

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

30th GIMBE 1996-2026 EVIDENCE FOR HEALTH

Evimeter

Synthesizing, Grading and Visualizing Clinical Evidence on Devices and Diagnostics using a Structured Mixed-Method Approach

Leeza Osipenko
Consilium Scientific CEO

CONSILIUM SCIENTIFIC

Background

CONSILIUM SCIENTIFIC



Aims

CONSILIUM SCIENTIFIC

Visualize clinical evidence underlying a particular medical device or a diagnostic test in a given indication

Test the methodology for Evimeter development

Collect feedback from stakeholders on Evimeter's application/utility

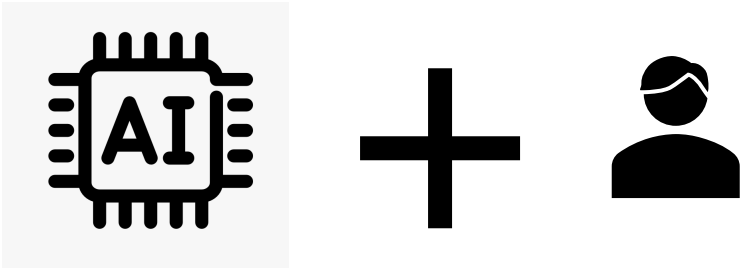
US-focused project
Exploring impact on Medicare Administrative Contractors

Employ AI tools and assess accuracy and gains in productivity

Establish use cases

Methods

CONSILIUM SCIENTIFIC



Evidence searches

↓

Data cleaning and structuring

↓

Technology-specific critical appraisal

↓

Label and supporting documents development

Results

CONSILIUM SCIENTIFIC

10 Labels

Supporting documents provide essential information for users on underlying data, methodology, detailed approach to scoring, missing data, suggested research to close evidence gaps

Stakeholder evaluation -> constructive criticism

Evimeter:

- novel
- labels are easy to understand
- supporting documents are highly informative

Results: Use of AI

- Indispensable
- Human oversight mandatory

Pro Versions

- **GPT**
- **Grok**
- **Claude**
- **SemanticScholar**
- **ConsensusAI**

- **Poor** performance in searches
- **Satisfactory** performance in evidence evaluation
- **Excellent** performance in data synthesis/presentation

Results: Use of AI

CONSILIUM SCIENTIFIC

Step	What humans did	What AI did	Our Satisfaction with AI Performance
Methodology interrogation	Developed methodology Reviewed AI outputs	Reviewed and criticised our approach	70%
Evidence searches	Developed search terms, conducted manual database searches, and checked for studies missed by AI	Searched for and screened potentially relevant clinical trials and publications	40%
Data cleaning and structuring	Reviewed manually each trial for inclusion	Search for duplicates, classification of studies, identification of suitable studies (indication/intervention)	70%
Technology-specific critical appraisal	10% checklists done manually for comparison; reviewed all AI outputs, corrected prompts, errors	Filled in check lists	70%
Label and supporting documents development	Developed labels in Tableau and Canva, developed supporting documents' structure and drafts	Produced multiple versions on labels, helped with visualization, synthesized data for supporting documents	90%

Key to Success

CONSILIUM SCIENTIFIC

High-Quality Prompting!

Are you asking the right questions?
Are you providing clear instructions?
Are you feeding top-quality examples?

Limits

CONSILIUM SCIENTIFIC

Evimeter

- Methodology requires further scrutiny, perhaps simplification
- Perception varies by audience
- Missing data

AI Use

- Closed system training of AI models for evidence assessment was out of scope – e.g. development of dedicated agent
- Human/AI process may have resulted in omissions/mistakes
- Poor AI performance in searchers (identification of studies)

Conclusions

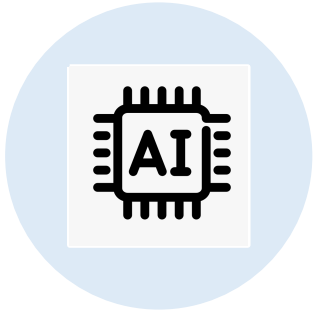
CONSILIUM SCIENTIFIC



Evimeter is independent, fully transparent and can be used by decision-makers in different domains and jurisdictions



Evimeter is a discovery tool for unveiling inconsistencies



AI agent development for evidence interrogation is feasible and necessary. It can produce high-quality reliable results for appraisal of clinical evidence



Human review remains essential at each step of the process (so far…)

Learn More and Get in Touch

CONSILIUM SCIENTIFIC

See actual Evimeter labels and reports

Visit: https://consilium-scientific.org/projects/evimeter

Contact: management@consilium-scientific.org