

Could we use the Elicit AI search tool to retrieve evidence to inform surveillance and prioritisation decisions at NICE?

Andrea Heath, Senior Information Specialist,
NICE

Also in the project team:

Evidence and Knowledge: Monica Casey, Daniel Tuvey, Jayne Jefferies

Surveillance: Kathryn Hopkins, Claire McLeod, Mike Raynor

NICE National Institute for
Health and Care Excellence



Surveillance at NICE

- **Guidance published**
 - But more studies and evidence
- **Decision: update, amend, retire**
 - Information needed to support decisions
- **Prioritisation**
 - Which topics?



Producing useful and usable guidance for health and care practitioners.



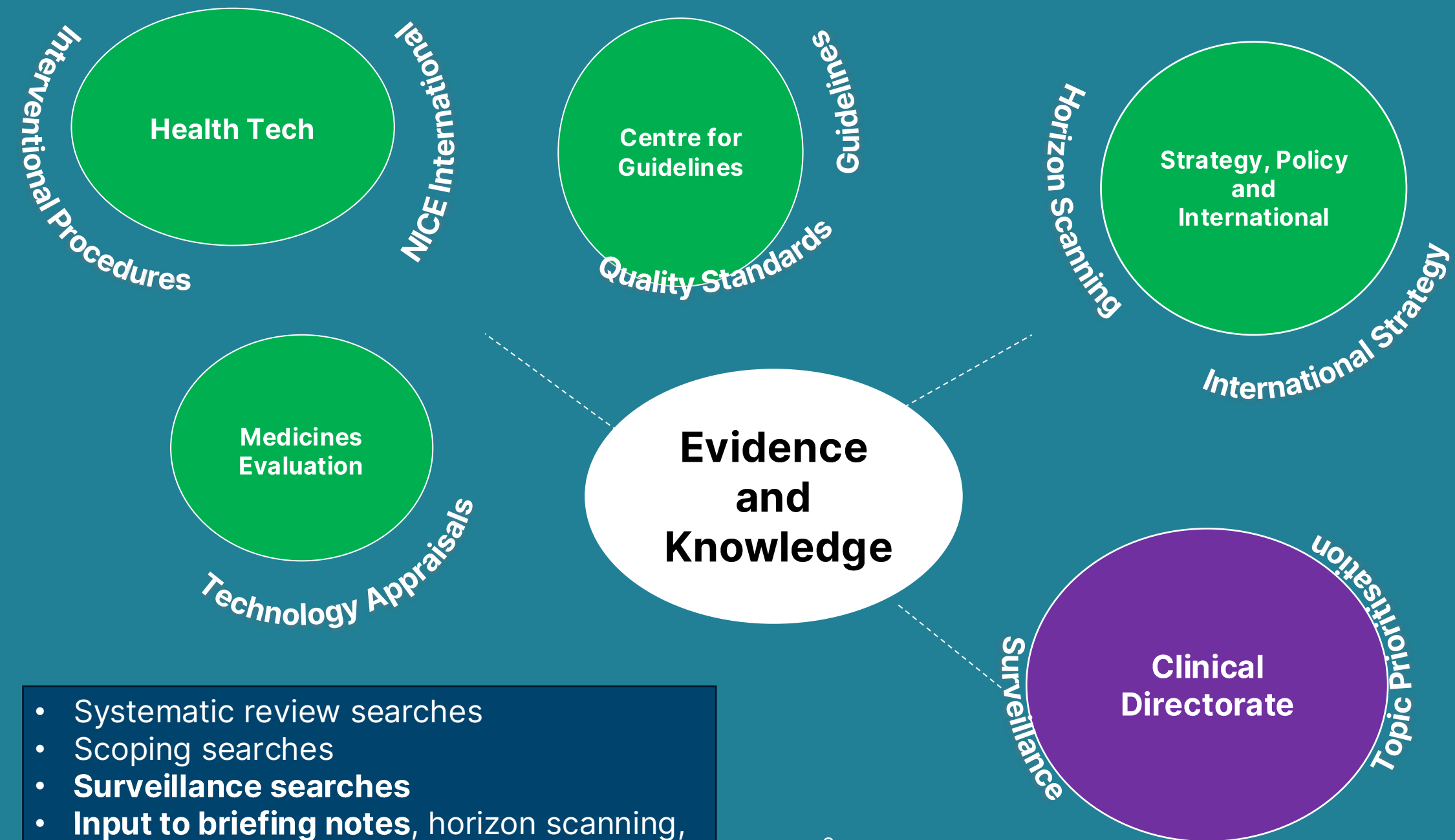
Focusing on what matters most by prioritising topics that are most important to the health and care system or address an unmet need.



