AI Agent for Automated Quality Check of MS Excel Based Cost-Effectiveness Models

MSR146

Tushar Srivastava, MSc[1]**;** Hanan Irfan, MSc[1]; Kunal Hriday, MSc,[2]; Vikas Badola, BTech[2]; Shilpi Swami, MSc[1]

Affiliations: [1]ConnectHEOR, London, UK;[2] ConnectHEOR, Delhi, India;
Email: tushar.srivastava@connectheor.com

connectHEOR

KEY MESSAGE

Near-perfect error detection. 90% faster. An AI QC agent delivered consistent, traceable quality assurance across Markov and Partitioned Survival models - at a scale manual review can't match.

BACKGROUND

Quality control (QC) of Excel-based cost-effectiveness (CE) models is essential for HTA credibility but remains **manual, time-consuming, and inconsistently documented**.

Current challenges include:

- Manual review processes prone to human error
- Inconsistent documentation of QC findings
- Time-intensive validation of complex model logic
- Difficulty testing dynamic model recalculations [1,2]

OBJECTIVES

1. Evaluate an AI-driven QC system that ingests CE models in Excel and executes a predefined QC checklist
2. Compare the AI agent's error detection performance against human validators with 3 years of industry experience in health economic (HE) modeling

METHODS

AI Agent Architecture

1. **Checklist Interpretation**
   AI agent interprets each checklist item to infer expected behavior and determines applicability to the model type
2. **Reasoning Layer**
   Generates a stepwise plan: identify worksheets/ranges, extract values, compare outputs, perform independent calculations
3. **Executor Layer**
   Performs static data extractions and dynamic changes (e.g., changing dropdowns/parameters, recalculating) to validate response patterns
4. **Output Generation**
   For each check: pass/fail result, supporting observations, and corrective recommendations

Model Testing

Tested on two complex models:

- Markov Model
- Partitioned Survival Model (PSM)
- 20 errors of varying difficulty manually seeded across both models, ranging in difficulty

RESULTS

100% (20/20)
Markov Model
Error Detection

95% (19/20)
PSM
Error Detection

>90%
Time Reduction
vs Manual Review

Error Detection Accuracy: AI Agent vs Human Review

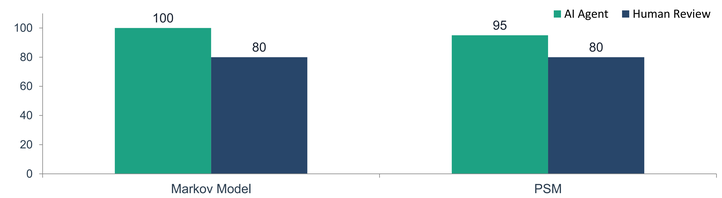


Fig 1: Accuracy of error detection in human vs AI review

Sample Error Findings:

Markov Model: Moderate-to-severe rheumatoid arthritis | Markov Model (8 health states) | 36 worksheets | ~150,000 cells | Lifetime horizon | Quarterly cycles | Healthcare payer perspective

- In Transitions_LongTerm, cells F112:F160 (cycles 21–40) referenced Inputs!D27 (single value) instead of the normalised probability range Inputs!D27:D34. Outgoing probabilities for the Moderate state summed >1.0 in multiple cycles.
- In Trace_Population!M45:M520, total cohort summed C45:J520, excluding the Death state in column K. From cycle 18 onward, total population declined below initial cohort of 1,000 patients.
- in Trace_Population!F45:F520 applied at cycle start (before state transitions), causing premature patient removal and understating accrued costs and QALYs.

PSM: Metastatic NSCLC | Partitioned Survival Model (PSM: PF, PD, Death) | 32 worksheets | ~120,000 cells | 20 VBA modules | 30-year lifetime horizon | Weekly cycles | Payer perspective

- In the Trace_QALY sheet, early-state QALY cells G45:G520 referenced Inputs!C18 (advanced disease utility) instead of Inputs!C15 (early disease utility).
- In Trace_Costs, administration cost cells H60:H520 applied a constant per-cycle cost without conditioning on treatment exposure or restricting to cycle 0
- Discounting was applied within some per-cycle cost formulas (drug acquisition) while other cost components were discounted only at accumulation stage.

Notable: The agent tested complex logic requiring model recalculation: a task typically challenging for even human auditors. Automated QC completed in <2 hours per model.

DISCUSSION

Across Markov and PSM case studies with controlled fault injection, the AI-driven, reasoning-based QC approach demonstrated:

- Detected both surface-level spreadsheet errors and deeper logic failures
- Strong performance across both model types, with 100% detection in Markov and 95% in PSM
- System outperformed manual review, which identified 80% of seeded errors
- Significant efficiency gains (>90% time reduction)
- Ability to test complex dynamic logic requiring model recalculation
- Structured outputs with observations and recommendations improve explainability and audit readiness
- Scalable, transparent, and reproducible QC outputs

These findings support the potential role of agentic AI systems in complementing traditional expert review in HTA workflows.

CONCLUSIONS

AI Agent
95–100% accuracy
<2 hrs per model

Human Review
~80% accuracy
20+ hrs per model

While AI surfaces issues at scale, human experts need to apply clinical and methodological judgement. This combination preserves accountability while dramatically increasing throughput and consistency.

These findings support broader adoption of AI-first QC pipelines in HEOR, particularly for complex models submitted to HTA bodies such as NICE, where model errors carry significant regulatory and reimbursement consequences.

Key Takeaways:

- AI-driven QC demonstrated strong accuracy and efficiency gains
- This approach complements , rather than replaces, traditional expert review in HTA workflows.
- AI-powered QC represents a paradigm shift from expert-bottlenecked manual checking to a scalable, structured, and reproducible validation workflow.
- Both PSM and Markov model case studies confirm that the agent consistently identifies high-severity issues that materially affect ICERs and QALY estimates.

References

1. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, et al. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value Health*. 2003;6(1):9–17.
2. Abraham J, et al. Under-reporting of validation efforts for health economic models persists despite the availability of validation tools: a systematic review. *Pharmacoeconomics*. 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

Public