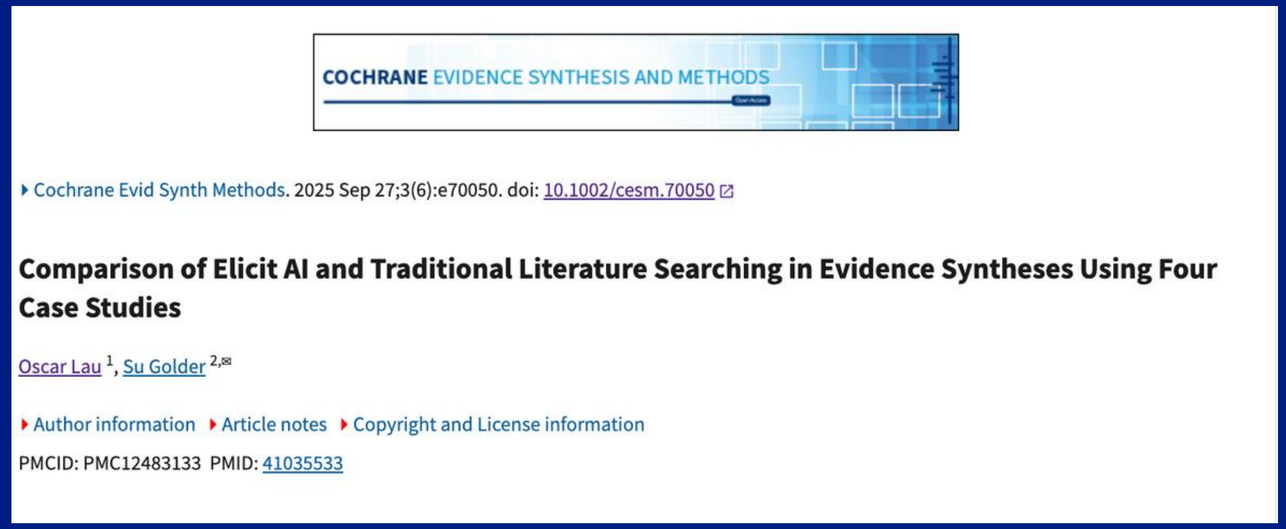
Evaluating Elicit for use in Systematic Review Searching



COCHRANE EVIDENCE SYNTHESIS AND METHODS

Cochrane Evid Synth Methods. 2025 Sep 27;3(6):e70050. doi: 10.1002/cesm.70050

Comparison of Elicit AI and Traditional Literature Searching in Evidence Syntheses Using Four Case Studies

Oscar Lau [1], Su Golder [2,✉]

