Automating Model Parameterisation for Health Economic Models: Validation of a Gen-AI RAG Pipeline

MSR247

Tushar Srivastava, MSc[1]; Hanan Irfan, MSc[1]; Hemansh Sridhar, BTech[2]; Shilpi Swami, MSc[1]

[1]ConnectHEOR Limited, London, UK; [2]ConnectHEOR, Delhi, India.

Email: tushar.srivastava@connectheor.com

connectHEOR

KEY MESSAGE

With 60–70% faster health economic model parameterisation and 88%+ exact-match accuracy, AI-assisted evidence extraction shows real promise, not as a replacement for expert judgement, but as a scalable, traceable foundation for the same,

BACKGROUND

Parameterisation of health economic models requires systematic extraction, synthesis, and validation of inputs from RCTs and published literature, a process that is resource-intensive and prone to inconsistency. [1]

In health technology assessment, parameterisation typically requires:

- Transparent identification of evidence sources,
- Reproducible transformation of published values into model-ready inputs, and
- Explicit handling of heterogeneity and uncertainty.[1]

RAG systems combine document retrieval with grounded generation, making them more suitable than standalone LLM workflows for evidence-intensive healthcare tasks where provenance, traceability, and reduced hallucination risk are essential. [2,3]

OBJECTIVES

This study evaluated a Gen-AI RAG pipeline designed to support automated parameterisation of health economic models, benchmarking its performance against manual extraction and validation.

1. Evaluate data accuracy and completeness of AI-extracted parameters.
2. Assess traceability to source evidence.
3. Determine suitability for direct use in cost-effectiveness model parameterization.

METHODS

Pipeline Design

- Evidence repository restricted to analyst-approved documents: trial publications, supplementary appendices, HTA dossiers, and SLRs.
- Documents ingested with section-level chunking and metadata tagging (study ID, intervention, comparator, population, line of therapy, follow-up period, endpoint type, source).
- Retrieved passages passed to generation layer only when satisfying predefined relevance thresholds.

Validation Framework

- Case studies spanning partitioned survival and state-transition models across multiple therapeutic areas.
- Gold-standard manual reference set with model-ready values, source locations, and contextual qualifiers.
- Three evaluation domains: (1) accuracy & completeness, (2) traceability, (3) suitability for direct use.
- Two independent reviewers with third-reviewer adjudication.
- Errors categorized: retrieval failure, partial extraction, incorrect contextual interpretation, source misattribution, numerical transformation error, duplication/omission.
- Time-and-motion exercise comparing manual vs. AI-assisted workflows



Fig 1: Gen-AI RAG Pipeline Workflow

RESULTS

>88%	60-70%	>95%
Exact-match rate for clinical efficacy	Time reduction vs. manual workflow	Citation traceability

Exact-Match Accuracy by Some Sample Parameter Types (%)

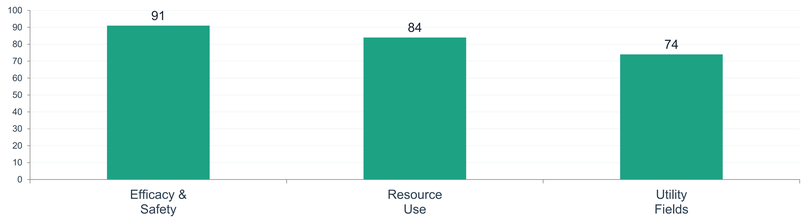


Fig 2: Exact match accuracy (%) by parameter type

Model-Readiness Assessment

Model readiness assessment evaluated whether AI-extracted parameters could be directly used in economic models, with or without minor analyst correction, based on accuracy, contextual alignment with source evidence, and completeness of required inputs for decision-making.

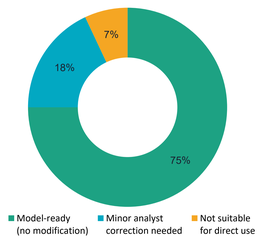


Fig 3: Model readiness assessment of extracted parameters

Analyst Time Comparison (in hours)

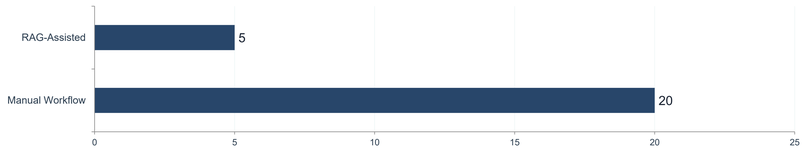


Fig 4:Analysts' time spent in manual vs RAG assisted workflow (hours)

MOST FREQUENT RESIDUAL ERRORS

- Incorrect subgroup carryover.
- Failure to distinguish median from mean follow-up.
- Omission of derivation notes for utility or cost inputs.
- Discrepancies primarily related to contextual interpretation (e.g., subgroup definitions, outcome timing).

RESULTS (CONTINUED)

Pipeline Performance Summary (Illustrative)

Metric	Value
Case studies	4
Prespecified parameters	312
Relevant source retrieval	94–97%
Efficacy and safety exact-match	89–93%
Resource-use exact-match	80–88%
Utility exact-match	68–79%
Overall citation traceability	>95%
Model-ready without modification	72–78%
Model-ready after minor correction	15–22%
Not suitable for direct use	5–10%

DISCUSSION and CONCLUSIONS

Key Findings

The Gen-AI RAG pipeline demonstrated potential to substantially accelerate model parameterisation while maintaining traceability to published evidence.

When used as a decision-support tool alongside expert review, such approaches may improve efficiency and consistency in health economic model development without compromising methodological rigor.

Greatest Near-Term Value

- Semi-automated evidence structuring.
- Citation-linked extraction.
- Quality-controlled prepopulation of parameter tables.

Future Directions

- Test generalizability across broader therapeutic areas
- Non-English evidence sources
- Indirect treatment comparison inputs
- Complex tasks: survival extrapolation assumptions, utility mapping

References

1. Briggs AH, Weinstein MC, Fenwick EAL, et al. Model parameter estimation and uncertainty analysis: ISPOR-SMDM modeling good research practices task force-6. Value Health. 2012;15(6):835-842.
2. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. 2020.
3. Bitterman DS, Ong JCL, Ting DSW, Liu N. Retrieval-augmented generation for generative artificial intelligence in healthcare. npj Health Systems. 2025;2:2.

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