# Title Page

## Title:

Assessing the Feasibility of Automated Single Database Searches (PubMed, Embase, OpenAlex, and Semantic Scholar) for Living Guideline Maintenance and Evidence Surveillance

## First author: Darren Rajit1

## Other authors: Lan Du2, Helena Teede1,3, Joanne Enticott1,3\*

## Affiliations:

## 1Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing, and Health Sciences, Monash University, Clayton, Victoria, Australia

## 2Monash Faculty of Information Technology (Machine Learning Group)

## 3Monash Partners Academic Health Sciences Centre, Clayton, Victoria, Australia

## *\*Senior and corresponding author: Joanne Enticott, E: joanne.enticott@monash.edu, Add: Level 1, 43-51 Kanooka Grove Clayton VIC 3168*

## 

## Keywords: Evidence synthesis, learning health systems, evidence retrieval, living evidence, living guidelines

## Target Journal:

* Lancet EClinicalMedicine (IF: 9.6, $5,250, Discounts can be requested if accepted)
* Research Synthesis Methods (IF: 5.0, Fee waiver available via Wiley Agreement)
* Journal of Clinical Epidemiology (IF: 7.3, Open Access Fee: $4030)

# Introduction

##### Link Evidence Based Guideline Development to LHS, and Highlight importance of Living Evidence

* Evidence Based Guidelines (EBGs) are crucial in pragmatising both Evidence Based Medicine and Person-Centred Medicine within Learning Health Systems (LHS) by ensuring stakeholder priorities are aligned with downstream evidence synthesis.
* This results in recommendations that are more likely to be of practical relevance to both frontline healthcare practitioners and consumers
* Thus, the transition of EBGs towards a living format, where the available research evidence is continuously synthesised as it is produced; is paramount to ensuring that recommendations and guidance remain relevant to frontline care.
* This is further crucial in realising high functioning and transformative learning health systems that are able to efficiently leverage the research evidence as it is produced.

##### Sustainability of Living EBG is An Issue – Automation is route for Addressing this

* However, the production of EBGs is highly time-consuming and expensive. Transition towards a living format utilising current (mostly manual) approaches compounds this, and ongoing maintenance could be potentially cost-prohibitive.
* As such, both automation and artificial intelligence enabled methods have a clear role to play towards enabling the goal of living, continually relevant guidelines.
* Thus, the first step in transitioning EBGs towards a living format firstly requires a method to continuously monitor and retrieve the current state of the literature, prior to downstream screening, data extraction and meta-analysis.

##### Highlight rise of next generation databases that allow for programmatic access

* The past 5 years have seen the rise of bibliographic databases that allow for programmatic (and by extension, automated) access to research evidence through application programming interfaces (API).
* However, limited work has been done in understanding the extent of coverage of such databases, and whether focusing search or evidence retrieval efforts on a single bibliographic database is sufficient for the purposes of continually searching the literature, and triggering updates when relevant. Indeed the Australian Living Stroke Guidelines recently transitioned to such an approach, only searching PubMed for updates after uncovering that 93% of articles could be retrieved from PubMed alone.
* As such, we build on our prior work developing the 2023 Edition of the International PCOS Guidelines by beginning the transition of the guidelines to a living format. The first step being understanding the extent of the coverage of this new generation of databases, and whether a single database search would be sufficient for evidence surveillance purposes, thus allowing savings in terms of effort.
* Here the assumption is that favourable coverage of articles included in the current edition of the guidelines would predict favourable coverage of potentially relevant articles in the future.

##### Context of International PCOS Guidelines as Dataset

* The International PCOS Guidelines involved the production of 54 evidence reviews, across 5 themes or Guideline Development Groups (GDGs), resulting in a 6000-page technical report which is the main source of data for this study. The guideline was NHMRC approved, conducted in collaboration across 40 societies in 6 continents, adapted by NICE UK due to its use of rigorous methods, and was the first guideline to include a research integrity check as part of guideline production
* As a result, the International PCOS Guidelines is a particularly robust gold standard for the development of datasets geared towards the development and evaluation of evidence-based guideline development automation techniques.
* Of these 54 evidence reviews, 50% (27/54) of these evidence reviews were systematic review updates, 37% (20/54) were new systematic reviews, and a further 11% (6/54) were narrative reviews. Here we tested coverage of articles that were included in both the systematic review updates and new systematic reviews that were carried out in the guidelines.
* 4 databases were tested. Firstly, PubMed and Embase, comprising the 2 most common databases that were searched. Secondly, OpenAlex and Semantic Scholar, which represent the new generation of field agnostic, bibliographic databases that i) aggregate multiple sources ii) and allow for programmatic access to the research corpora via publicly available APIs.
* All databases were also chosen on the basis of API availability that would enable automated and programmatic access to the research evidence.