Author information · Article notes · Copyright and License information

PMCID: PMC12483133 PMID: 41035533

Su Golder, Helen Fulbright and Oscar Lau

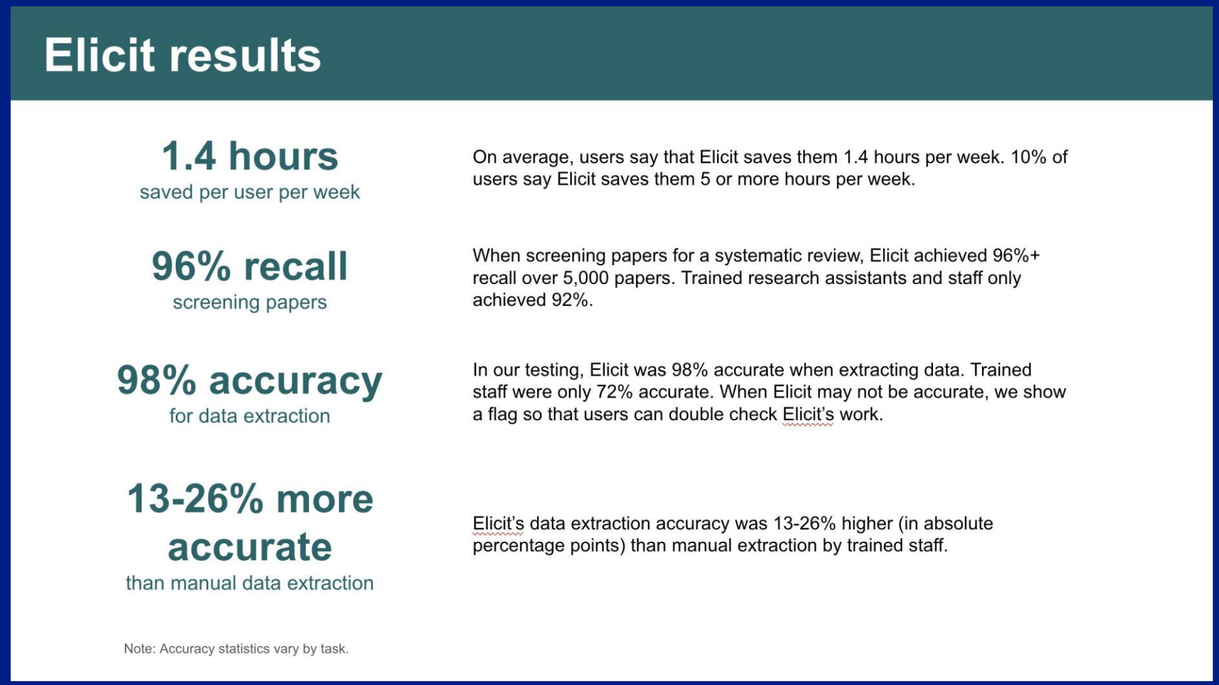
Evaluations of Tools

Most evaluations of AI tools for conducting systematic reviews