Transforming the Assessment of Uncertainty in Health Economic Models Using Generative AI

William Rawlinson,[1] Elizabeth Vinand,[1] Oliver Pople,[1] Sarah Cudworth[1]

[1] Estima Scientific, London, United Kingdom

EE459

Introduction

- There are often many plausible model structures, assumptions, and data sources for a health economic model. The sensitivity of model results to these methodological choices can be described as 'modelling approach uncertainty'.
- However, systematically exploring this uncertainty is challenging. Implementing and evaluating alternative modelling approaches can be time-consuming and resource-intensive, particularly when changes to the underlying model engine are required. Consequently, only a limited subset of plausible approaches is typically assessed.
- Generative AI offers the potential to overcome these practical barriers, enabling more comprehensive and systematic exploration of modelling approach uncertainty.

Aim

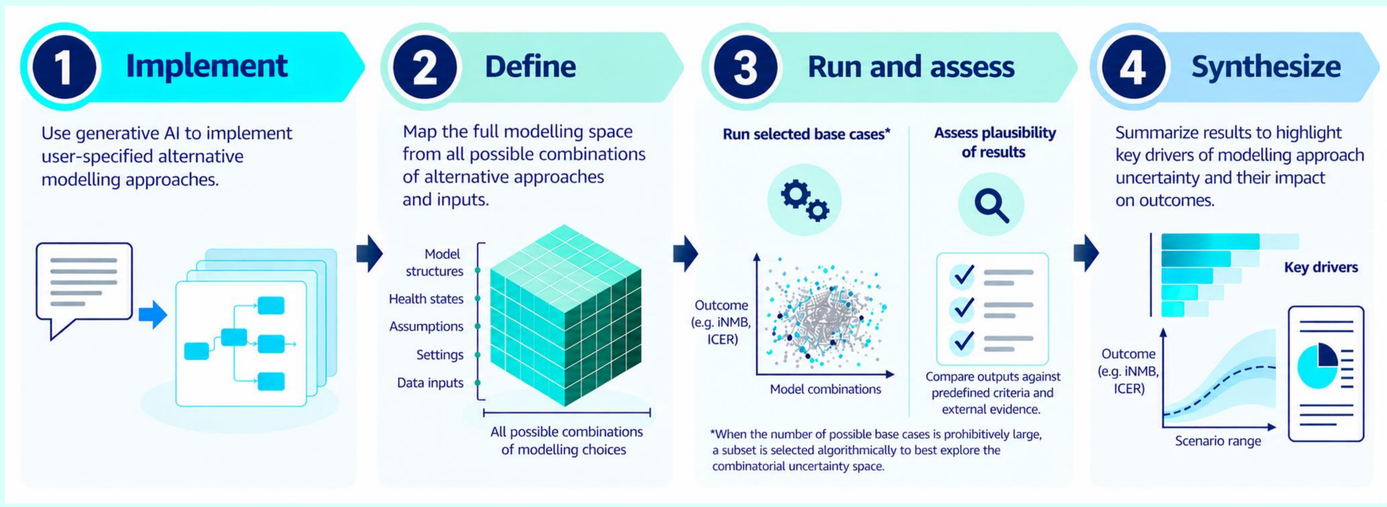
To evaluate the ability of economEST, a generative AI-based pipeline, to:

1. Automatically **implement** alternative modelling approaches specified by the user in natural language, including modifications to model structures, health states, assumptions, and settings.
2. **Define** the full modelling space resulting from all possible combinations of alternative model approaches and inputs.
3. **Run** the resulting base cases, using algorithmic selection of an informative subset where the number of combinations is prohibitively large, to identify key model drivers and explore best- and worst-case scenarios.
4. **Assess** the plausibility of resulting base-case outputs against predefined criteria and external evidence.
5. **Synthesize** findings to identify key drivers of modelling approach uncertainty.

Methods

- economEST was evaluated using two previously published health economic models.
- For each model, a health economist specified alternative model structures, assumptions, and settings in natural language, which economEST automatically implemented using generative AI. External data sources and other criteria were also provided to inform plausibility assessment.
- From the full set of possible base cases, generated by combining all user-specified modelling options, a subset was selected algorithmically to explore the range of possible outcomes and identify influential modelling choices, including effects arising from interactions between multiple modelling decisions.
- The accuracy of model implementations and plausibility assessments was evaluated through manual review by two health economists.
- The automated economEST workflow is illustrated in **Figure 1**.

