# FULLY REPLICATING PUBLISHED MARKOV HEALTH **ECONOMIC MODELS USING GENERATIVE AI**

Jagpreet Chhatwal<sup>1,2</sup>, Sumeyye Samur<sup>2</sup>, I. Fatih Yildirim<sup>2</sup>, Elif Bayraktar<sup>2</sup>, Tugce Ermis<sup>2</sup>, Turgay Ayer<sup>2,3</sup> <sup>1</sup>Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, <sup>2</sup> Value Analytics Labs, Boston, MA, USA, <sup>3</sup>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**

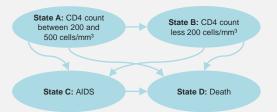
We replicated the HIV/AIDS Markov model from Chapter 2 of Andrew Briggs et al.'s "Decision Modeling for Health Economic Evaluation".<sup>1</sup>

- · 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.Al<sup>2</sup>, a GPT-4-based platform utilizing multi-agent pipelines, including CrewAI<sup>3</sup>, LangChain<sup>4</sup>, and OpenAI<sup>5</sup> libraries, was used for parameter extraction.
- · Model Development and Runs: The extracted data was implemented in the R Heemod <sup>6</sup> package to build and run the Markov model
- · Evaluation: Life years, costs, and ICER were calculated.
- · Comparison: Al-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. Al-extracted and Generated Model Schematic



# **RESULTS** (cont.)

able 1. Al-extracted Transition Probabilities matched those in Briggs et al.			
From	То	Probability	
State A: CD4 count between 200 and 500 cells/mm <sup>3</sup>	State A: CD4 count between 200 and 500 cells/mm <sup>3</sup> 0.7		
State A: CD4 count between 200 and 500 cells/mm <sup>3</sup>	State B: CD4 count less 200 cells/mm <sup>3</sup> 0.2		
State A: CD4 count between 200 and 500 cells/mm <sup>3</sup>	State C: AIDS	0.067	
State A: CD4 count between 200 and 500 cells/mm <sup>3</sup>	State D: Death	0.01	
State B: CD4 count less 200 cells/mm <sup>3</sup>	State B: CD4 count less 200 cells/mm <sup>3</sup>	0.581	
State B: CD4 count less 200 cells/mm <sup>3</sup>	State C: AIDS	0.407	
State B: CD4 count less 200 cells/mm <sup>3</sup>	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. Al-extracted Costs matched those in Briggs et al.

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

### Table 3. Al-extracted Utilities matched those in Briggs et. al.

State	Utility
State A: CD4 count between 200 and 500 cells/mm <sup>3</sup>	0.99
State B: CD4 count less 200 cells/mm <sup>3</sup>	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
Al-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

ars(ggslet2) ars((ldyobras	
wry Tsiatia-Yaki	
rery(jscelling)	
STATES A.	
	a court perimer callaber", "state #; eps court cost callaber", "state #; Arms", "State #; Beath")
shate_brands	secure se
CCOLFELL 0.20 Drug COSTS 5-	· · · · · · · · · · · · · · · · · · ·
	C(0, R, 0, R) + C(200, NHL2, NH07, R)
sittenet, rate	
	d <- "covenation_10"
num_agertia vi	1 minutes
state transit.	m.probabilities to an anti-information probabilities)
	midsolv(mag costs)
CFC144 ev an.	http://www.tcyclesp
restances of	langer(status)
trans_matrix	
	te_mrans(tion_probabilities.
	rum_states,
Barrow.	+ 74.50
	metric) to states
	marries a status
trong, autrix\$	Tate comes c- states
	controletion transition. trans matrix)
strategy_para	eters ( list(sat_mess) s.status)
	m states)
about the	older string as
	ofin_sisterout_brains - firewart(", fealth_costs[1],
	.001;cont_drags = discourt(", drag.contol1];
	<pre>.est.cost_total - cost_wealth - cost_stogs.life_pear - ". life_pears(1). ")"</pre>
the second second	r «- mulliphipheningatata enfinition.string);
strutegy.	and and the second se
	<pre>(strategy_baramaturs.llst(state_irpst))</pre>
	prosterio - ul'insultion, menu
strat according	de call'addae dramar, status presentari
ran maint	
ran, model	
	Come Aperts, reach, see states - 111.
san le	a martine, and the second
C047.	cost_total.
#1fm	s LLFs_years

<sup>1</sup>Briggs, A., Sculpher, M., & Claxton, K. (2006). Decision Modelling for Health Economic Evaluation. Oxford University Press REFERENCES

- <sup>2</sup>ValueGen.Al, https://valuegen.ai/
- 3CrewAl, https://github.com/crewAlInc/crewAl?form=MG0AV3
- <sup>4</sup>OpenAl, https://github.com/openai/openai-dotnet?form=MG0AV3
- <sup>5</sup>LangChain, https://pvthon.langchain.com/docs/how to/ga\_citations/?form=MG0AV3
- <sup>6</sup>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