are conducted by the developers or authors of the AI tools themselves

Much more akin to marketing

Frequently use simple reviews of RCTs of drug interventions

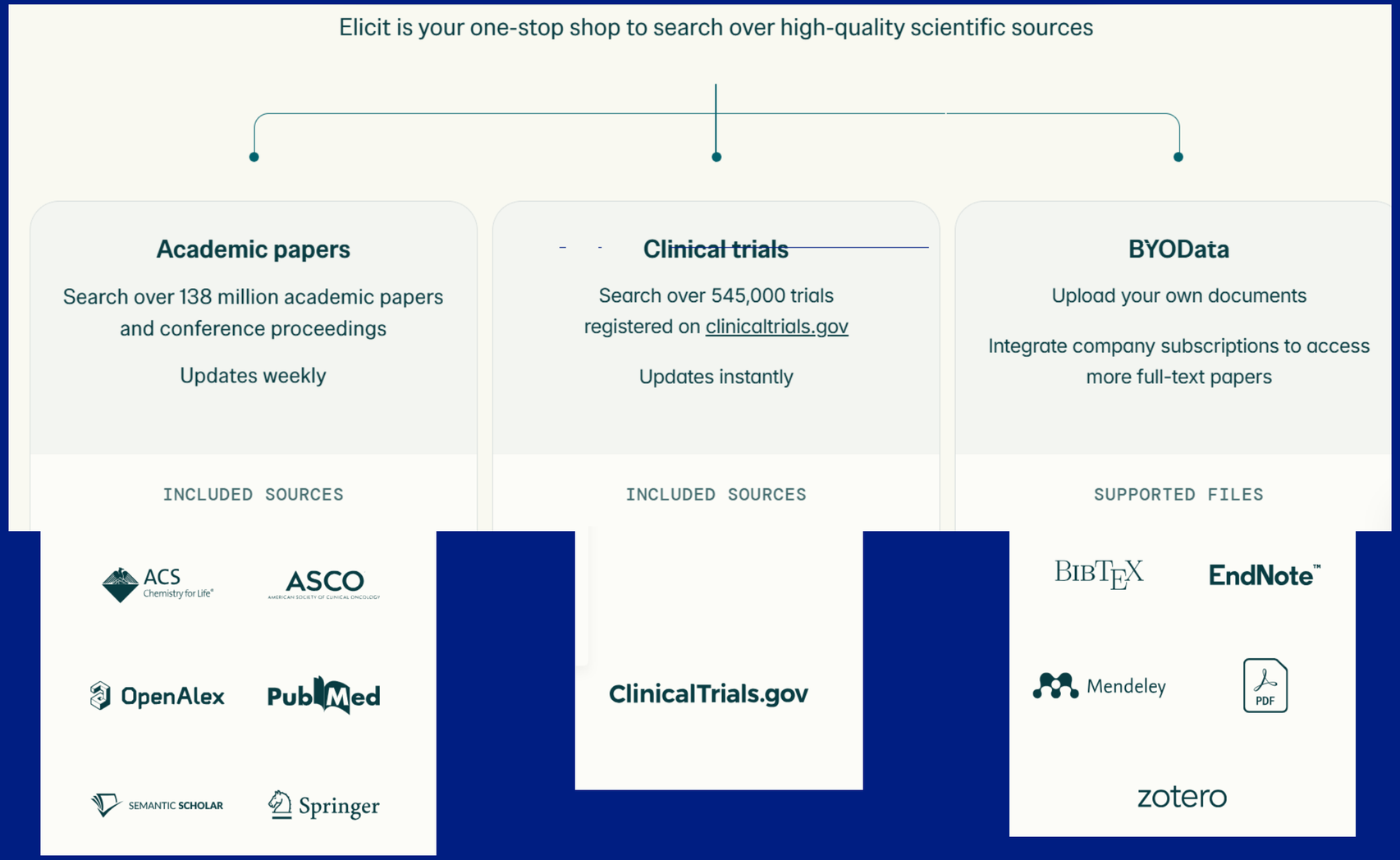
Limited published independent academic research



What is Elicit?

