A Model Explainer Agent to Reduce the Comprehension Burden on HTA Assessors Reviewing Economic Model Submissions: A Case Study in Rare Disease

MSR166

Tushar Srivastava, Hanan Irfan, Thaison Tong, Shilpi Swami

ConnectHEOR, London, UK. Email: tushar.srivastava@connectheor.com

> **In a simulated first-pass HTA review of one rare disease Excel model, with one agent-assisted and one manual reviewer, the explainer agent located and explained 29 of 30 model elements and cut time to familiarisation from about 8 hours to 1.5 hours, with every explanation traced to a cell or document; appraisal stayed with the assessor.**

1 BACKGROUND

An assessor receives a workbook they did not build and must work out how it operates, within a fixed procedural window [1,2], before critique can begin; in our simulation this familiarisation took about 8 hours.

- Technical errors are common in company models: only 2 of the 41 NICE single technology appraisals completed in 2017 had no reported errors [3], so this step is effortful and error-prone [4], and assessor capacity is a recognised constraint on HTA bodies.[5]
- In **rare disease**, evidence is thin and structural assumptions carry more of the result, so the demand is higher still.[6]

Evidence gap: NICE and ISPOR frame AI as augmentation under human oversight [7,8], but whether it can compress model comprehension without touching appraisal is untested.

2 OBJECTIVES

We tested a **model explainer agent**, a conversational AI co-worker that interrogates and explains a submitted model in natural language, against three aims.

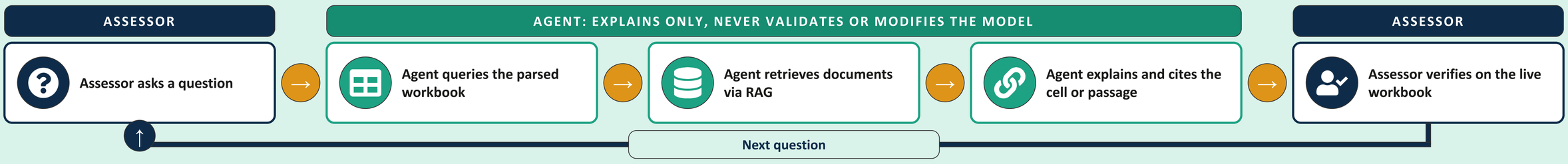
1 COMPREHENSION
Test whether the agent eases assessors' comprehension of an unfamiliar model.

2 CLARIFICATION
Measure drafting speed and coverage of a senior assessor's question list.

3 BOUNDARY
Show that these gains need not displace methodological appraisal.

3 METHODS

We simulated an HTA body's **first-pass review**, putting the same submission and the same reference standards in front of an agent-assisted and a manual reviewer. The agent parsed the workbook into an interrogable representation and indexed the documentation through **retrieval-augmented generation (RAG)**; it explained and validated the model but never modified it.



- The test case was a production-grade Excel **rare disease** model: a Markov cohort with 5 states, a lifetime horizon and about 150 000 cells.
- Parsing let answers resolve against the workbook itself, while RAG grounded generated text in retrieved passages rather than parametric memory.[9]
- Reference standards were fixed in advance: **30 model elements** an assessor must understand, and clarification questions set by a senior assessor.
- Outcomes were familiarisation time, drafting time, coverage and traceability, the last primary because an uncheckable explanation is of no use to an assessor.

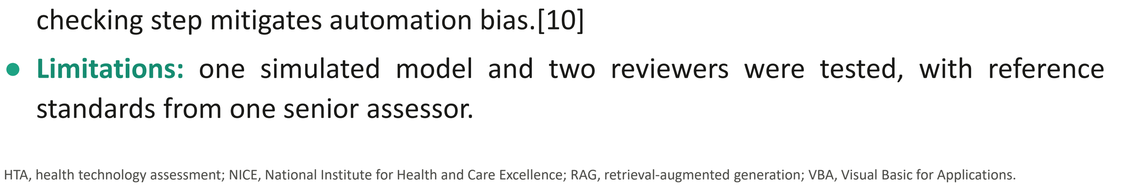
4 RESULTS

The agent-assisted reviewer located and explained **29 of 30** model elements (97%) on first pass, including elements in Visual Basic for Applications (VBA)-automated logic, where the calculation sequence is not visible in the cells.

97% of model elements explained on first pass

~80% less time to confirmed familiarisation

~70% less clarification-question drafting time



Familiarisation fell from about **8 hours to 1.5 hours** on the same submission: a working day of orientation against part of a morning, with confirmation as the endpoint.

Safeguard	Finding
Traceability	Every explanation traced to a real cell or documented passage
Fabrications	None
Residual error	One, caught on verification
Appraisal	Stayed with the assessor

Answers were built from the parsed workbook and indexed documents, so each assertion came with a location to open. The single residual error was **caught on verification**.

- Drafting time fell by about 70% while coverage of the senior reference question list rose to **29 of 30**, so faster drafting did not buy a thinner question set.
- Appraisal stayed with the assessor: nothing in the run substituted for the methodological judgements that constitute assessment.

5 DISCUSSION

The familiarisation gain follows from what the bottleneck is: orienting to an unfamiliar workbook is a **search problem**, so the cost is navigation, not judgement.

- An indexed, queryable model removes most of that cost while leaving judgement untouched, so appraisal can stay where it was.
- Grounding limits assertions to what can be retrieved [9], which matters in HTA, where a reviewer must reproduce a single statement.
- The explain-only boundary fits NICE's augmentation principle [7], and a retained checking step mitigates automation bias.[10]
- **Limitations:** one simulated model and two reviewers were tested, with reference standards from one senior assessor.

HTA, health technology assessment; NICE, National Institute for Health and Care Excellence; RAG, retrieval-augmented generation; VBA, Visual Basic for Applications.

6 CONCLUSION

The agent gave fast, source-traceable understanding of an unfamiliar submission in a simulated review, cutting familiarisation by about 80%.

- **Time moved to critique:** drafting fell ~70% and question coverage reached 29 of 30.
- **Traceability kept it safe:** no fabrications, and the one residual error was caught on verification.
- **Judgement stayed human by design:** it speeds comprehension, not appraisal.

Next step: prospective evaluation with HTA reviewers across submissions.

REFERENCES
1. Paulden M. PharmacoEconomics. 2026;44.
2. NICE. Technology appraisal manual (PMG36). 2022.
3. Radeva D, et al. Int J Technol Assess Health Care. 2020;36(4):311–316.
4. Chilcott J, et al. Health Technol Assess. 2010;14.
5. Desmet T, et al. Front Pharmacol. 2024;15.
6. Nestler-Parr S, et al. Value Health. 2018;21.
7. NICE. Use of AI in evidence generation (ECD11). 2024.
8. Fleurence RL, et al. Value Health. 2025;28.
9. Lewis P, et al. Adv Neural Inf Process Syst. 2020;33.
10. Goddard K, et al. J Am Med Inform Assoc. 2012;19.

Financial disclosure: All authors are employees of ConnectHEOR Limited. No external funding was received for this work. The authors have no conflict of interest to declare.

Poster presented at ISPOR Europe 2026 in Vienna, Austria, 8-11 November 2026