Closer, but Not There Yet: Evaluating Large Language Models (LLMs) for Scoping Searches in Oncology Systematic Literature Reviews (SLRs)

MSR300

Oikonomidi T,[1] Gerna D,[1] Asteriadis A,[1] Punekar Y[2]
1. IQVIA Athens, Greece
2. IQVIA London, United Kingdom

Objectives

Scoping literature searches are critical for SLR planning, because they are used to estimate the volume of published evidence and associated resource requirements.

Scoping searches developed by Microsoft (MS) Copilot® previously required substantial subject matter expert (SME) corrections for major technical errors [1].

This case study investigated the performance of newer LLMs and skill-based agentic approaches, by comparing searches generated by MS Copilot® and Claude Cowork Fable 5.1® with SME searches for three previously conducted SLRs.

Methods

Three SLRs were evaluated: one clinical SLR in acute myeloid leukemia (AML), one economic and one clinical SLR in bladder cancer. All searches were developed for Embase® via OVID SP®.

First, two AI-based approaches were used to generate searches:

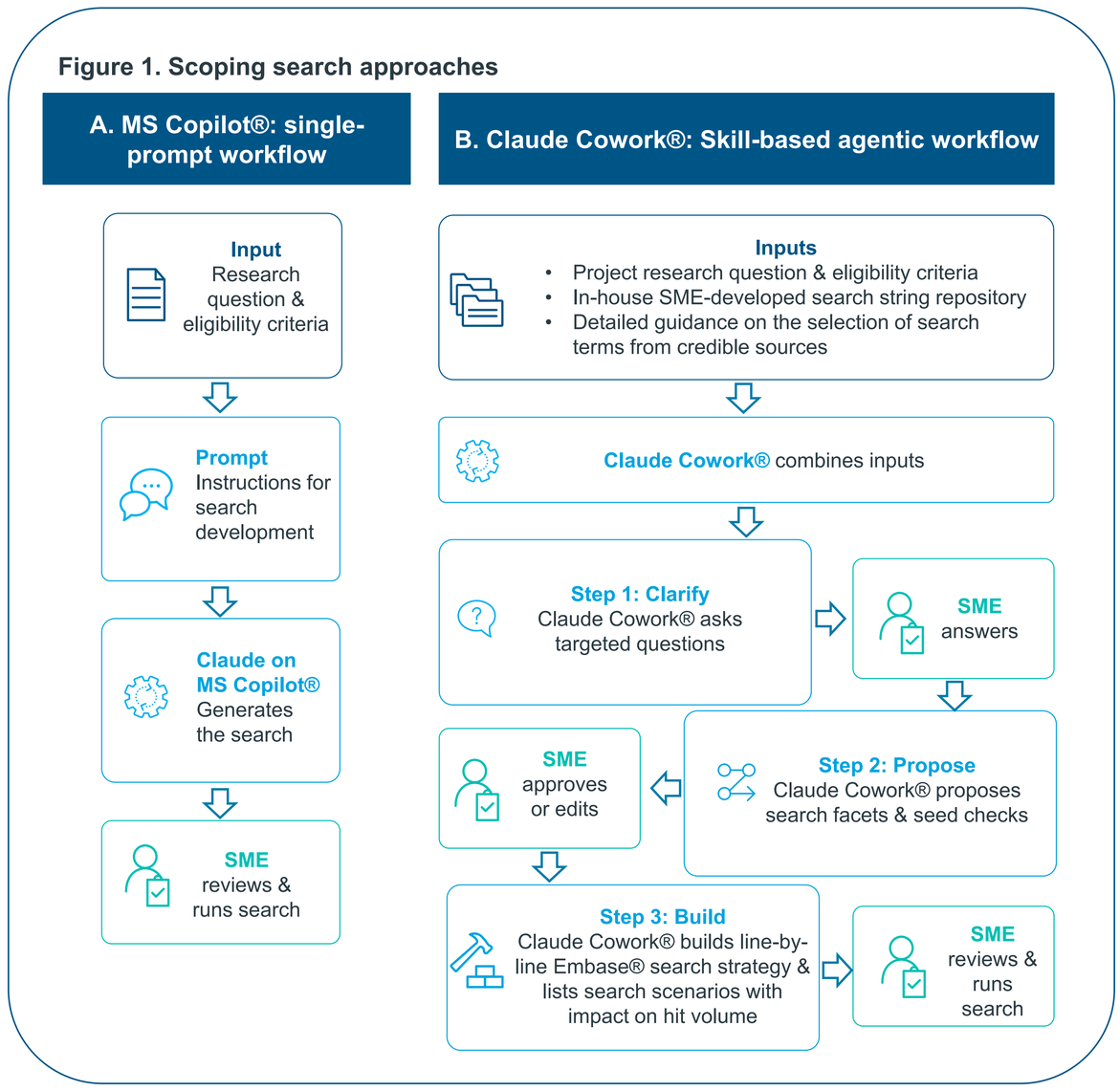
- A single prompt was used with MS Copilot® (Claude Opus 4.7, Figure 1A).
- A Skill was developed with Claude Cowork® (Fable 5.1, Figure 1B).