- Elicit is an AI research assistant
- Launched by Ought in 2021
- Elicit can automate systematic reviews to save researchers time (80% of the time required according to Elicit)
- Elicit offers 3 different workflows: Research agent, Report, and **Systematic Review**
- Can find papers, screen, extract data, and write report

Elicit Search



Elicit subscription and pricing

Basic
For casual exploration

Free

- Limited access to Research Agent
- 2 Automated Reports per month
- Unlimited search across more than 138 million papers

Plus
For deeper research

$7 Save 42%

Per user/year, billed as $84 annually

Everything in Basic and...

- Increased daily access to Research Agent
- Export to RIS, CSV,

Pro
For systematic reviews

$29 Save 41%

Per user/month, billed as $348 annually

- Extended access to Research Agent
- Dedicated Systematic Review Workflow that can screen 5,000

Scale
For collaboration

$49 Save 38%

Per user/month, billed as $588 annually

Everything in Pro and...

- Full access to Research Agent
- Extract and interpret from

Enterprise
For companies & schools

Custom

- Full Access to Research Agent
- Unlimited paper alerts
- No training on your data by default
- Greater scale: screen 40,000

How we tested Elicit

Objective

- To evaluate Elicit's ability to find included studies for systematic reviews across various interventions

- We calculated the **sensitivity** and **precision** of each Elicit search and compared the results to those of the original reviews.

$$\text{Sensitivity (\%)} = \frac{\text{Number of included records retrived}}{\text{Total number of included records}} \times 100$$

$$\text{Precision (\%)} = \frac{\text{Number of included records retrived}}{\text{Total number of records retrived}} \times 100$$

Methodology

- Four reviews
  - Pharmacology - 1 review
  - Surgery - 1 review
  - Public Health - 2 reviews

- We compared the included studies by Elicit against the included studies in the four reviews

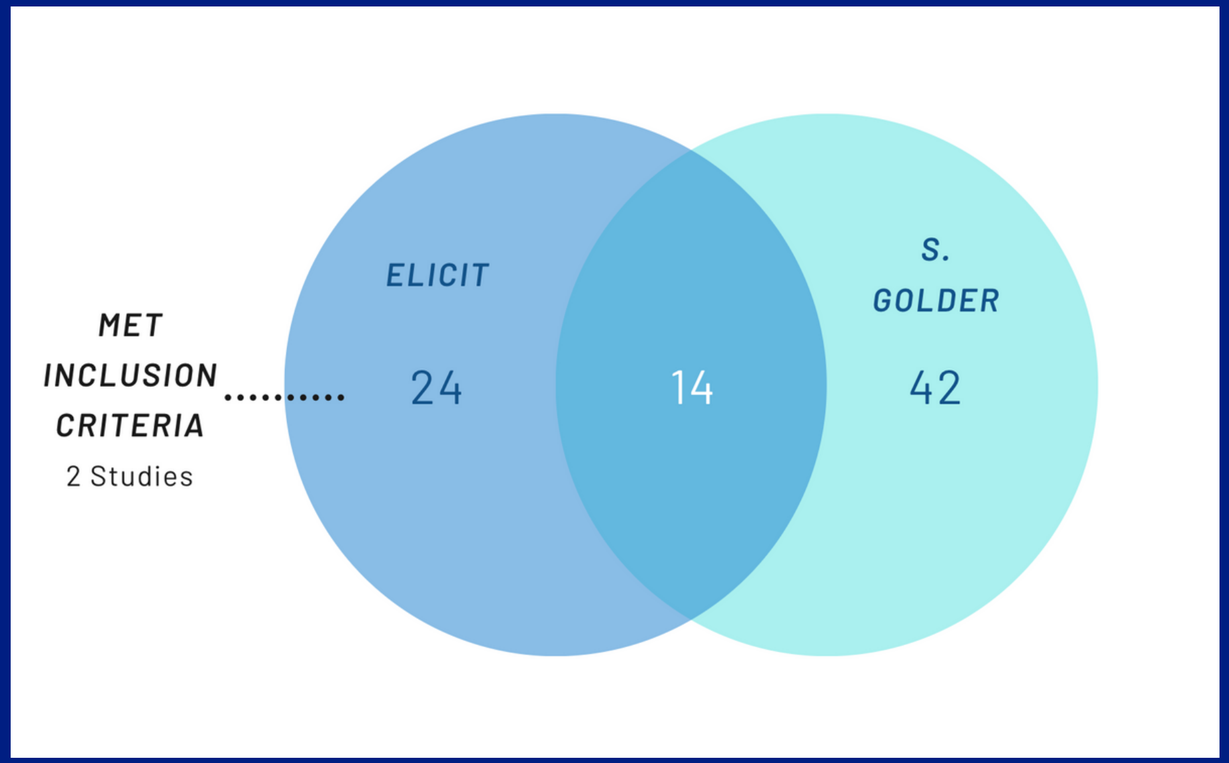
- We calculated the no. of additional studies identified by Elicit that would've been included in the original review by emailing the authors.