# Aims

* To evaluate the coverage and feasibility of using only PubMed, Embase, OpenAlex or Semantic Scholar in programmatically retrieving articles that were included in the 2023 Edition of the International PCOS Guidelines.

# Methods

Code and data required to replicate the retrieval process and subsequent analyses is available as a public GitHub repository: <https://github.com/darrenkjr/International_PCOS_EBG_Surveillance_ResearchLetter>

1. Articles that were included after full text screening for each evidence review conducted as part of the 2023 International PCOS Guideline, were extracted from the technical report of the guidelines.
2. The Digital Object Identifier (DOI) of each included article was then retrieved for each article. Where DOIs were not retreivable, PubMed IDs were retrieved instead. If this failed, then the title of the article was extracted.
3. DOIs and PubMed IDs were then used to programmatically retrieve article metadata from the PubMed, OpenAlex, and Semantic Scholar APIs respectively to confirm coverage.
4. As the Embase API required an additional subscription for access, searching within Embase was instead conducted via the Ovid interface.
5. A second round, title only search was then conducted for all databases in instances where retrieval by DOI or PMID failed, or if a DOI or PMID was not retrievable for a particular article.
6. Article metadata was then consolidated for each database and the extent of coverage for each database was then evaluated.
7. The extent of overlap across each database was also evaluated
8. An exploratory sensitivity analysis of articles that were not retrievable for each database was also conducted via extracting the risk of bias assessment that was conducted for each non-retrievable article as part of the 2023 International PCOS Guideline. This is then presented as a proportion by database.

# Results

A total of 1249 articles were included in the international PCOS guidelines across all systematic reviews and systematic review updates. Table 1 below depicts the extent of coverage of databases that were tested. OpenAlex had the best coverage at 98.6%, followed by Semantic Scholar (98.2%), Embase (96.4%), and lastly, PubMed at 92.0%.

Table 1: Percentage coverage of articles that were included (n=1249) as part of the 2023 International PCOS Guidelines that were retrievable for each tested Database.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | PubMed (%) | Embase (%) | OpenAlex (%) | Semantic Scholar (%) |
| Overall | 92.0 | 96.4 | 98.6 | 98.2 |
| GDG 1 | 89.8 | 97.0 | 98.2 | 96.1 |
| GDG 2 | 91.4 | 94.7 | 99.2 | 99.2 |
| GDG 3 | 93.5 | 92.7 | 97.6 | 97.6 |
| GDG 4 | 93.0 | 97.3 | 98.8 | 99.5 |
| GDG 5 | 93.9 | 98.6 | 99.3 | 98.6 |

Figure 1 below present the overall assessed risk of bias for articles that were not retrieved from each tested database. These assessments were based on the original assessments that were conducted as part of the quality assessment stage of the production of the International PCOS Guidelines 2023. OpenAlex demonstrated the highest proportion of high risk of bias articles amongst the articles that were not retrieved (47.1%, n=8), followed closely by moderate-risk articles (52.9%, n=9). Semantic Scholar showed a similar pattern, with 31.8% (n=7) high-risk and 68.2% (n=15) moderate-risk articles that were not retrieved.

Notably, both PubMed and Embase had articles that were unretrieved that were originally assessed as low risk of bias. With 13.0% (n=13) of unretrieved articles being low risk for PubMed, and 17.8% (n=8) of unretrieved articles being low risk for Embase. These low risk articles may have had a greater contribution in downstream meta-analyses that were conducted during guideline production, indicating that missing these articles may have an impact in downstream recommendations.

