Simulating Sponsor-HTA Conversations: A Multi-Agent GenAI Framework for Reimbursement Processes

Presented by:

Barinder Singh, Director, Pharmacoevidence
barinder.singh@pharmacoevidence.com

AUTHOR(S): **Barinder Singh, RPh**[1], Gagandeep Kaur, MPharm[1], Pankaj Rai, MS[1], Rajdeep Kaur[1], Nicola Waddell, HNC[2], Shubhram Pandey, MSc[1].
[1]Pharmacoevidence Pvt. Ltd., Mohali, India, [2]Pharmacoevidence Pvt. Ltd., London, United Kingdom.

OBJECTIVES: Health technology assessments (HTA) agencies like National Institute for Health and Care Excellence (NICE) evaluate sponsor-submitted evidence to inform reimbursement decisions, with a key focus on identifying evidence gaps through clarification questions. This study aimed to develop and validate a multi-agent Generative artificial intelligence (GenAI) framework that simulate sponsor HTA conversations by generating clarification questions using sponsor-submitted data.
METHODS: A role-based, multi-agent large language models (LLMs) framework was developed to simulate sponsor-HTA interactions, featuring specialized AI agents representing HTA secretariat (lead), clinical, economic, PRO and patient/public involvement reviewers. The HTA secretariat coordinated the process and compiled clarification questions based on HTA guidelines. Using sponsor-submitted data for hepatocellular carcinoma (HCC), the tool generated questions across clinical, economic, and textual domains. Subject matter experts (SME) validated the outputs for relevance, accuracy, traceability, and alignment with HTA expectations.
RESULTS: Two NICE HTA submissions in first-line advanced/metastatic HCC were selected for sponsor-HTA simulation. The multi-agent GenAI system successfully simulated interactions by generating structured clarification questions based on sponsor-submitted evidence. A total of 104 questions were generated including 29 clinical, 62 economic, and 13 textual. The majority of questions focused on data-related gaps, while a smaller proportion sought clarification on methodological aspects. The AI-generated output was validated by a SME, demonstrating 80-85% agreement with actual clarification questions previously raised by NICE in the technology appraisals. The remaining 15-20% of questions, although not identical to historical NICE examples, were still relevant and evidence-based, reflecting potential gaps that would warrant further clarification within the HTA process.
CONCLUSIONS: The multi-agent GenAI framework provides a scalable, regulatory-aligned method for simulating NICE HTA reviews by generating structured clarification questions. It enhances sponsor preparedness and may streamline assessments, with future updates to include budget impact and real-world evidence.

Imagine if we could anticipate the HTA's questions – *before* the assessment begins

Pharmaco Evidence

HTA submissions are complex and time intensive

It involves processing vast amounts of clinical, economic, and real-world data within tight timelines to meet submission deadlines

The HTA submission process is highly time- and labor-intensive, requiring extensive preparation of documentation and analyses

Post-submission, EAG review and feedback may lead to additional iterations, extending the time to final decision

The EAG review process

Step	Stage	Key Output
1	Scoping	Final scope
2	Company submission	Evidence dossier
3	EAG review	Assessment report
4	Technical engagement	Resolution of issues
5	Appraisal committee	Draft → Final guidance
6	Publication	NICE final guidance
7	Implementation	NHS adoption

Can we streamline this process, without compromising scientific rigor?

Pharmaco Evidence

Reimagining HTA conversations

- GenAI was used to develop and validate an LLM model that could identify evidence gaps in an evidence submission and predict clarification questions from the EAG
- A role-based, multi-agent LLMs framework was developed to simulate sponsor-HTA interactions, featuring specialized AI agents representing EAG lead, clinical, economic, ITC experts and advisors