Case Study 1

Umbrella review: the acute and long-term harms of vaping in young people

Identification
Elicit identified 25/56 (45%) of the review's included studies

Inclusion
Elicit included 14/56 (25%) of the review's included studies



Case Study 1

Sensitivity: 27.6%

Precision: 42.1%

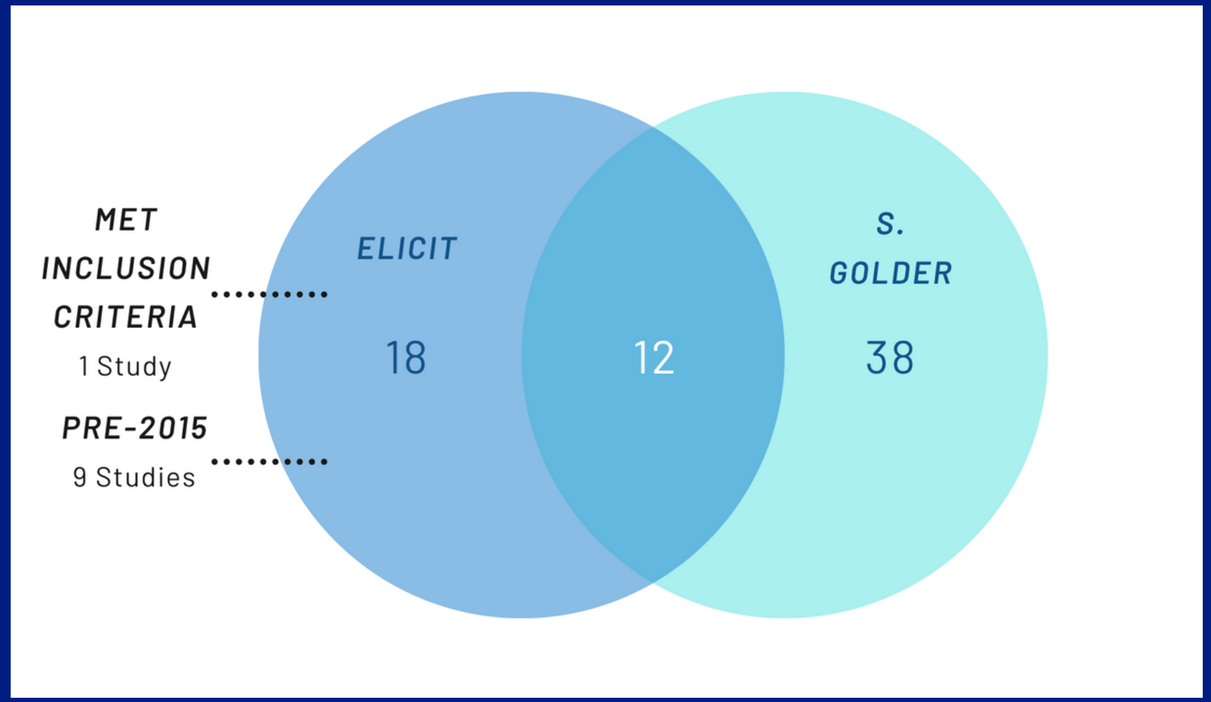
No. of additional studies identified by Elicit: 2

Case Study 2

Scoping review : cost-effectiveness of local authority-funded interventions aimed at increasing physical activity to support policy decisions