Second, two tests were conducted:

- Test 1 (25 June 2026) compared MS Copilot® to SME searches.
- Test 2 (13 September 2026) compared Claude Cowork® Fable 5.1 to SME searches.

For each test, LLM and SME searches were run on the same date. For each LLM search, the SME assessed: (1) whether the search could run in OVID SP® and (2) the number of records retrieved compared with the corresponding SME search.

Figure 1. Scoping search approaches



Results

Search performance: In all three cases, Claude Cowork® Fable 5.1 estimated the number of hits more accurately than MS Copilot® Claude Opus 4.7.

- MS Copilot® Claude Opus 4.7 searches overestimated in all three test cases the number of hits compared with the original SME searches (Figure 3, +370 to +24,413 hit discrepancy). Overestimation was explained primarily by use of non-indication-specific terms (e.g. “UC” for urothelial carcinoma capturing ulcerative colitis records). Of note, rerunning the same prompt with the same input produced a different search strategy, with different limits, indicating low reliability.
- Claude Cowork® Fable 5.1 retrieved similar numbers of hits to the SME searches in two of three test cases (Figure 4, +30 to +518 hit discrepancy). The exception was the economic SLR case, where it overestimated the search hits by +11,683. This was attributed to use of broad economics keywords that attracted substantial noise.

Error risk: A qualitative analysis of LLM errors (Figure 2) suggests that Claude Cowork® Fable 5.1 made primarily low/moderate risk errors.

Cost: Claude Cowork® cost approximately 10 USD per scoping search, while MS Copilot® had no additional per-search cost beyond the existing Microsoft 365 Copilot license.

Figure 2 Errors requiring SME correction by severity: Claude Cowork® Fable 5.1

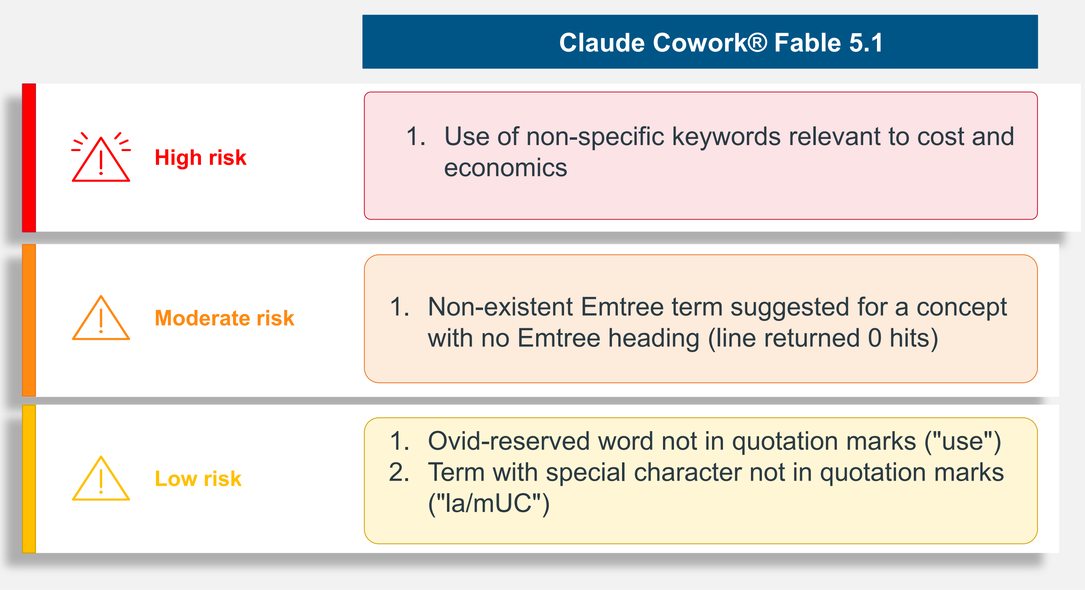


Figure 3. Number of results (titles/abstracts) retrieved by SME and MS Copilot® searches for the same SLR

