

THE ECOSYSTEM OF EVIDENCE

Navigating the future of the ecosystem of evidence

Artificial intelligence and digital transformation in research and healthcare

11th EBHC International Conference
Taormina, 21st – 24th October 2026

#EBHC2026

GIMBE EVIDENCE FOR HEALTH

International Society for Evidence-Based Health Care

McMaster University | DEPARTMENT OF HEI Health Research Methodology

UNIVERSIDAD DE LA FRONTERA

EXPLORING THE USE OF AN INTEGRATED END-TO-END PLATFORM, THE SK PLATFORM, IN EVIDENCE SYNTHESIS: A Case Study from one recommendation question informing the ASH AML Guideline

María José Oliveros, PhD, MSc, PT
Health Research Methodology, McMaster University
Universidad de La Frontera, Departamento de Ciencias de la Rehabilitación, Chile

Romina Brignardello-Petersen, Yetiani Roldan
Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University

SYSTEMATIC REVIEWS (SRs) ARE ESSENTIAL, BUT RESOURCE-INTENSIVE

>12

MONTHS

of manual, multi-stage effort for conducting a typical SR



Artificial intelligence (AI) has shown promise in supporting SR tasks, with encouraging results across different stages of evidence synthesis.

58%

Records auto-excluded (Cochrane classifier)

100% / 90%

Study-design accuracy (Elicit / ChatGPT)

97%

Population accuracy (Elicit)

However, most tools address individual tasks, creating interoperability gaps between platforms and highlighting the need for integrated, end-to-end solutions.

Thomas J, et al. J Clin Epidemiol. 2021;133:140–151; Helms Andersen T, et al. Cochrane Evid Synth Methods. 2025;3(4).

STUDY OBJECTIVES

To **compare** the **performance** and **resource utilization** of an AI tool against traditional SR methods used in a completed SR that informed the 2025 ASH guidelines for older adults with AML.

1 Performance

Agreement and completeness (consistency) of outputs across SR stages

2 Resource utilization

Time spent, reviewer workload, and task distribution at each stage.

Process insights

Strengths, limitations, and opportunities identified through reviewer debriefs.

PLATFORM AND CASE SELECTION RATIONALE

WHY THE SK PLATFORM

1 **Existing tools**
Cochrane classifier, Laser-AI, Elicit, ChatGPT — each covers one disconnected stage.

2 **Key limitations**
Fragmented workflows, Data rarely transfers cleanly; teams must learn multiple systems.

3 **An end-to-end alternative**
The SK Platform connects search through synthesis in one system.

SK PLATFORM BY EPISTEMONIKOS

BUILT ON EPISTEMONIKOS

A nonprofit medical knowledge organization maintaining the world's largest database of health-related systematic evidence reviews.

Epistemonikos. The Sustainable Knowledge (SK) Platform.

PLATFORM AND CASE SELECTION RATIONALE

WHY THIS GUIDELINE QUESTION

ASH AML guideline

Older adults with newly diagnosed AML considered candidates for antileukemic therapy

Conventional induction + post-remission therapy
vs HMA/LDAC ± venetoclax

→ **2 direct comparisons / 2 recommendations**

We selected a challenging, real-world guideline question

- **RCTs** and **NRSs**
- Evidence from a **previous guideline + an update**
- **Groups of therapies** rather than single agents
- **Scoping required** before data extraction
- **Heterogeneous outcome-data** formats

