera™ Health Economics & Market Access, Thermo Fisher Scientific.

Acknowledgments

The authors would like to acknowledge Kawthar Nakayima and Michael Grossi of Thermo Fisher Scientific for graphic design and editorial support.

Presented at ISPOR—The Professional Society for Health Economics and Outcomes Research | 17–20 May 2026 | Philadelphia, PA, USA