Identification
Elicit identified 23/50 (46%) of the review's included studies

Inclusion
Elicit included 12/50 (24%) of the review's included studies



Case Study 2

Sensitivity: 25.5%

Precision: 43.3%

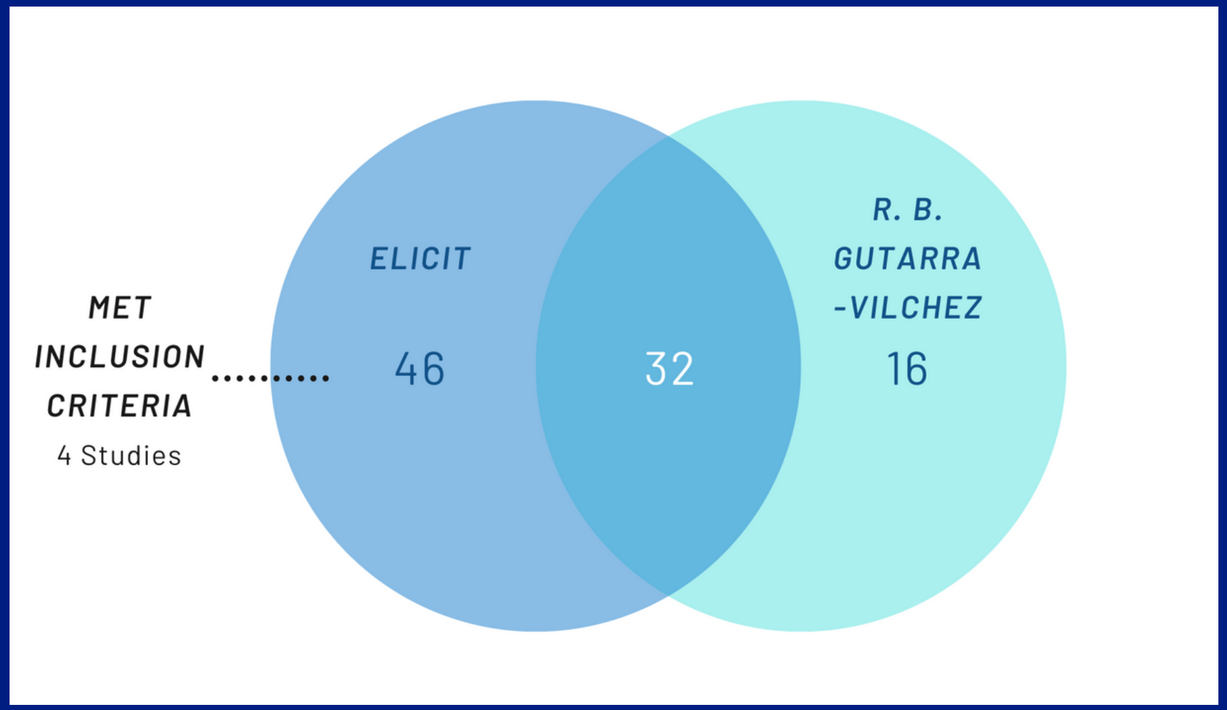
No. of additional studies identified by Elicit: 1

Case Study 3

Systematic review: evaluating the effectiveness and safety of vasodilators in women undergoing fertility treatment.

Identification
Elicit identified 36/48 (75%) of the review's included studies

Inclusion
Elicit included 32/48 (67%) of the review's included studies



Case Study 3

Sensitivity: 69.2%

Precision: 46.2%

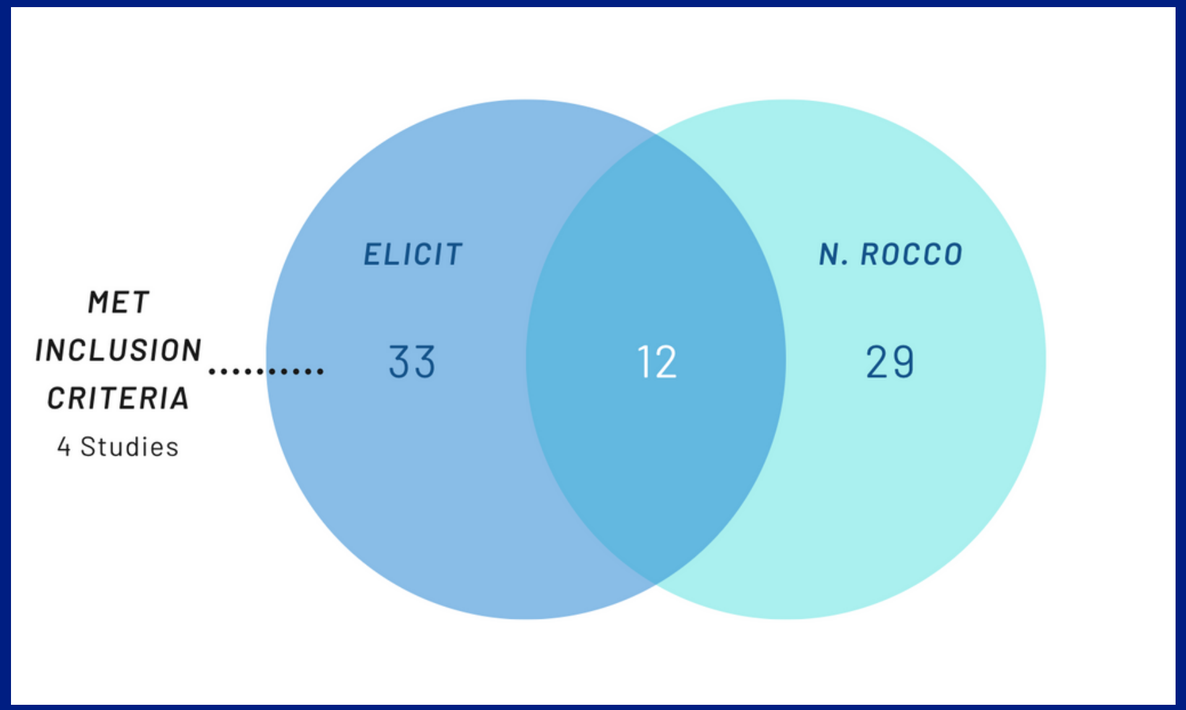
No. of additional studies identified by Elicit: 4

Case Study 4

Systematic review: assessing the effects of implants vs autologous tissue flaps for post-mastectomy breast reconstruction.

