Can Generative Artificial Intelligence (GenAI) be Leveraged to Automate Scoping Searches of Systematic Literature Reviews (SLRs)?

MSR53

IQVIA

Mavrigiannaki A M,[1] Oikonomidi T,[1] Kaur N[2]

1. IQVIA Athens, Greece
2. IQVIA Bengaluru, India

Objectives

SLRs are foundational to evidence-based healthcare, yet their planning and execution demand substantial resources and could benefit from partial automation. A recent review found that most automation efforts have concentrated on the screening stage, followed by data extraction and risk of bias assessment [1]. These advancements are largely driven by progress in genAI [2].

An essential yet often underexplored component of SLR planning is scoping, which precedes protocol development and aims to estimate the volume of search results and the effort required to complete the SLR [3]. Scoping searches are typically crafted by subject matter experts (SMEs), and research into how AI can support this stage remains limited. This study investigates the capabilities of Microsoft (MS) Copilot® in assisting human SMEs with the development of SLR scoping searches.

Methods

This cross-sectional study compared the search strategies originally developed by SMEs for three economic SLRs with those that were developed by MS Copilot®. The SLRs focused in three different disease areas (thalassemia, muscle-invasive bladder cancer and Sjögren's syndrome).

First, a prompt was devised by SLR SMEs. The prompt was tested and revised with one SLR. The final prompt is presented in Figure 1. The prompt asked MS Copilot® (Web version) to develop a search strategy for Embase® via OVID SP®, using three inputs: the SLR research question, the SLR PICOS criteria and a test set of studies the search should retrieve.

Once MS Copilot® produced the search strategy, SMEs revised it for critical errors (e.g., incorrect syntax that would prevent the search from running on OVID SP®). Critical errors were recorded and analysed. The search was then conducted, and the results were exported on MS Excel®. For each search strategy, SMEs evaluated 1) output volume (i.e., the total number of results with the MS Copilot® search versus the original SME search), and 2) sensitivity (i.e., the proportion of included articles in the original SLRs that was retrieved by the MS Copilot® search). Tests were conducted on 19-20 June 2025. The original SLR searches were repeated on 19-20 June to ensure the number of search results represented the same time period.

Figure 1. Prompt used to generate scoping searches in the study; terms in bold brackets should be updated to match the SLR protocol

Initialisation

Acting as an information retrieval specialist, develop a search strategy for a systematic literature review.

Database

Adapt your search strategy for **[Embase] using [OVID]** syntax. The output must be a structured search using Boolean logic, free text words (including synonyms) and **[Emtree]** terms. Please prioritise **[sensitivity]** to ensure all relevant studies are identified.

Protocol inputs

The review aims to answer the following question**: [insert research question from SLR protocol]**

The PICOS of the review are:

- Patients**: [insert from SLR protocol]**
- Intervention/comparison: **[insert from SLR protocol]**
- Outcome: **[insert from SLR protocol]**
- Study design: **[insert from SLR protocol]**
- Limits: **[English]** language, publication date: **[insert limit]** for full text journal articles and **[insert limit]** for conference abstracts; geography: **[insert restrictions]**

Test set

Ensure your search strategy retrieves the studies in the following test set: **[List at least two complete references that the search should retrieve]**

Results

In all three test cases, the search strategy developed by MS Copilot® required SME corrections in order to run on OVID SP® and provide usable and valid output.

A qualitative analysis, presented in Figure 2, showed that errors ranged from minor (e.g., use of unsupported characters) to use of incorrect search logic (i.e., blocks of keywords were incorrectly connected with the Boolean term 'and' instead of 'or', thus changing the approach to the research question).

In two of three SLRs, MS Copilot® searches overestimated the number of hits (by 388-614 hits, i.e., 34%-45% more than the original SLR search) (Figure 3). In the third case, the Copilot® search underestimated the number of hits by 244 (13%).

In terms of sensitivity, results varied across cases; MS Copilot® search retrieved 90% and 91% of the reports included in the original SLRs for 2 of 3 test cases (Table 1). In the third case, the Copilot® retrieved only 23% of the reports included in the original SLR for Sjogren's Syndrome.

Table 1. Sensitivity of MS Copilot® search (proportion of SLR includes retrieved by the search)

Disease	Inclusion rate
Thalassemia	90%
Bladder cancer	91%
Sjogren's syndrome	23%

Figure 3. Number of results (titles/abstracts) retrieved by SME and MS Copilot® searches for the same SLR