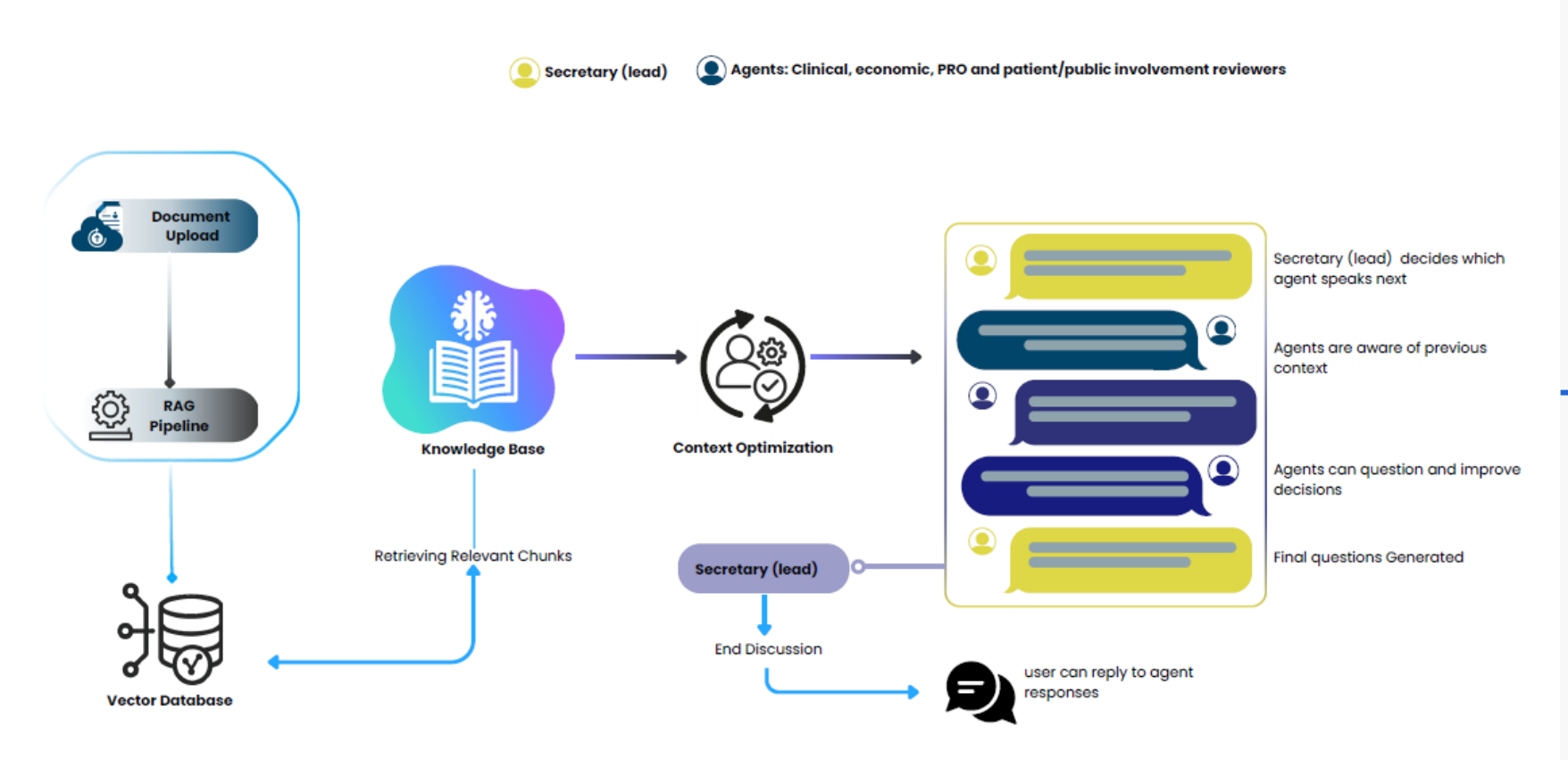
Evidence Pharmaco®

Reimagine the HTA assessments

Generative AI

How the conversation takes shape

- The lead sets the main objective guiding the agents' discussion. Each agent is given a distinct role and context



SME validated the outputs for *relevance*, *accuracy*, *traceability*, and *alignment* with HTA expectations

How the conversation takes shape

NICE 1L advanced/mHCC submissions were used for the simulation

- The multi-agent GenAI system successfully simulated interactions by generating structured clarification questions based on sponsor-submitted evidence for two submission
- A total of 104 questions were generated, including 29 clinical, 62 economic, and 13 general
- The majority of questions focused on data-related gaps, while a smaller proportion sought clarification on methodological aspects

Questions generated by the agents

ERG Economic Reviewer

Based on the provided information, here are some key clarification questions regarding how the clinical evidence influences the cost-effectiveness model for sorafenib in advanced hepatocellular carcinoma:

1. **Please clarify the rationale for using the lognormal distribution to extrapolate overall survival data beyond the SHARP trial period.**

Rationale: The choice of parametric distribution for extrapolation can significantly impact long-term survival estimates and cost-effectiveness results. While the lognormal distribution provided a slightly better statistical fit, the Appraisal Committee previously noted uncertainty around accepting it as definitive.

2. **Please provide a detailed comparison of the extrapolated overall survival curves (lognormal and Weibull) against the long-term observational data from the GIDEON study and UK real-world evidence (Palmer et al. 2013).**

Rationale: Validating extrapolations against long-term real-world data is crucial for assessing the plausibility of modeled survival estimates and their impact on incremental QALYs gained.