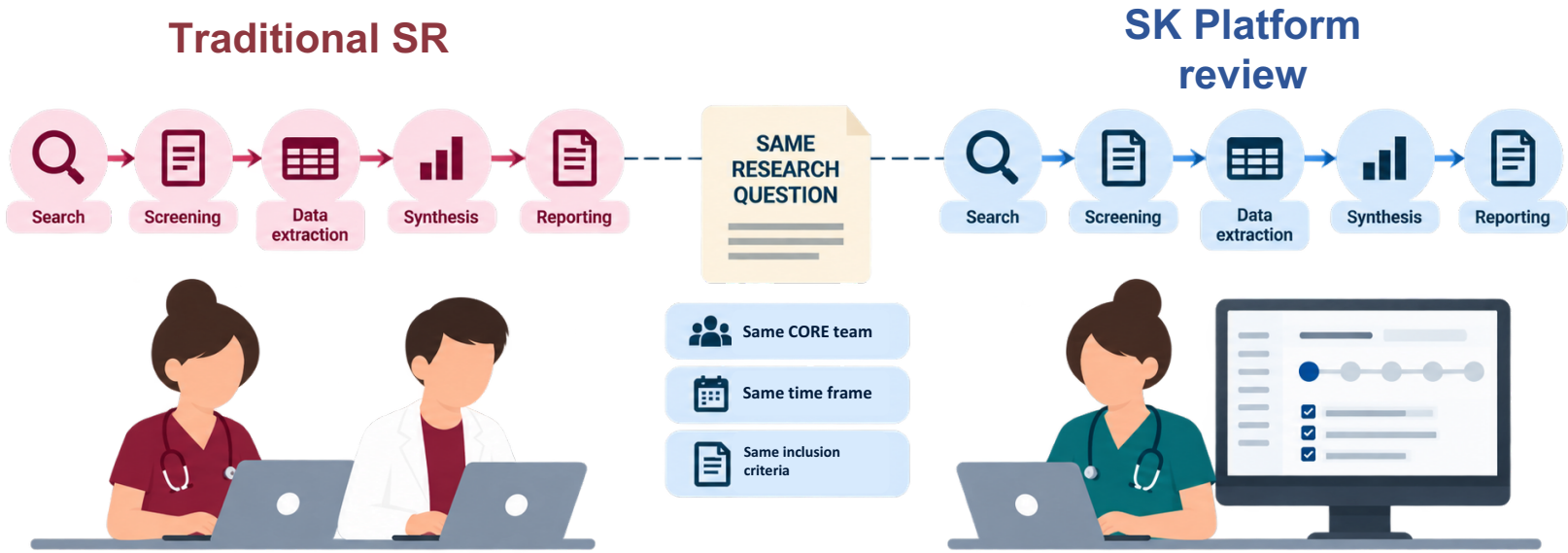
The test case reflects common methodological challenges in guideline evidence synthesis.

Sekeres MA, et al. American Society of Hematology 2025 guidelines for treating newly diagnosed acute myeloid leukemia in older adults. Blood Adv. 2026;10(6):1897–1928.

AML: acute myeloid leukemia; HMA: hypomethylating agents; LDAC: low-dose cytarabine; VEN: venetoclax, allo-HCT: allogeneic hematopoietic cell transplant; RCTs: randomized controlled trials; NRSs: non-randomized studies

HOW WE COMPARED THE TWO APPROACHES

We conducted the same SR de novo using the SK Platform, while keeping the key methodological parameters consistent.