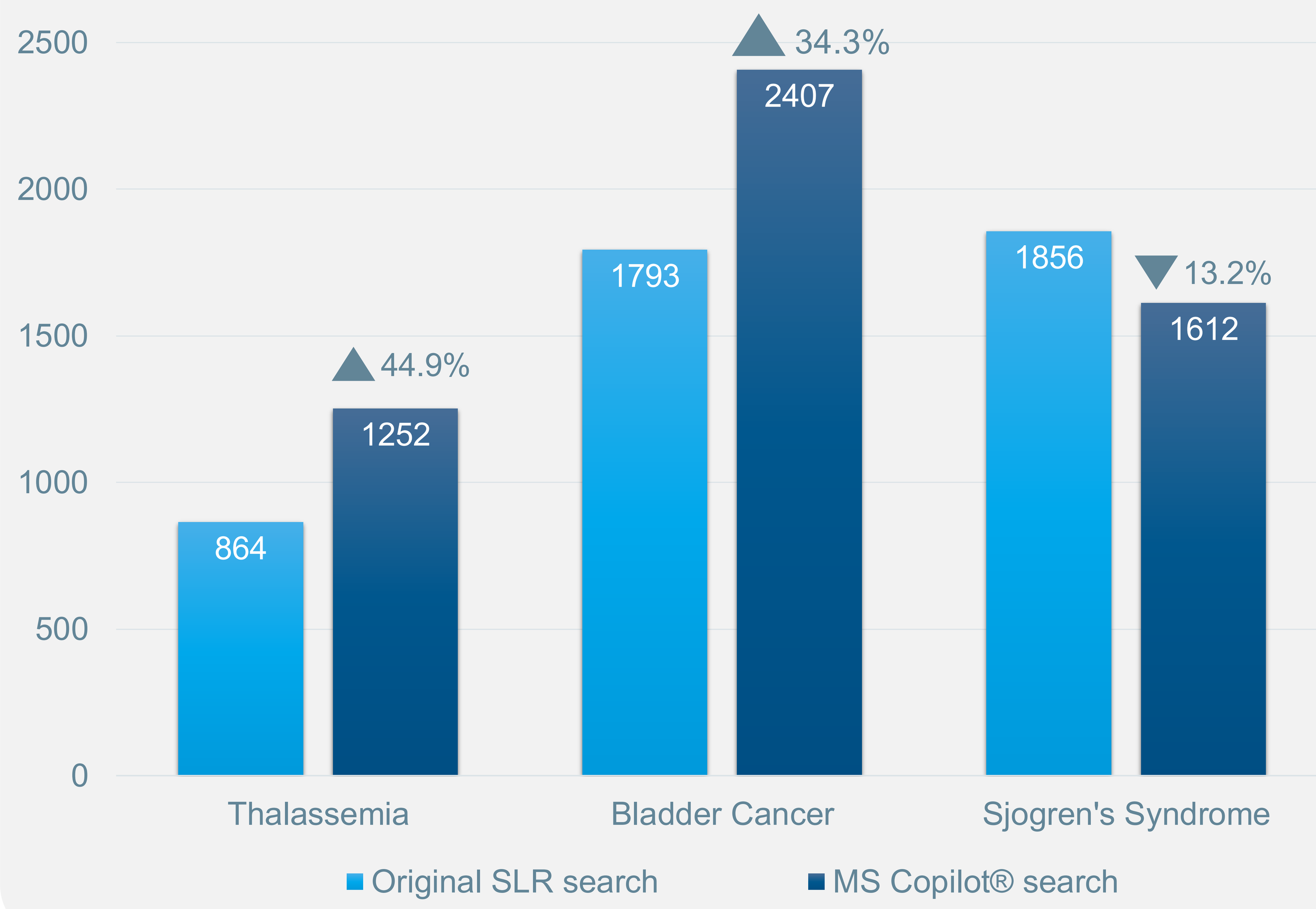
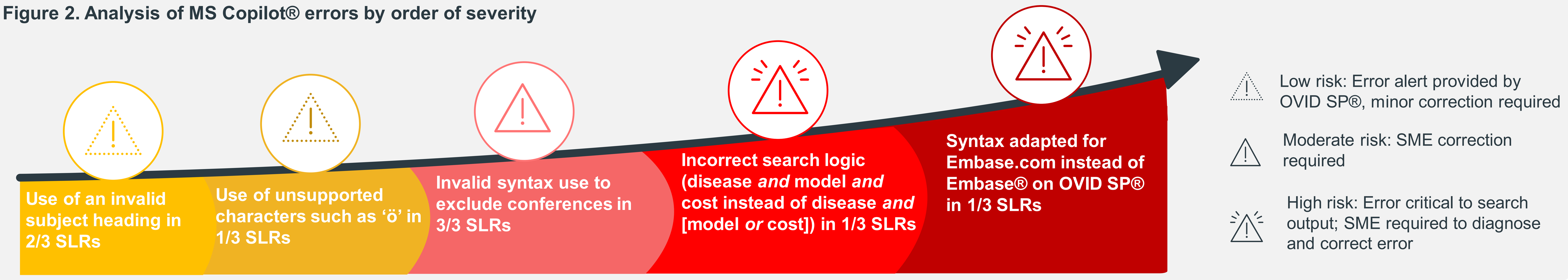


Figure 2. Analysis of MS Copilot® errors by order of severity



Conclusions

Across the three test cases, MS Copilot® performed poorly at estimating the number of final search results; search results were overestimated by 34%-45% or underestimated by 13%. The scoping searches performed variably at identifying relevant studies, with the proportion of final includes retrieved from the search ranging from 23% to 91%. As all SLRs evaluated focused on economic evidence, the observed variability in performance may be attributed to the disease area, or to randomness inherent to web-accessible versions of LLMs, including MS Copilot®. MS Copilot® could be valuable in assisting reviewers with limited knowledge in developing search strategies, particularly for targeted literature reviews, but revisions by SME remain indispensable.

References: 1.Ofori-Boateng, R., Aceves-Martins, M., Wiratunga, N. et al. Towards the automation of systematic reviews using natural language processing, machine learning, and deep learning: a comprehensive review. Artif Intell Rev 57, 200 (2024). 2.de la Torre-López, J., Ramírez, A. & Romero, J.R. Artificial intelligence to automate the systematic review of scientific literature. Computing 105, 2171–2194 (2023). 3. Peters, M.D.J., Marnie, C., Colquhoun, H. et al. Scoping reviews: reinforcing and advancing the methodology and application. Syst Rev 10, 263 (2021)

International Society for Pharmacoeconomics and Outcomes Research Europe 2025, 9-12 November, Glasgow, UK.