Identification
Elicit identified 20/41 (49%) of the review's included studies

Inclusion
Elicit included 12/41 (29%) of the review's included studies



Case Study 4

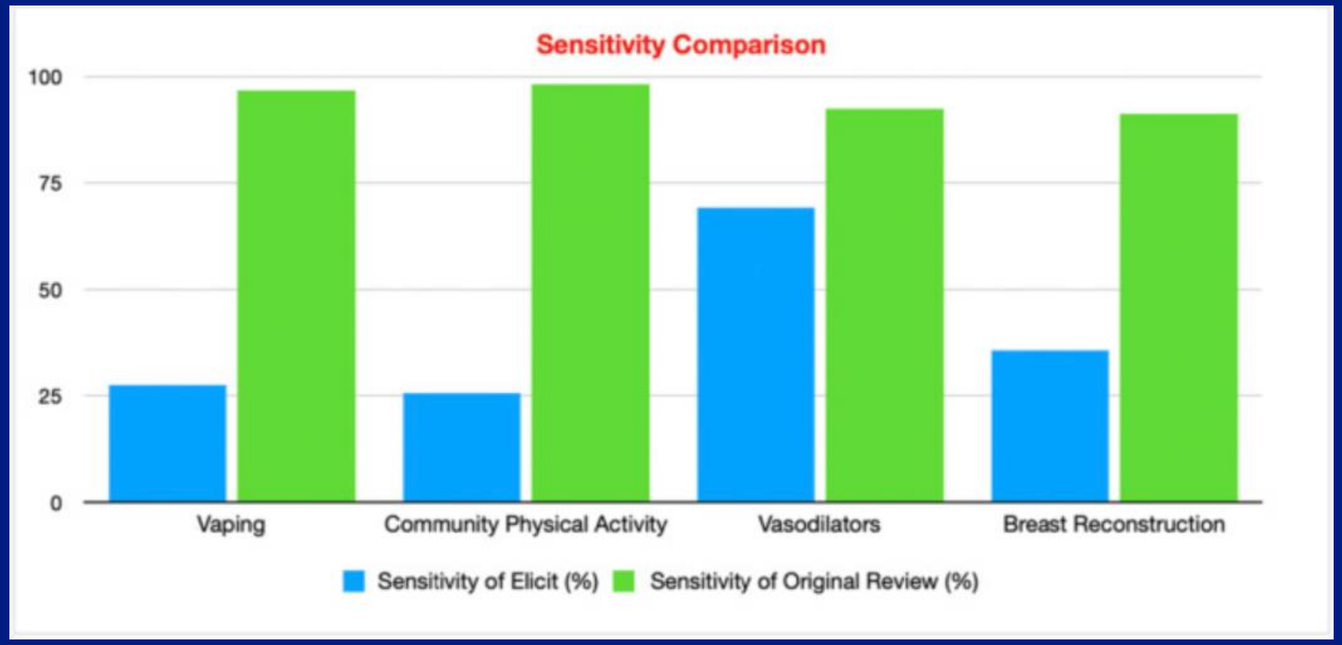
Sensitivity: 35.6%

Precision: 35.6%

No. of additional studies identified by Elicit: 4

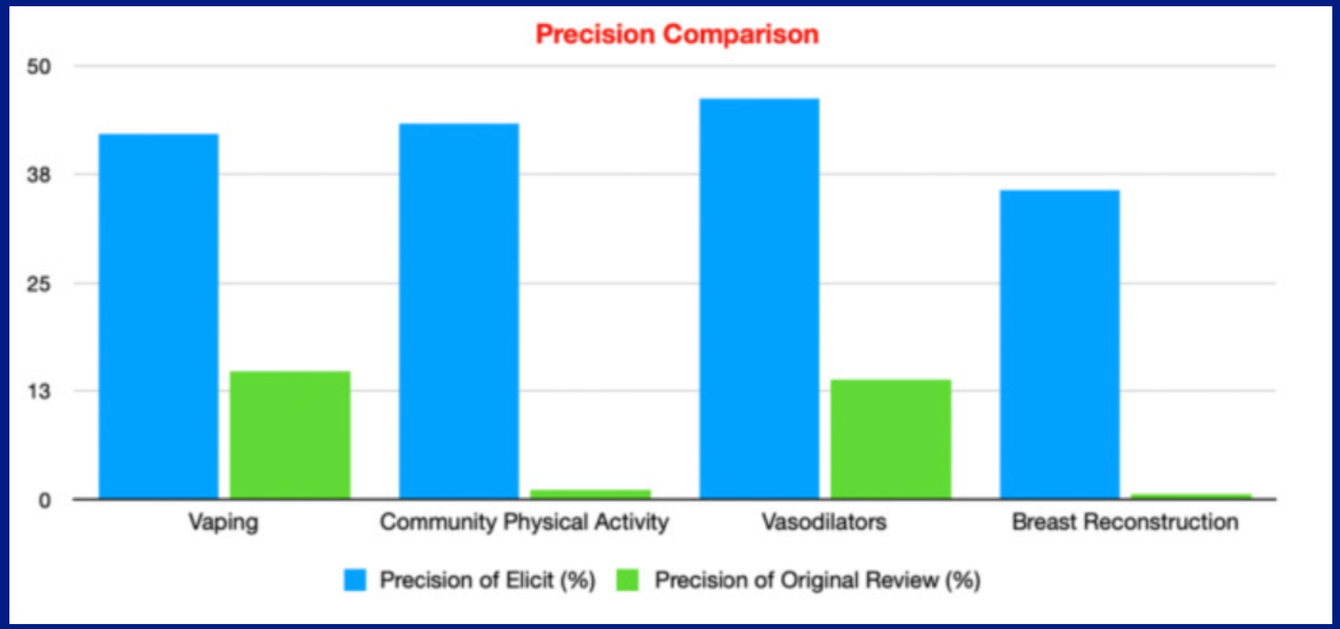
Sensitivity Comparison

- Elicit : **39.5%** (25.5% - 69.2%)
- Traditional : **94.5%** (91.1% - 98.0%)
- Elicit sensitivity lowest in public health reviews
- Elicit sensitivity highest in drug intervention review



Precision Comparison

- Elicit : **41.8%** (35.6% - 46.2%)
- Traditional : **7.55%** (0.65% - 14.7%)
- **Higher precision in Elicit**, making it useful for grant writing, scoping searches, and as supplementary source



Strengths and Limitations of Our Research

Strengths

- Independent researchers
- Real world data
- Four heterogeneous case studies
- Original authors involved

Limitations

- Single tool
- Retrospective
- No funding
- Out of date

Conclusions so far...

- **Very quick results with high precision**, making it useful for grant proposals and conducting scoping and preliminary searches
- **Unique studies** identified in Elicit demonstrate it can be a useful adjunct to traditional searching
- Currently, Elicit is best used as a **supplementary tool**, not a standalone method, due to its low sensitivity

