Advancing Survival Analysis Automation for HTA Submissions: A Multi-Agent Generative AI Framework

MSR41

Shubhram Pandey[1], Rashi Rani[2], Sameer Mansoori[1], Rajdeep Kaur[1], Marjana Bharali[1], Barinder Singh[1]
[1]Pharmacoevidence, Mohali, India, [2]Heorlytics, Mohali, India

Evidence Pharmaco®
Science-led HEOR & Tech-driven AI

The Status Quo

- Survival extrapolation underpins life-year and quality adjusted life years (QALY) estimation in cost-effectiveness models submitted to health technology assessment (HTA) bodies worldwide, making methodological choices directly consequential for reimbursement decisions[1]
- Yet the workflow remains a specialist bottleneck: selecting and fitting parametric and spline-based models, assessing goodness-of-fit against HTA guidelines, justifying extrapolation choices, and producing audit-ready outputs demands significant analytical time and expertise that limits throughput at scale
- No existing framework automates this process end-to-end while remaining concordant with published HTA guidance - leaving teams to rebuild bespoke pipelines for each submission
- Multi-agent generative artificial intelligence (GenAI) architectures where discrete, specialized agents handle modular tasks such as data preparation, model fitting, evidence retrieval, and report generation, represent a promising approach to address this gap by combining analytical rigor with improved reproducibility and scalability

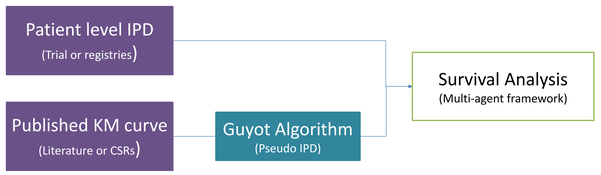
What We Set Out to Solve

- To **design and implement a multi-agent GenAI framework** in which specialized agents handle discrete analytical tasks across the survival modelling pipeline, with human-in-the-loop checkpoints at methodologically critical decision nodes
- To **automate end-to-end survival analysis workflows** encompassing data preparation, parametric and flexible model fitting, goodness-of-fit assessment, and long-term extrapolation, in alignment with **HTA guidelines (NICE DSU, CADTH)**
- To generate structured, reproducible, **HTA-submission-ready outputs** including model comparison summaries, extrapolation justifications, and survival estimates suitable for direct integration into health economic models

How We Solved It

- The framework accepts **patient-level time-to-event data** from trials or registries. Where only published Kaplan-Meier curves are available, **pseudo-individual patient data (IPD)** is reconstructed via the validated Guyot algorithm, removing dependency on proprietary datasets (**Figure 1**)

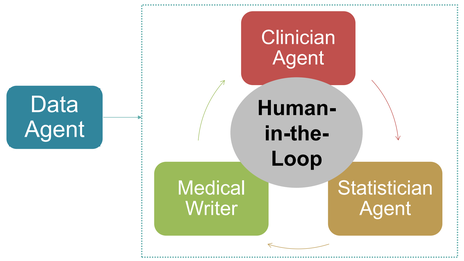
Figure 1: Data flow architecture



CSR: Clinical Study Reports; IPD: Individual Patient Data; KM: Kaplan–Meier

- Users specify analytical intent in **plain language**. Where instructions are incomplete, the framework identifies candidate model types from data characteristics - balancing automation with analyst oversight
- The framework uses a RAG pipeline to stay aligned with HTA guidance. At each analytical decision point, it retrieves relevant sections from NICE DSU technical support documents, CADTH guidance, and published survival analysis literature - grounding every model recommendation in current methodological standards
- Each agent operates on **domain-specific knowledge** with defined responsibilities: the **Clinician Agent** evaluates biological and clinical plausibility; the **Statistician Agent** drives model fitting, selection, and goodness-of-fit assessment against HTA benchmarks; the **Medical Writer Agent** produces structured, submission-ready outputs - functioning as an orchestrated pipeline with explicit handoffs (**Figure 2**)

Figure 2: Agent architecture



- Available models include **standard parametric distributions**, flexible **spline-based** models (odds, hazard, probit scales; up to 3 knots), **cure models** for long-term survivors, **piecewise** models for non-proportional hazards, and **parametric** mixture models for heterogeneous populations. Model selection is informed by AIC/BIC, visual fit assessment, and hazard scale inspection across all model types

Figure 3: List of survival models

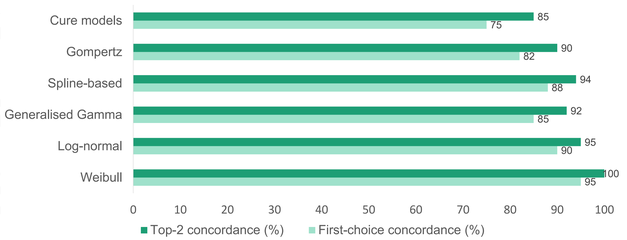
STANDARD PARAMETRIC MODEL	SPLINE BASED MODEL	OTHER MODELS
• Weibull • Log-normal • Log-logistic • Exponential • Gompertz • Gamma • Generalised Gamma	• Hazard scale • Odds scale • Probit scale • Up to 3 knots	• Cure models • Piecewise • Mixture models

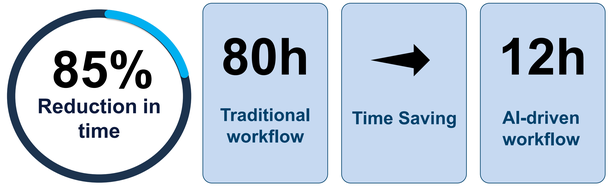
- A **closed-loop feedback mechanism** enables iterative model refinement - users incorporate clinical plausibility judgements and external validation data such as registry-based long-term estimates, ensuring final extrapolations are both statistically robust and clinically meaningful
- Every **analytical decision** - model selection rationale, goodness-of-fit assessment, extrapolation justification, and human review inputs is logged in a structured, version-controlled output, ensuring full reproducibility and a transparent audit trail suitable for HTA dossier submission

