

# REPLICATING A PARTITIONED SURVIVAL MODEL FROM AN ICER REPORT USING GENERATIVE AI

HTA77

Turgay Ayer<sup>1,2</sup>, Sumeyye Samur<sup>1</sup>, Elif Bayraktar<sup>1</sup>, Tugce Ermis<sup>1</sup>, I. Fatih Yildirim<sup>1</sup>, Jagpreet Chhatwal<sup>1,3</sup>

<sup>1</sup> Value Analytics Labs, Boston, MA, USA, <sup>2</sup> Georgia Institute of Technology, Atlanta, Georgia, USA, <sup>3</sup> Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA,

## KEY FINDINGS

We demonstrate the feasibility of replicating published partitioned survival models using Generative-AI.

ValueGen.AI estimated costs and QALYs with error margins ranging from 22% to 50% and 0.4% to 2%, respectively.

Of note, incremental costs and QALYs were within 1.5% margin of the reported values and ICER was within 3.2% margin.

Future research should aim to understand which components of health economic model development can be accurately replicated by AI and where human oversight remains essential.

## BACKGROUND

- Generative AI has shown potential in automating complex tasks through advanced natural language processing; however, its integration into health economic modeling remains in the early stages of development.
- Building partitioned survival models (PSMs), with their requirements for precise parameter extraction, poses unique challenges due to their inherent complexity.

## OBJECTIVE

Our objective was to assess the feasibility and accuracy of using Generative AI to build a PSM, utilizing a published benchmark as a reference.

## METHODS

We replicated the PSM from the 2022 Institute for Clinical and Economic Review (ICER) Multiple-Myeloma Report<sup>1</sup> using ValueGen.AI, a customized Generative-AI platform.

**Automated Model Generation:** We uploaded the published report to ValueGen.AI<sup>2</sup> and prompted it to build a PSM for Ide-cel and its comparator (Triple- or Quad Refractor) using the report.

**Data Extraction:** We developed multi-agent pipelines for parameter extraction leveraging GPT-4o in combination with CrewAI<sup>3</sup>, LangChain<sup>4</sup>, and OpenAI<sup>5</sup> libraries in Python.

**Model Development and Run:** We developed a general PSM template using the Heemod<sup>6</sup> library in R and an API with the Plumber<sup>7</sup> library to automate PSM model construction and execution.

**Evaluation:** We compared the AI model outputs, including costs, QALYs and Incremental cost-effectiveness ratio (ICER) against the values in the ICER report to assess accuracy.

## RESULTS

- Generative AI successfully extracted PSM model structure (Figure 1), including the health states and the distributions with their parameters for estimating overall survival (OS) and progression-free survival (PFS) curves (Table 1).

Figure 1. AI-extracted Model Structure



## RESULTS (cont.)

Table 1. AI-extracted Distribution Parameters Matched Those in the ICER Report.

	Distribution	Parameter 1	Parameter 2
<b>Ide-cel</b>			
OS	Lognormal	3.24	0.93
PFS	Lognormal	2.31	1
<b>Comparator</b>			
OS	Lognormal	2.13	1.3
PFS	Lognormal	1.22	1

- Generative AI extracted all cost components effectively, including treatment, administration, monitoring, and adverse event costs for Ide-cel and its comparator. However, because the language model could not link each cost to specific health states, human oversight was required to review and confirm the costs associated with each health state.
- Generative AI extracted the utilities and linked them to the health states correctly (Table 2).

Table 2. AI-extracted Utilities Matched Those in the ICER Report.

Utility	Value
Progression-free on Therapy and Responding	0.78
Progression-free off Therapy and Responding	0.82
Progressed Disease/not Responding to Treatment	0.71
Death	0

- The AI-based PSM estimated the cost of Ide-cel at \$503,023 (compared to \$646,000) and the comparator at \$138,582 (compared to \$276,000). The discrepancy in costs can be attributed to Generative AI's inability to accurately distinguish treatment durations and apply the associated costs to the relevant cycles. Additionally, it estimated QALYs at 2.239 for Ide-cel (compared to 2.24 in the report) and 1.06 for the comparator (compared to 1.08) (Table 3).

Table 3. AI-extracted Distribution Parameters Matched Those in the ICER Report.

	AI-based PSM	ICER Report PSM
<b>Cost</b>		
Ide-cel	\$503,023	\$646,000
Comparator	\$138,582	\$276,000
<b>QALY</b>		
Ide-cel	2.239	2.24
Comparator	1.06	1.08

- The cost and QALY findings yielded an ICER of \$308,744 (compared to \$318,966 in the report). The AI-based PSM estimated the reported outcomes, with error margins of 1.5% for cost differences, 1.6% for QALY differences, and 3.2% for ICER (Table 4).

Table 4. Delta Values and Error Margins for AI-based PSM vs. ICER Report PSM.

	AI-based PSM	ICER Report PSM	Error Margin
Delta Cost	\$364,441	\$370,000	1.5%
Delta QALYs	1.179	1.16	1.6%
ICER for Ide-cel vs Comparator	\$308,744	\$318,966	3.2%

## REFERENCES

- Chimeric Antigen Receptor T-Cell Therapy for B-Cell Cancers: Effectiveness and Value, [https://icer.org/wp-content/uploads/2020/10/ICER\\_CAR\\_T\\_Final\\_Evidence\\_Report\\_032318.pdf](https://icer.org/wp-content/uploads/2020/10/ICER_CAR_T_Final_Evidence_Report_032318.pdf)
- ValueGen.AI, <https://valuegen.ai/>
- CrewAI, <https://github.com/crewAI/crewAI?form=MG0AV3>
- LangChain, [https://python.langchain.com/docs/how\\_to/qa\\_citations/?form=MG0AV3](https://python.langchain.com/docs/how_to/qa_citations/?form=MG0AV3)
- OpenAI, <https://github.com/openai/openai-dotnet?form=MG0AV3>
- Filipović-Pierucci, A., Zarka, K., & Durand-Zaleski, I. (2017). Markov Models for Health Economic Evaluation: The R Package heemod. ArXiv e-prints. R package version 0.8.0, 1702.03252