Figure 1. Overview of the economEST automated pipeline for exploring modelling approach uncertainty.

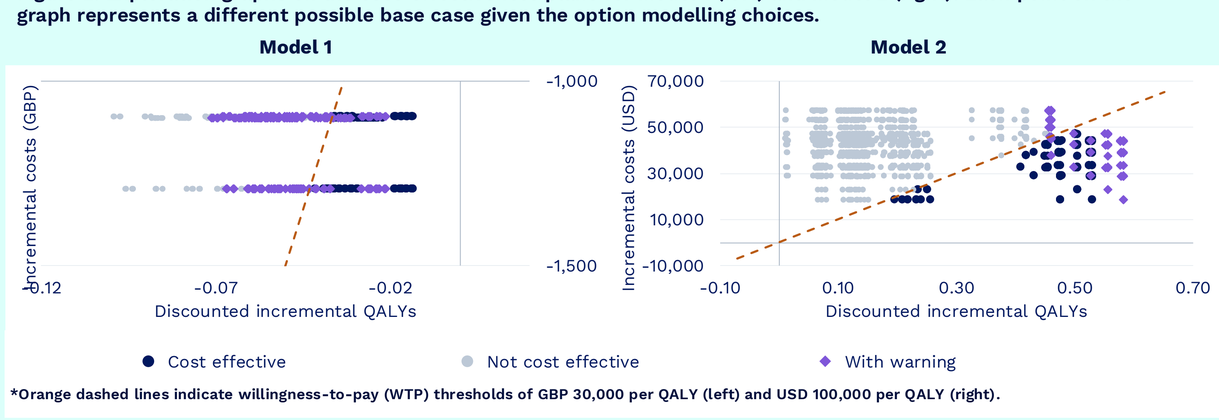


Results

Exploration of modelling approach uncertainty

- For both models, 7 modelling factors and a range of structural implementations were assessed, including alternative health state structures and model options.
- 1,500 base cases algorithmically selected and evaluated per model.
- **Model 1:** The different modelling options resulted in 2,916 possible base cases, all 1,500 evaluated were associated with lower costs and fewer QALYs (**Figure 2, left**). Of these, 231 (15.4%) generated sufficient cost savings per QALY forgone to meet the £30,000/QALY willingness to pay threshold, suggesting potential cost-effectiveness. Plausibility warnings were identified in 417 base cases (27.8%), primarily where utility following second-line treatment failure exceeded that following first-line treatment failure.
- **Model 2:** Of 6,048 possible base cases, 294/1500 (19.6%) of those evaluated were cost-effective at $100,000/QALY (**Figure 2, right**). Plausibility warnings were identified in 105 base cases (7.0%), where estimated QALYs exceeded a plausible threshold.

Figure 2. Exploration graphs on the cost-effectiveness plane for Model 1 (left) and Model 2 (right). Each point on the graph represents a different possible base case given the option modelling choices.



*Orange dashed lines indicate willingness-to-pay (WTP) thresholds of GBP 30,000 per QALY (left) and USD 100,000 per QALY (right).

Best- and worst-case scenarios

- Best- and worst-case scenarios were identified for each model exploration as the evaluated base cases with the highest and lowest incremental net monetary benefit (iNMB), respectively (**Figure 3**).
- **Model 1:** Six of seven modelling factors differed between scenarios, with ICERs of £11,019/QALY (worst case) and £95,173/QALY (best case), compared with a base-case ICER of £20,475/QALY [SW quadrant – less costly and less effective].
- **Model 2:** All seven modelling factors differed between scenarios, with ICERs of US$5,399,036/QALY (worst case) and US$31,954/QALY (best case), compared with a base-case ICER of US$83,857/QALY.

Figure 3. Comparison of modelling aspects in the best- and worst-case scenarios for Model 1 (left) and Model 2 (right).