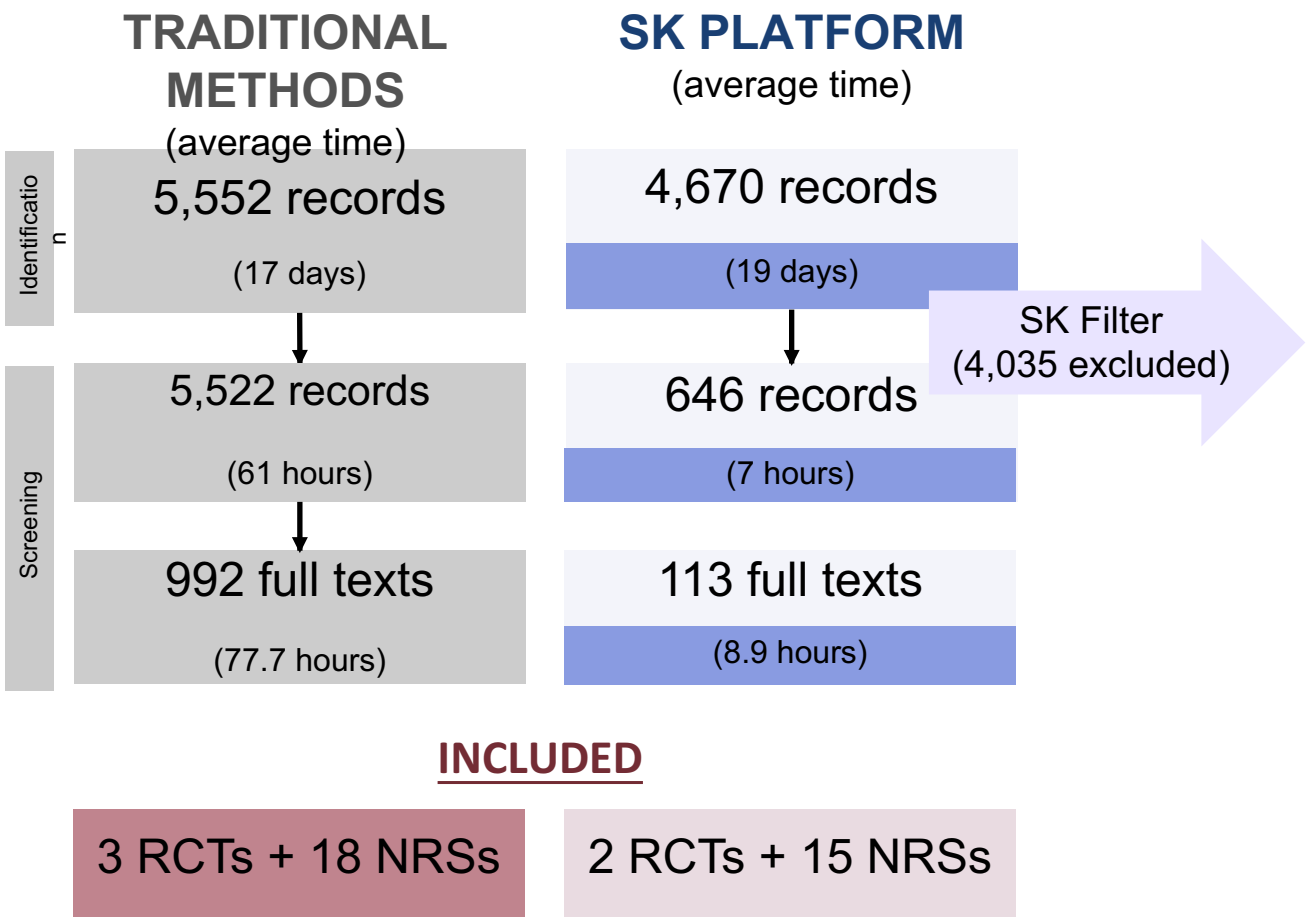


Experienced SR reviewers

Experienced SR reviewers who did not participate in the traditional SR

Consistency | Resource use

IDENTIFICATION AND SCREENING



SK PLATFORM BY EPISTEMONIKOS

CONSISTENCY

- 16% fewer records retrieved
- 66.7% agreement RCT identification
- 61% agreement NRS identification

DATA EXTRACTION AND SYNTHESIS

DATA AVAILABLE FOR ANALYSIS	RCTs 46% agreement | NRSs 50% agreement
EFFECT ESTIMATES	Consistent effect estimates for 50% of outcomes in RCTs and 0% in NRSs
DIRECTION OF EFFECT	Consistent direction of effect for 83.3% (5/6) of outcomes in RCTs and 66.7% (4/6) in NRSs.
CERTAINTY OF EVIDENCE (CoE)	Consistent CoE ratings for 66.7% (4/6) of outcomes in RCTs and 100% in NRSs.

SK PLATFORM BY EPISTEMONIKOS

CONSISTENCY

Differences in data extraction affected effect estimates, but both approaches yielded consistent directions of effect and CoE ratings for most outcomes.

