FULLY REPLICATING PUBLISHED MARKOV HEALTH ECONOMIC MODELS USING GENERATIVE AI

EE247

Jagpreet Chhatwal[1,2], Sumeyye Samur[2], I. Fatih Yildirim[2], Elif Bayraktar[2], Tugce Ermis[2], Turgay Ayer[2,3]

[1]Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, [2] Value Analytics Labs, Boston, MA, USA, [3]Georgia Institute of Technology, Atlanta, Georgia, USA

KEY FINDINGS

We demonstrate that generative AI can fully replicate – conceptualization, parameterization, and coding – simple health economic models with accuracy. This study serves as a basis for future research on fully replicating more complex health economic models.

BACKGROUND

- Generative AI holds significant promise for automating complex tasks, such as developing health economic models.
- Although its application in this field is still in the early stages, it has the potential to streamline model development by reducing the required time and expertise, offering substantial benefits to stakeholders.

OBJECTIVE

Our objective was to evaluate the feasibility and accuracy of Generative AI in fully replicating health economic models by utilizing a well-established benchmark.

METHODS

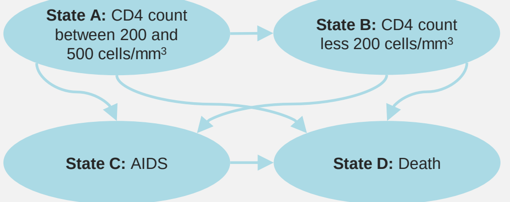
We replicated the HIV/AIDS Markov model from Chapter 2 of Andrew Briggs et al.'s *"Decision Modeling for Health Economic Evaluation".*[1]

- **Data Extraction :** Python was used for interactions with a large language model to extract model structure and parameter values from the chapter PDF.
- **Data Processing:** ValueGen.AI [2], a GPT-4-based platform utilizing multi-agent pipelines, including CrewAI[3], LangChain[4], and OpenAI[5] libraries, was used for parameter extraction.
- **Model Development and Runs:** The extracted data was implemented in the R Heemod [6] package to build and run the Markov model.
- **Evaluation:** Life years, costs, and ICER were calculated.
- **Comparison:** AI-generated model structure and outcomes were compared with those from Briggs et al.

RESULTS

- Briggs et al. reported cost and life year outcomes only for the monotherapy arm, along with the ICER for comparing monotherapy to combination therapy.
- Generative AI successfully extracted key model components, such as health states, transition probabilities, costs, and utilities (Figure 1, Tables 1-3).

Figure 1. *AI-extracted and Generated Model Schematic*



RESULTS (cont.)

Table 1. *AI-extracted Transition Probabilities matched those in Briggs et al.*

From	To	Probability
State A: CD4 count between 200 and 500 cells/mm^3	State A: CD4 count between 200 and 500 cells/mm^3	0.721
State A: CD4 count between 200 and 500 cells/mm^3	State B: CD4 count less 200 cells/mm^3	0.202
State A: CD4 count between 200 and 500 cells/mm^3	State C: AIDS	0.067
State A: CD4 count between 200 and 500 cells/mm^3	State D: Death	0.01
State B: CD4 count less 200 cells/mm^3	State B: CD4 count less 200 cells/mm^3	0.581
State B: CD4 count less 200 cells/mm^3	State C: AIDS	0.407
State B: CD4 count less 200 cells/mm^3	State D: Death	0.012
State C: AIDS	State C: AIDS	0.75
State C: AIDS	State D: Death	0.25
State D: Death	State D: Death	1.0

Table 2. *AI-extracted Costs matched those in Briggs et al.*

State	Cost
State A: CD4 count between 200 and 500 cells/mm^3	$2,756
State B: CD4 count less 200 cells/mm^3	$3,052
State C: AIDS	$9,007
State D: Death	0.0

Table 3. *AI-extracted Utilities matched those in Briggs et. al.*

State	Utility
State A: CD4 count between 200 and 500 cells/mm^3	0.99
State B: CD4 count less 200 cells/mm^3	0.964
State C: AIDS	0.911
State D: Death	0.0

- The AI-based model closely aligned with the reported outcomes, displaying an 8% error in costs, 0.1% in life-years, and 2% in ICER compared to Briggs et al. (Table 4).
- We repeated the experiments 20 times, and the error margins remained consistent.
- ValuGen.AI platform successfully generated the model code in R (Figure 2).

Table 4. *Generative AI-based model outcomes compared to Briggs et.al.*

	Life years for Monotherapy	Costs for Monotherapy	ICER for Monotherapy vs Comb. Therapy
AI-based Model	8.47	$48,400	$6,400
Briggs et.al.	8.45	$44,663	$6,276
Error Margin	0.1%	8%	2%

Figure 2: *Screen shot from the AI-generated Model Code in R*

REFERENCES

[1]Briggs, A., Sculpher, M., & Claxton, K. (2006). Decision Modelling for Health Economic Evaluation. Oxford University Press

[2]ValueGen.AI, https://valuegen.ai/

[3]CrewAI, https://github.com/crewAIInc/crewAI?form=MG0AV3

[4]OpenAI, https://github.com/openai/openai-dotnet?form=MG0AV3

[5]LangChain, https://python.langchain.com/docs/how_to/qa_citations/?form=MG0AV3

[6]Filipović-Pierucci, A., Zarca, 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

Contact: Jagpreet Chhatwal, PhD: jagchhatwal@mgh.harvard.edu