When Not to Use Artificial Intelligence in Health Economics and Outcomes Research

MSR94

Tushar Srivastava, MSc[1]; Shilpi Swami, MSc[1]
[1]ConnectHEOR Limited, London, UK.
Email: tushar.srivastava@connectheor.com

connectHEOR

KEY MESSAGE

AI in HEOR is defensible when it assists, traceable, and governed. It isn't when it decides, obscures, or operates without accountability. Hence, AI as a productivity aid in HEOR is conditionally acceptable. However, AI as a decision-maker is not. The line is whether human judgement remains traceable and accountable.

BACKGROUND

As artificial intelligence (AI) adoption expands within health economics and outcome research (HEOR), the risk of inappropriate application increases, particularly in contexts requiring transparency, traceability, and methodological defensibility for HTA.

Recent guidance from NICE states that AI should be used in evidence generation only where there is demonstrable value and where risks such as reduced human oversight, lack of transparency, and added methodological complexity are adequately managed. [1]Evolving frameworks such as hybrid intelligence stress the importance of human-AI managed collaboration. [2]

WHO guidance emphasizes transparency, accountability, safety, and the protection of public interest. ISPOR good-practice work in modeling continues to treat transparency and validation as central to the credibility of decision models.[3]

OBJECTIVES

This study aims to:

1. Define explicit exclusionary criteria for AI use in HEOR
2. Identify where AI should not be used because the workflow, evidence standard, or decision consequence is incompatible with current AI limitations
3. Distinguish AI as a productivity aid from AI as a decision-maker
4. Establish a threshold standard for acceptable AI use in HTA-facing work, where outputs must be explainable and open to scrutiny by assessors and reviewers

FIVE ASSESSMENT DIMENSIONS

Decisional Criticality	Transparency	Data Quality	Stability	Consequences of Error
Impact on reimbursement / policy	Documented & reproducible	Completeness & provenance	Consistency across settings	Impact on submissions and patient access

METHODS

Conceptual Analysis Approach

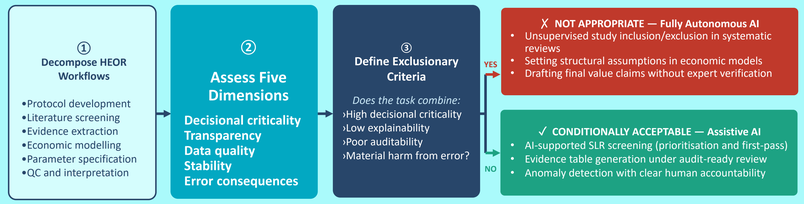
A conceptual analysis was conducted drawing on HTA methods, literature, regulations, and AI governance frameworks.

Core HEOR workflows were decomposed into discrete decision points and AI suitability assessed across five dimensions.

Workflows mapped include protocol development, literature screening, evidence extraction, evidence synthesis, economic model structuring, parameter specification, quality control, and interpretation of results.

Exclusionary criteria were defined when a task combined high decisional criticality with low explainability, poor auditability, unstable source inputs, or potentially material harm from undetected error.

Assessment Framework



AI Admissibility Conditions

AI was considered conditionally admissible only where:

- ✓ **Outputs could be verified against source documents**
- ✓ **Reviewed by domain experts**
- ✓ **Embedded within documented governance process**
- ✓ **Clear human accountability maintained**

Consistent with current evidence-synthesis guidance permitting AI/automation only when methodological rigor is not compromised and human oversight is maintained.

KEY FINDING

The central policy question is not whether AI can perform a task, but whether its use preserves the standards of transparency, reproducibility, accountability, and methodological defensibility expected in HTA and related decision settings.

RESULTS

6
Inappropriate AI Use Cases

✗ Inappropriate AI Uses (Fully Autonomous)

Unsupervised study inclusion/exclusion
Final include/exclude decisions in systematic reviews require traceable human judgement

Undisclosed evidence interpretations
Generating evidence interpretations without disclosure risks opaque HTA submissions

Structural assumptions in economic models
Setting model structure autonomously undermines reproducibility and peer review

Resolving conflicting clinical data
Autonomous conflict resolution masks clinical uncertainty critical to payer decisions

Drafting final value claims
Unverified AI-drafted value claims cannot meet HTA evidentiary standards

Sparse or heterogeneous source material
Unstable inputs amplify AI hallucination risk and compromise evidence validity

5
Conditionally Acceptable Uses

✓ Conditionally Acceptable (Assistive AI)

Evidence Synthesis & SLR
Literature triage and first-pass screening; evidence table pre-population. **Condition:** Final inclusion decisions and data verification retained by trained reviewers.

Model Quality Control
Automated error detection and consistency checks; cross-checking inputs against source documents. **Condition:** Human sign-off required before submission.

Model Parameterisation
AI-assisted extraction of clinical efficacy data with source traceability; sensitivity analysis scenarios for expert review. **Condition:** Analysts retain structural control.

Technical Report Generation
AI-assisted drafting of HTA technical reports from structured outputs; template population under post-processing review. **Condition:** Expert review essential

Workflow Execution
Repetitive formatting, reference checking, and data structuring tasks. **Condition:** Low decisional criticality, reversible outputs, documented governance process.

DISCUSSION & CONCLUSIONS

The analysis distinguished between AI as a productivity aid and AI as a decision-maker [2]:

Productivity aid : may be defensible when it accelerates repetitive tasks under audit-ready review.

Decision-maker : problematic when it obscures rationale, weakens reproducibility, or prevents assessors from understanding how a conclusion was reached.

A precautionary boundary is warranted for any application in which AI would replace, rather than support, expert judgment on high-consequence evidentiary or modeling decisions.

Framing AI adoption around exclusion criteria can help organizations deploy assistive tools where they add value while protecting the integrity of reviews, models, and submissions intended to inform payer or policy decisions.

References:

1. NICE. *Use of AI in evidence generation: NICE position statement*. National Institute for Health and Care Excellence; 2024.
2. Srivastava T, Irfan H, Babiy V, Swami S. Integration of generative AI with human expertise in HEOR: a hybrid intelligence framework. *Adv Ther.* 2025;42:4103–4130. doi:10.1007/s12325-025-03273-w
3. World Health Organization. *Ethics and governance of artificial intelligence for health*. Geneva: WHO; 2021.
4. Flemyng E, Noel-Storr A, Macura B, et al. *Position statement on artificial intelligence (AI) use in evidence synthesis across Cochrane, the Campbell Collaboration, JBI and the Collaboration for Environmental Evidence*. Cochrane Database Syst Rev. 2025.

Financial Disclosure:
Authors are employees of ConnectHEOR Limited. No external funding received. No conflict of interest to declare.

Poster presented at ISPOR in Pennsylvania, Philadelphia, USA, May 17-20, 2026