Providing rigorous, independent assessment of complex evidence for new health technologies.



Encouraging the uptake of best practice to improve outcomes for everyone.



- Systematic review searches
- Scoping searches
- **Surveillance searches**
- **Input to briefing notes**, horizon scanning, knowledge dissemination

Sort: Most relevant		
Search		
Filters		
+ Column		
Export		
UPGRADE		
Save all		
Get full-texts?		
Paper	10 found	Summary
Economic evaluations of implantable cardioverter defibrillators: a systematic review L. García-Pérez, Pilar Pinilla-Dominguez, A. García-Quintana, E. Caballero-Dorta, F. Javier García-García, and 2 more European Journal of Health Economics, 2015, 11 citations Abstract only		There are studies comparing the cost-effectiveness of implantable cardiac defibrillators, * with ICDs being cost-effective for secondary prevention in high-risk patients, * but uncertainty remains for primary prevention. * The evidence is based on a systematic review of 24 studies. *
The Cost Effectiveness of Implantable Cardioverter Defibrillators: A Systematic Review of Economic Evaluations		There are studies comparing the cost-effectiveness of implantable cardiac defibrillators, * with a systematic review of 34 economic evaluations showing ICDs to be

Chat with papers

Edit columns

Summarize these papers

Copy

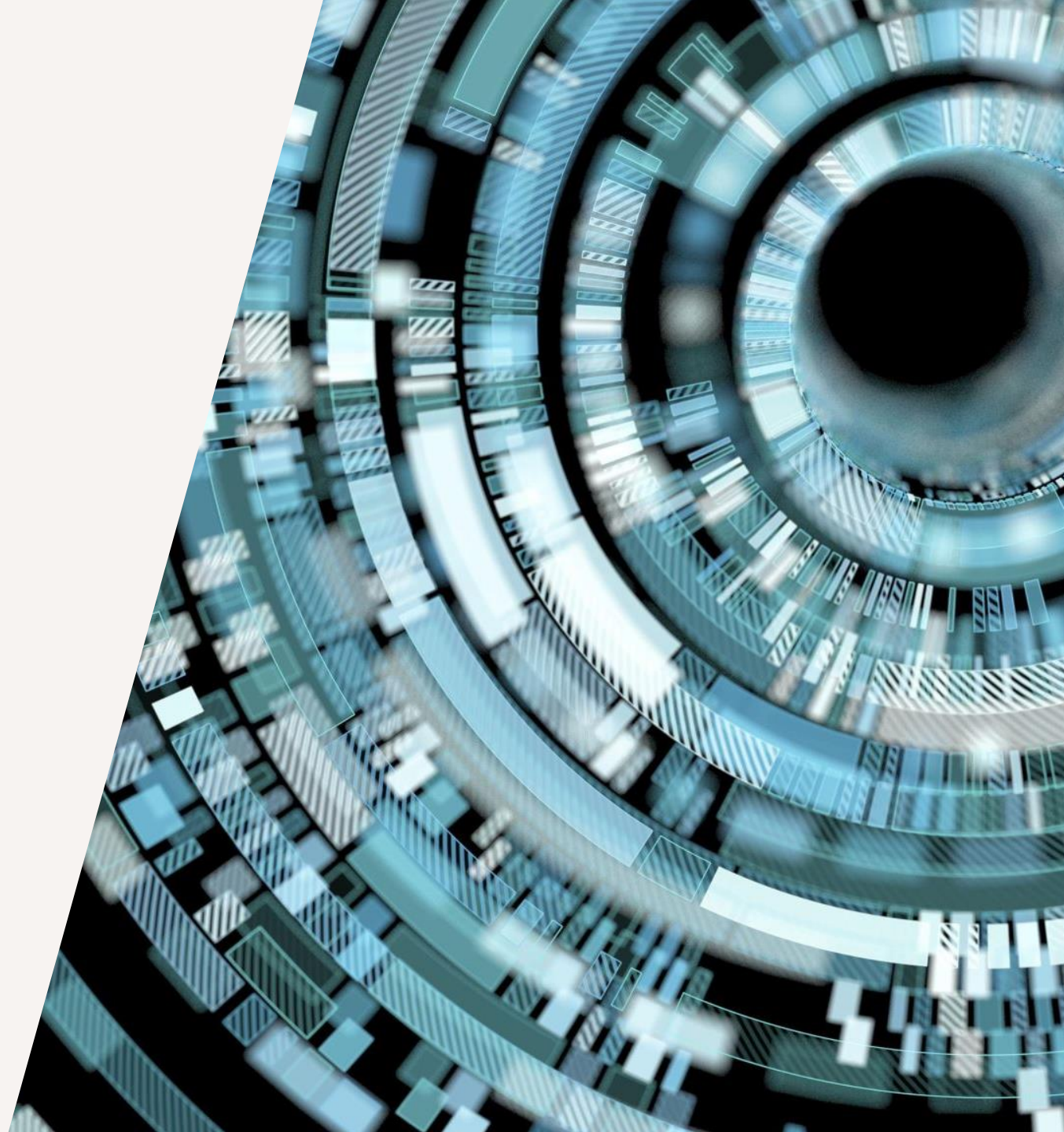
Yes, there are extensive studies comparing the cost-effectiveness of implantable cardiac defibrillators (ICDs), with multiple systematic reviews and individual economic evaluations available.