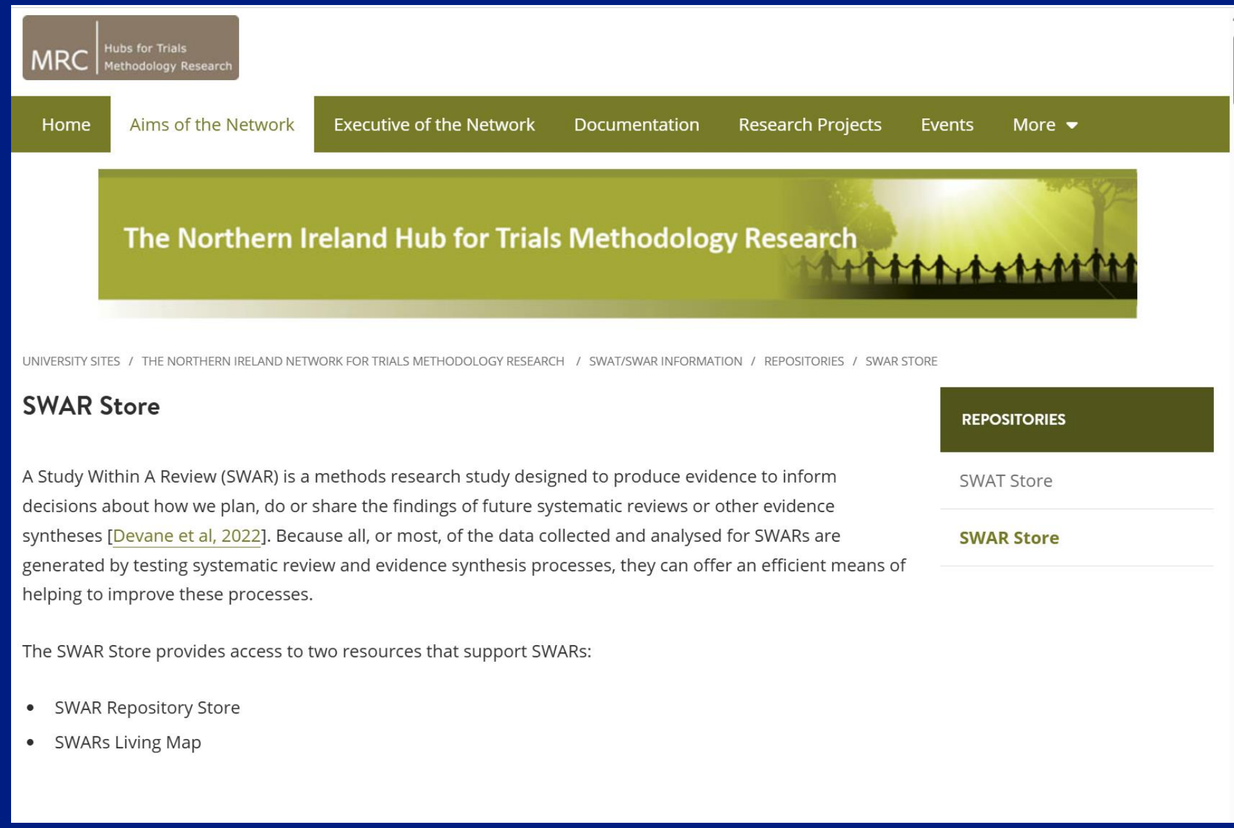
Next steps…

Evaluation of Elicit and Undermind.ai

Case examples: cancer diagnosis in primary care (PenARC, Exeter), childhood cancer (CRD), blood cultures (CRD), HTLV (CRD), PJP guidelines (CRD), spine (CRD), alcohol and CVD (DoHS), cosmetic procedures (DoHS), dermal fillers (DoHS) and social media and gambling (DoHS)

Future evaluations

Prospective study and include further tools, such as ASTA, Consensus, EMBASE AI, Scispace etc.



References

Bernard, N., Jr, Y.S., Bier, N., Lihoreau, T., Pazart, L. and Tannou, T. (2025). Using artificial intelligence for systematic review: the example of elicit. BMC Medical Research Methodology, [online] 25(1). doi:https://doi.org/10.1186/s12874-025-02528-y.

Étienne Fortier-Dubois (2025). How we evaluated Elicit Reports. [online] The Elicit Blog. Available at: https://blog.elicit.com/elicit-reports-eval/

Fenske, R.F. and Otts, J.A.A. (2024). Incorporating Generative AI to Promote Inquiry-Based Learning: Comparing Elicit AI Research Assistant to PubMed and CINAHL Complete. Medical Reference Services Quarterly, 43(4), pp.292–305. doi:https://doi.org/10.1080/02763869.2024.2403272.

Golder S, Hartwell G, Barnett LM, Nash SG, Petticrew M, Glover RE. Vaping and harm in young people: umbrella review. Tob Control. 2025 Aug 19:tc-2024-059219. doi: 10.1136/tc-2024-059219. Epub ahead of print. PMID: 40829950.

Golder S, Castro A, Dale V, Moe-Byrne T, Jefferson L, Afonso D, Hillcoat L, Barlett C, Bloor K. Population-based physical activity interventions potentially provided by local authorities in the UK: A scoping review of economic analyses. Newcastle, UK. HSR UK Conference. 2-3 July 2025

Gutarra-Vilchez, R.B., Vazquez, J.C., Demián Glujovsky, Lizaraso, F., Viteri-García, A. and Martinez-Zapata, M.J. (2025). Vasodilators for women undergoing fertility treatment. Cochrane library, 2025(3). doi:https://doi.org/10.1002/14651858.cd010001.pub4.

Hooson, M. (2025). UK Artificial Intelligence (AI) Statistics And Trends In 2025. Forbes. [online] 7 Jan. Available at: https://www.forbes.com/uk/advisor/business/software/uk-artificial-intelligence-ai-statistics/.

Moens, M. Nagels, G. Wake, N. Goudman, L. Artificial intelligence as team member versus manual screening to conduct systematic reviews in medical sciences, iScience, Volume 28, Issue 10, 2025,

Rocco, N., Catanuto, G.F., Accardo, G., Nunzio Velotti, Chiodini, P., Cinquini, M., Privitera, F., Rispoli, C. and Nava, M.B. (2024). Implants versus autologous tissue flaps for breast reconstruction following mastectomy. Cochrane library, 2024(10). doi:https://doi.org/10.1002/14651858.cd013821.pub2.

Woodhams, J. (2024). Research your AI research tools - UK Research Integrity Office. [online] Ukrio.org. Available at: https://ukrio.org/ukrio-resources/research-your-ai-research-tools/.