3. **Please clarify how post-progression treatments received in the SHARP trial were incorporated into the overall survival extrapolation and cost-effectiveness model.**

Rationale: Post-progression treatments can influence overall survival results and need to be appropriately

Questions generated by the agents

Evidence Pharmaco®

HTA Lead

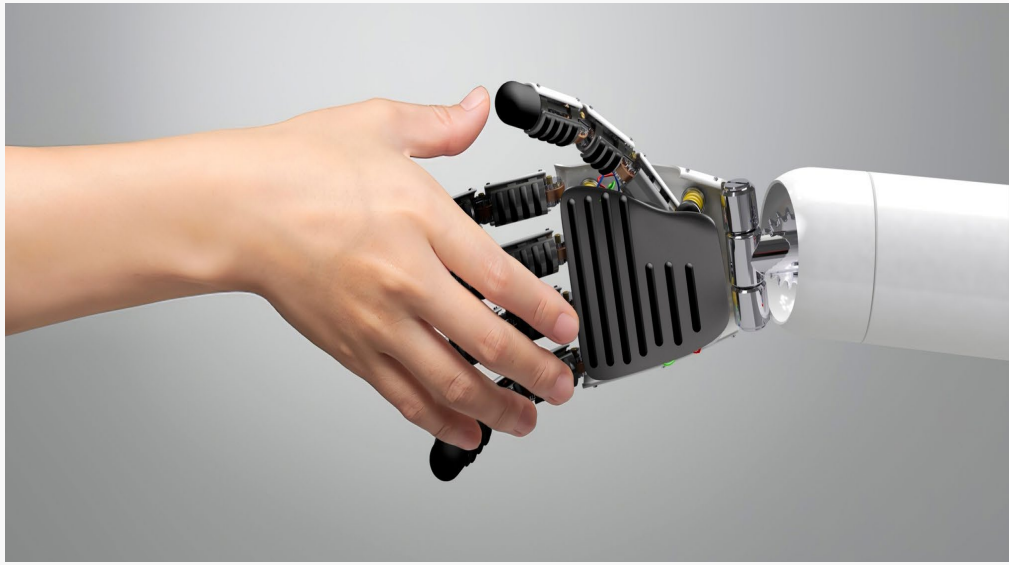
Based on the contributions from the ERG reviewers, I will now consolidate and finalize the clarification questions for Section A - Clarification on Effectiveness Data:

1. **Priority question:** Please provide full details of the systematic literature review methodology, including search strategies, databases searched, inclusion/exclusion criteria, and PRISMA flow diagram. **Rationale:** To assess the comprehensiveness and quality of the evidence base informing the submission.
2. **Priority question:** Please clarify the definitions and censoring rules used for all time-to-event endpoints in IMbrave150, including PFS, OS, and time to deterioration of QoL. **Rationale:** To understand how outcomes were measured and ensure consistency with NICE methodological standards.
3. **Priority question:** Please provide Kaplan-Meier plots and hazard ratios (with 95% CIs) for OS and PFS in IMbrave150 by key subgroups, including BCLC stage, presence/absence of macrovascular invasion or extrahepatic spread, and baseline AFP level. **Rationale:** To assess consistency of treatment effect across important prognostic subgroups.
4. **Priority question:** Please provide details of post-progression therapies received in each arm of IMbrave150, including proportions of patients and specific treatments. **Rationale:** To evaluate potential impact on OS and assess generalizability to UK practice.
5. **Priority question:** Please clarify how adverse events were defined, graded and attributed to study treatment in IMbrave150. Provide exposure-adjusted incidence rates for key adverse events. **Rationale:** To fully characterize the safety profile and enable appropriate modeling of adverse event impacts.
6. **Priority question:** Please provide full details of the network meta-analysis methodology, including assessment of heterogeneity, inconsistency, and model fit statistics. **Rationale:** To evaluate the robustness of the indirect treatment comparisons informing the economic model.
7. **Priority question:** Please provide the statistical code and output for the parametric survival modeling of OS, PFS and time to treatment discontinuation. **Rationale:** To enable verification of the extrapolations used in the economic model.
8. **Priority question:** Please clarify how health-related quality of life data were collected, analyzed and mapped to EQ-5D utilities, including handling of missing data. **Rationale:** To assess the appropriateness of the utility values used in the economic model.
9. **Priority question:** Please provide sensitivity analyses exploring alternative assumptions for extrapolation of OS, PFS and time on treatment beyond the trial period. **Rationale:** To characterize uncertainty in long-term outcomes informing the economic model.
10. **Priority question:** Please provide external validation of the extrapolated OS and PFS projections against real-world data sources or clinical expert opinion. **Rationale:** To assess plausibility of modeled long-term outcomes in UK clinical practice