What We Got

- The framework was validated across **10 HTA case studies** spanning oncology (non-Small cell lung cancer, breast cancer, melanoma, chronic lymphocytic leukemia), rare disease (lysosomal storage disorder, Ph+ acute lymphoblastic leukemia), and cardiovascular indications, submitted to NICE, CADTH, and EU HTA bodies
- **First-choice model concordance** with expert-driven analyses was **87%; top-2 concordance exceeded 92%,** with discordance observed only in two cases involving non-proportional hazards where expert clinical judgement overrode statistical criteria
- On average, **6.4 survival model** families were evaluated per case study, with spline-based models selected in 4/10 cases, standard parametric in 5/10, and cure models in 1/10 - consistent with published HTA submission patterns
- End-to-end analysis time reduced from approximately 80 hours to 12 hours - an **85% reduction** - with the largest savings in model fitting and goodness-of-fit assessment (30h → 2h), data preparation (12h → 1.5h), and narrative report generation (20h → 5.5h)
- A first-draft HTA-ready survival report was generated within **35 minutes** of data upload across all 10 case studies, compared to an estimated 3-5 days for manual first drafts
- Human intervention was required for a median of 1.6 review cycles per case study before outputs were deemed submission-ready

Figure 4: Concordance by model family

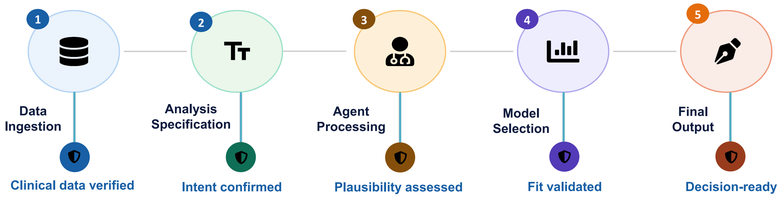




- All 10 case studies produced Excel-formatted survival outputs directly importable into health economic models without manual reformatting, **eliminating a typically 4–6-hour reformatting step**
- Automatically generated narrative justifications were rated "appropriate for HTA submission without revision" by clinical reviewers in **8/10 cases**; minor wording refinements were requested in **2/10 cases**
- Word reports included extrapolation rationale, goodness-of-fit commentary, and model selection justification aligned with NICE DSU TSD 14 reporting standards across **all case studies**

Interactive Dashboards | **Excel-formatted economic model inputs** | **Dynamically generated Word reports with interpretations**

Figure 5: Human-in-the-loop validation and clinical oversight framework



↻ HUMAN-IN-THE-LOOP OVERSIGHT - Expert review & external validation applied at each checkpoint

Iterative refinement
Model outputs reviewed after each stage. Expert clinical judgement incorporated to refine survival extrapolations before progression to the next step

External validation data
Registry-based long-term survival estimates, and published literature used to anchor AI-generated outputs to real-world clinical evidence

Decision-ready outputs
Only clinically plausible and methodologically robust models progress to final reporting ensuring suitability for NICE appraisals and HTA dossiers

- Framework extrapolations were reviewed against registry-based long-term survival estimates in **7/10 case studies** where external data were available; AI-generated extrapolations fell within clinically plausible ranges in **all 7 cases**
- The human-in-the-loop mechanism **successfully flagged one case** (rare disease, non-proportional hazards) where the statistically optimal model was clinically implausible, prompting expert override - demonstrating the system's ability to surface, not suppress, clinical judgement
- Final outputs across **all 10 case studies were deemed decision-ready**, with results suitable for direct use in NICE and CADTH dossier sections on survival analysis
- Validation was conducted against expert opinion as the **reference standard rather than observed long-term outcomes**; prospective validation against registry follow-up data remains an important next step before routine deployment in live HTA submissions

Why It Matters to You

- The multi-agent AI framework enables automated, accurate survival analysis with 87% concordance with expert model selection and maintained clinical validity across 10 HTA case studies - demonstrating both analytical reliability and submission readiness
- Critically, the framework is designed to augment rather than replace HTA expertise - human-in-the-loop checkpoints ensure that clinical judgement, contextual reasoning, and regulatory nuance remain central to every submission, with AI serving as a force multiplier for the analyst, not a substitute
- By reducing end-to-end analysis time by 85%, the framework significantly improves efficiency and scalability in HTA workflows, delivering consistent, audit-ready outputs suitable for direct integration into NICE, CADTH, and EU HTA dossiers
- Future development will incorporate Bayesian model averaging for uncertainty quantification, expand external validation across broader therapeutic areas, and enhance real-time quality assurance mechanisms to support wider HTA applicability

References
1. Clark, T., Bradburn, M., Love, S. et al. Survival Analysis Part I: Basic concepts and first analyses. Br J Cancer 89, 232–238 (2003)
2. Briggs A, Claxton K, Sculpher M. Decision Modelling for Health Economic Evaluation. Oxford University Press; 2006.
3. Latimer NR. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data. Med Decis Making)
4. NICE DSU Technical Support Document No. 14
5. Baio, G. (2020). survHE: Survival Analysis for Health Economic Evaluation and Cost-Effectiveness Modeling. Journal of Statistical Software

Correspondence:
Shubhram Pandey (Shubhram.Pandey@pharmacoevidence.com)

Disclosure:
SP, SM, RR, MB, RK and BS, the authors declare that they have no conflict of interest

ISPOR, May 17-20, 2026, Philadelphia, PA, United States