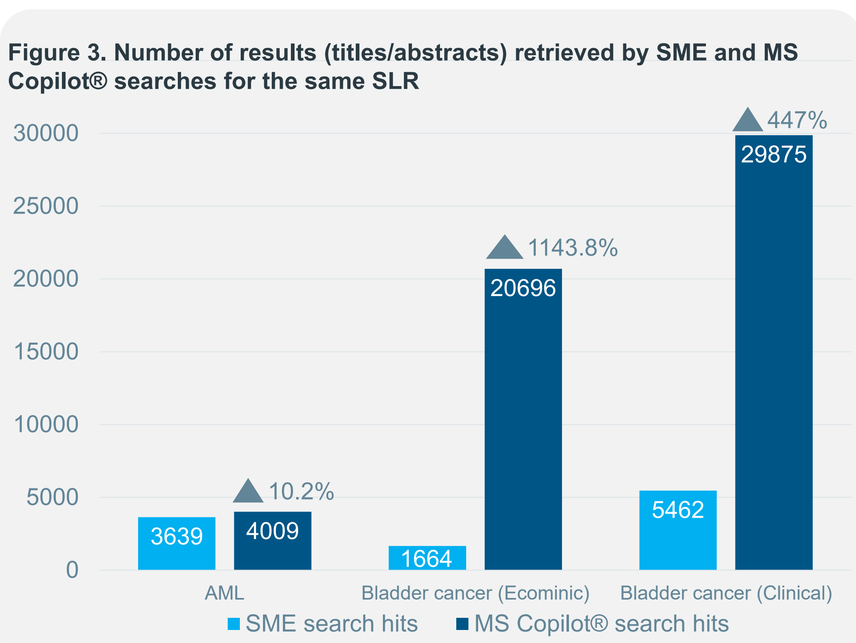
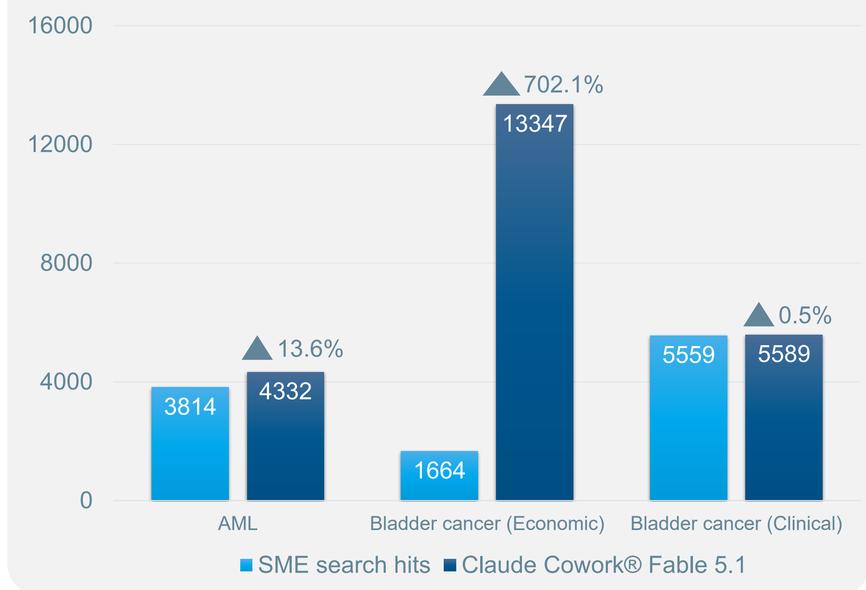


Figure 4. Number of results (titles/abstracts) retrieved by SME and MS Fable 5.1 Claude cowork® searches for the same SLR



Conclusions

- In two of three tests, searches developed with a bespoke Skill on Claude Cowork® Fable 5.1 retrieved results close to the SME searches
- In contrast, simple prompts on MS Copilot® led to significant overestimation of the search results volume
- These findings suggest agentic AI, developed based on bespoke inputs by SMEs, can be used as a starting point for search strategy development with SME QC, to reduce the effort required to scope and plan SLRs

References: 1. Mavrigiannaki, A.M., Oikonomidi, T., Kaur, N. Can Generative Artificial Intelligence (GenAI) Be Leveraged to Automate Scoping Searches of Systematic Literature Reviews (SLRs)? *Value Health.* 28(12 Suppl 1), S503 (2025)

International Society for Pharmacoeconomics and Outcomes Research Europe 2026, 8-11 November, Vienna, Austria.

© 2026. All rights reserved. IQVIA® is a registered trademark of IQVIA Inc. in the United States, the European Union, and various other countries.