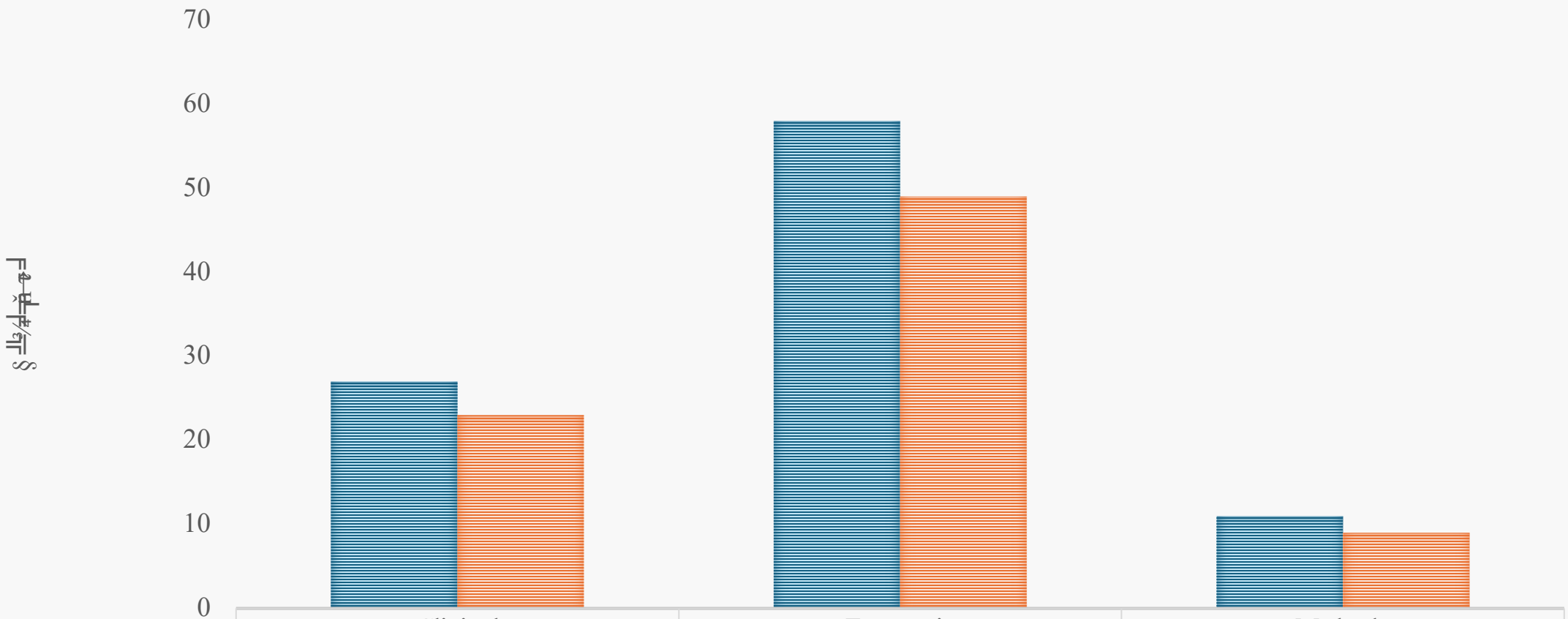
High level of concordance between AI-simulated and actual questions

- The AI-generated output was validated by an SME, demonstrating 80-85% agreement with actual clarification questions previously raised by NICE in the technology appraisals TA551 and TA666

- The remaining 15-20% of questions, although not identical to historical NICE examples, were still relevant and evidence-based



High level of concordance between AI-simulated and actual questions



	Clinical	Economic	Methods
Actual Questions	27	58	11
Gen AI Predicted	23	49	9

Key question missed by Gen AI

The EAG have sourced the relevant EMA public assessment report, but please clarify whether there is also a Medicines and Healthcare products Regulatory Agency (MHRA) Public Assessment Report. If yes, please provide a (confidential) copy

Please provide the extent to which each biomarker was elevated (mean/median/range) or changed due to each

Please clarify which of the following numbers are correct in the first SLR update. The Medline and Embase results are 484 on the final line Page 3 of 69 Clarification questions of the CS, Appendix D, Table 3 when limited to English language, but 494 is reported in the PRISMA flow diagram (CS, Appendix D, Figure 2). Medline in Process via PubMed is 178 in the CS, Appendix D Table 4, but 162 in the PRISMA flow diagram (Figure 2)

Please provide details of the number of patients in each group whose survival follow up was discontinued by the sponsor (company submission, page 20), and reasons why

Strengthened sponsor readiness and streamlined review

- The multi-agent GenAI framework offered a scalable approach to simulating NICE HTA review processes by generating structured clarification questions

- It enhances sponsor preparedness and may streamline future assessments, with planned extensions to incorporate additional analyses and real-world evidence

Thank you

For Questions or Clarifications please contact:

Barinder Singh, Director, Pharmacoevidence
barinder.singh@pharmacoevidence.com

Pharmaco Evidence