RESOURCE UTILIZATION

+2 days

Search setup

One additional meeting needed with developers

-54h

T&A screening

Range 28–625h saved vs. traditional

-69h

Full-text screening

Range 63–73h saved vs. traditional

+3 months

Extraction setup

vs. 2.5 weeks to build a traditional Excel form

≈ equal

Extraction time

3.6h / 2 RCTs (SK) vs. 5.4h / 3 RCTs (traditional)

6.7%

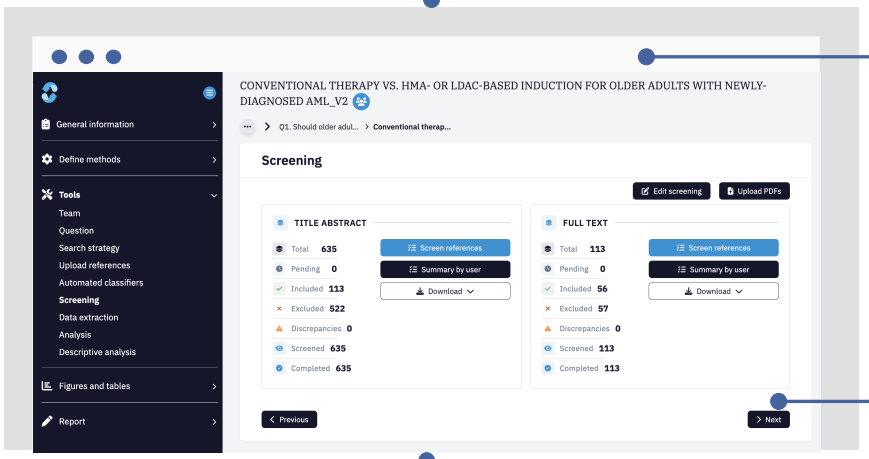
AI extraction contribution

f all required data items correctly extracted by AI without human intervention.

SK PLATFORM BY EPISTEMONIKOS

SK reduced reviewer time during screening but required substantial initial setup and did not eliminate manual tasks in later review stages.

OPPORTUNITIES FOR FURTHER PLATFORM DEVELOPMENT



Expand database integration

Improve extraction formats and supplementary-file access to accommodate different data structures and ensure access to all relevant study information.

Strengthen supervisory / quality-control functionality to allow oversight of reviewer decisions and real-time corrections.

Integrate GRADE and Summary of Findings workflows to support CoE assessments and evidence presentation within the platform.



- SK identified fewer studies and captured fewer outcomes, resulting in inconsistent effect estimates, although direction of effect and CoE were largely consistent.
- Need for substantial **human oversight** and **quality control** before using SK-generated evidence to inform ASH guideline recommendations.
- **Potential value** for scoping, rapid evidence surveillance, provisional guidance, living reviews, and detecting changes in evidence..
- Possible **next steps** include
  - evaluating newer SK platform versions,
  - comparing performance with alternative AI platforms, and
  - Assess panel trust and usability of AI-generated outputs.

We **thank** the **American Society of Hematology** for supporting this project and presentation.





THE ECOSYSTEM OF EVIDENCE

Navigating the future of the ecosystem of evidence

Artificial intelligence and digital transformation in research and healthcare

11th EBHC International Conference
Taormina, 21st – 24th October 2026

#EBHC2026

GIMBE EVIDENCE FOR HEALTH

International Society for Evidence-Based Health Care

McMaster University | DEPARTMENT OF HEI Health Research Methodology

UNIVERSIDAD DE LA FRONTERA

EXPLORING THE USE OF AN INTEGRATED END-TO-END PLATFORM, THE SK PLATFORM, IN EVIDENCE SYNTHESIS: A Case Study from one recommendation question informing the ASH AML Guideline

María José Oliveros, PhD, MSc, PT
Health Research Methodology, McMaster University
Universidad de La Frontera, Departamento de Ciencias de la Rehabilitación, Chile

Romina Brignardello-Petersen, Yetiani Roldan
Department of Health Research Methods, Evidence, and Impact (HEI), McMaster University