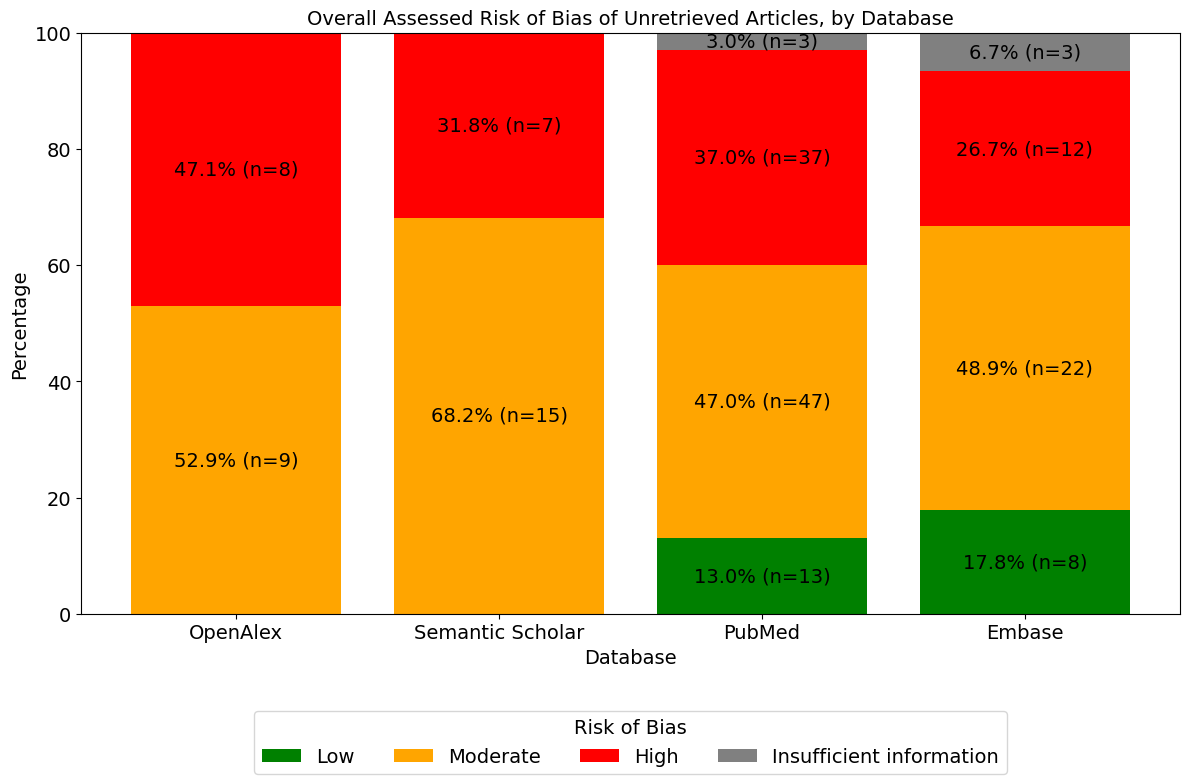


Figure 1: Overall Assessed Risk of Bias of Articles that were not retrieved, for each tested Database

The TopSet diagram in Figure 2 below illustrates the extent of overlap in successful article retrievals across all 4 tested databases. Notably, there was a high degree of overlap between all 4 databases, with 89.7% (n=1118) of all articles being retrievable from all 4 databases. The second largest overlap group was between Embase, Semantic Scholar and OpenAlex, with an overlap of 60 articles (4.8%) of all retrieved articles. Notably, there were 2 instances where articles were uniquely retrieved by a singular database. Embase, with 2 unique articles, and likewise with OpenAlex, also with 2 unique articles. Nonetheless, OpenAlex is represented in all coverage intersections, indicating that the database had the widest coverage, as also depicted in Table 1.

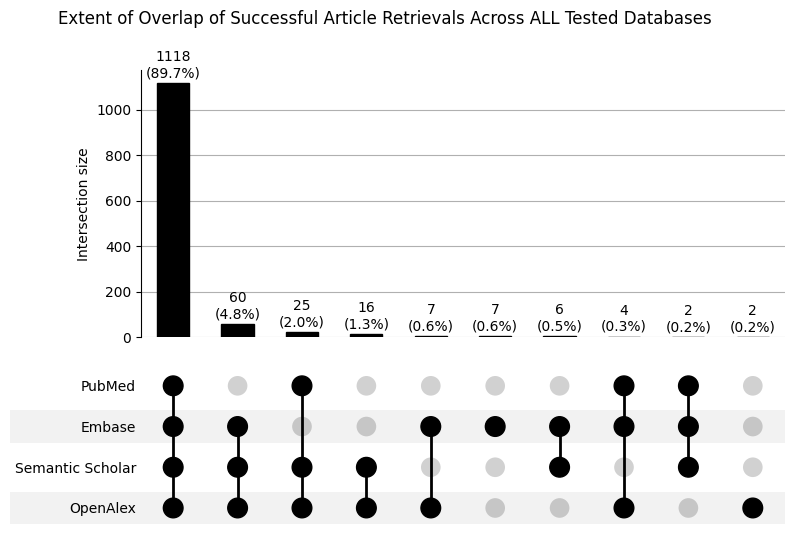


Figure 2: Extent of Coverage Overlap between all tested Databases

# Discussion

* Summary of results. Note that Embase API was planned to be used, however an institutional subscription was required, and due to time constraints, evaluation proceeded without waiting for API access to be granted. As such, Embase was searched via Ovid Medline instead for completion. Nonetheless, the other 3 options have publicly available APIs that are accessible free of charge for research / academic purposes. Thus, these are recommended for surveillance purposes.
* Placeholder - get co-authors first reaction

# Limitations

# Future work

* Reverse engineer a high sensitivity search strategy for best performing database. However, we expect that such approaches would result in low specificity (or precision). Thus such search strategies should be integrated with supplementary approaches to improve efficiency by constraining the articles required to be screened by a human. We leave this for planned future work. (DOI Protocol)

# Conclusion

* Placeholder