The evidence base is substantial, with systematic reviews identifying 24 studies [L. García-Pérez et al., 2015](#), 34 studies [Fotini Gialama et al., 2014](#), and 11 economic evaluations [Jackie Bryant et al., 2005](#) comparing ICDs to medical therapy or amiodarone. Individual cost-effectiveness analyses from major clinical trials include evaluations of 8 randomized trials covering thousands of patients [G. Sanders et al., 2005](#), MADIT with cost-effectiveness ratios of \$27,000 per life-year saved [Alvin I. Mushlin et al., 1998](#), and MADIT-II with 1,095 patients showing ratios of \$78,600-\$235,000 per life-year saved depending on time horizon.

Ask anything about the results

Project aim

Assess the ability of the AI search tool 'Elicit' to search for information related to the surveillance of NICE guidance and the prioritisation of topics for guidance development



Tool evaluation (Canada Drug Agency)

Criteria	Evaluation Criteria	Notes
Model details	Who is the developer? Which version?	Ought is the developer. No information on the version being tested. Only tells us that we are using the "Basic" version.
What does it do?		You ask Elicit a question. It then returns what it thinks are the most relevant papers that answer your question
Terms of use	Are there clear terms of use?	Yes
Services	Information retrieval, screening, data extraction?	Yes
Content: Source of citations	What is the underlying source?	Semantic Scholar, Pubmed and Open Alex (over 138 million academic papers).
Content: What kind/types of information is it good for	What type of study does it find?	Study types: Review, Meta-Analysis Systematic Review, RCT, Longitudinal Sematic Scholar does not index grey literature.
Content: comprehensiveness	How comprehensive is it of health and care literature?	Elicit is designed to cater to both biomedicine and social sciences. It cannot access scholarly work behind a paywall or from licensed journals.
Content: coverage	What is the date range?	The publication year filter goes back to 1971. However, there is no mention of a date range in the supporting information.



Quantitative element

- 12 topics selected
- Gold standard for comparison
- “Prevalence of [condition]”
- Evaluation by surveillance analysts



Pre-testing

- Test searches for familiarisation
- Repeatability testing

NICE



Prompts

- Prompt development
- Individual testing
- Discussion and agreement

Prompts



1. Are there any statistics showing the epidemiology, incidence or prevalence of [condition] in [population] in [country]? Filter to [date]
2. How many [people/more specific population term] have [condition] in [country]? Filter to [date]
3. What is the prevalence of [condition] in [population] in [country]? Filter to [date]



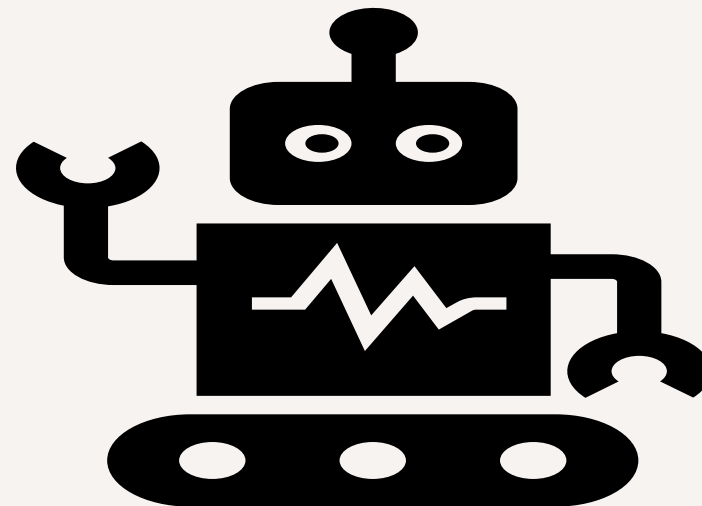
- Step 1: Country = UK, date=2020
- Step 2: Country = UK, date=2015
- Step 3: Country = 'countries similar to the UK', date=2015
- Step 4: No country or date specified

Evaluating the prompt results

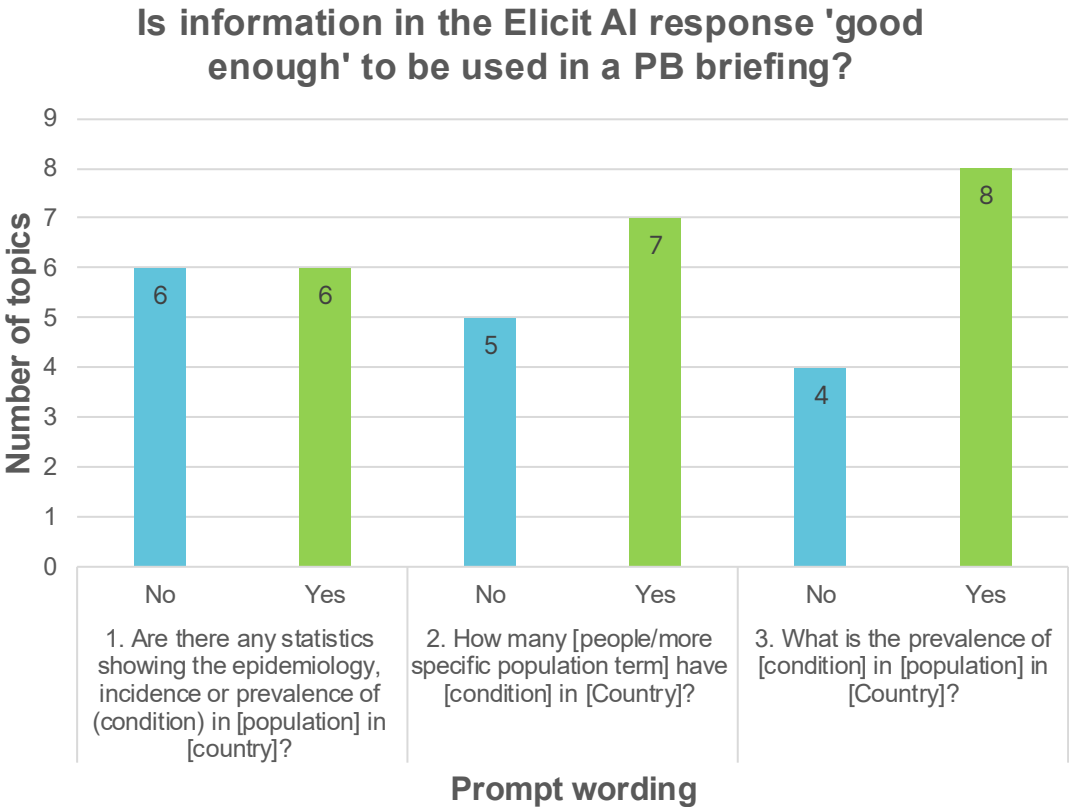
Gold standard= information included in original Priority Board briefing paper

Did the AI generated response:

- Identify the same source of information?
- ID information 'good enough' to include in a PB briefing?
 - Correct population?
 - UK data? (if ID'ed by gold standard)
 - Recent? (<5-years old)
 - Data of comparable quality ([Joanna Briggs prevalence study checklist](#))
- If the info was not good enough, why?



Quantitative results



For 8 of 12 topics, Elicit provided information judged sufficient/good enough for use in prioritisation board briefings.



For 4 topics, results were incomplete or outdated, meaning that supplementary searching would be needed.



2 topics, the data was too old, more recent data was available.
2 topics, the data was not from the UK, UK data was available.

Qualitative element

Aims: explore ease of use, currency, relevancy, transparency (Questions based on [REACT Framework](#)) N=3 Surveillance TAs

Topics : Chronic lymphocytic leukaemia, ankle surgery, meditation for PTSD, thoracic endometriosis, risk of stillbirth

Analysis

Interviews transcribed by Microsoft Co-pilot

Thematic analysis by Co-pilot, sense checked and nuanced by **very human research team**

NICE

Conclusions

- Useful for gaining basic knowledge quickly, but not in isolation

Themes

- Confidence (or lack of) in summaries
- Functionality
- Limitations
- Comparison with other search tools

Conclusions and recommendations

- Elicit may be useful as a supplementary tool for rapid topic familiarisation and preliminary evidence gathering.
- It should not replace conventional literature searches or be used as a sole source for surveillance reviews or briefing papers.
- To ensure accuracy and reliability, a human-in-the-loop approach should be used to validate and verify information outputs from Elicit.
- Procurement of the paid version is not recommended at present, as the free version meets current needs.