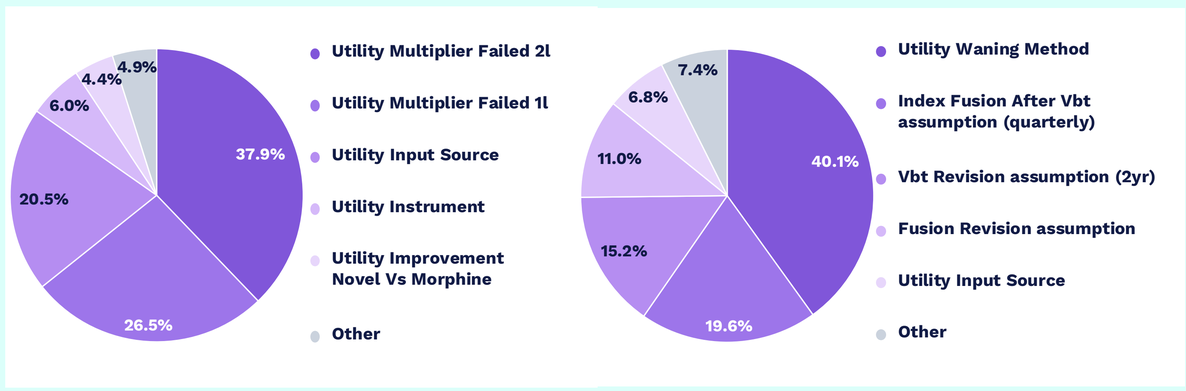
Aspect	Scenario comparison	
Model version	No titration period	Base model
Cycle Prob Discontinue After Failed 2L	0.00	0.15
Safety Profile Scenario	Base case	Decreased (-25%)
Utility Improvement Novel Vs Morphine	0.00	0.50
Utility Instrument	SF-6D	SF-6D
Utility Multiplier Failed 1L	1	0.70
Utility Multiplier Failed 2L	1	0.50
Utility Input Source	Pessimistic	Optimistic

Aspect	Scenario comparison	
Model version	Revision decrement for VBT	Alt health state structure
Fusion Revision assumption	0.20	NA (not available in this model version)
Index Fusion After Vbt assumption (quarterly)	No revision	0.01
Vbt Revision Assumption (2yr)	No revision	0.10
Time Horizon Years	10	5
Utility Input Source	Optimistic	base_case
Utility Waning Method	none	exponential
Utility Waning Period Years	10	5

Factor influence

- Factor influence was assessed across evaluated combinations of modelling choices, rather than relative to a single reference base case, giving a more comprehensive picture given the uncertainty around the chosen base case. Influence was defined as the difference between the highest and lowest mean incremental net monetary benefit (iNMB) across each factor's possible values over all base cases explored (**Figure 4**).
- **Model 1:** Utility multipliers following second- and first-line treatment failure were the most influential factors (37.9% and 26.5%, respectively), followed by the utility input source used (20.5%).
- **Model 2:** Utility waning method was the most influential factor (40.1%), followed by the quarterly index fusion after vertebral body tethering (VBT) assumption (19.6%) and the two-year VBT revision assumption (15.2%).

Figure 4. Relative influence of modelling aspects on iNMB for Model 1 (left) and Model 2 (right).



Pricing analysis

- Economically justifiable prices (EJPs) were calculated to assess the impact of modelling approach uncertainty on intervention pricing.
- Model 2 (case study): At a willingness-to-pay threshold of US$100,000/QALY, the EJP increased from US$58,763 in the original base case to US$90,206 in the best-case scenario (**Figure 5**).
- In the worst-case scenario, no non-negative EJP was identified, indicating that the intervention was not cost-effective even at a device price of US$0.
- The pricing analysis illustrates how alternative modelling approaches can substantially alter pricing headroom and cost-effectiveness conclusions.

Figure 5. Threshold prices and incremental results for best-, base- and worst-case scenarios.

Best case threshold price	Base case threshold price	Worst case threshold price
US$90,206	**US$58,763**	**No non-negative price**
At US$100,000 per QALY	At US$100,000 per QALY	At US$100,000 per QALY
iNMB: US$39,591	iNMB: US$8,148	iNMB: -US$52,360

Limitations

Although the case studies used previously published health economic models, hypothetical values were introduced to simulate alternative modelling scenarios. The study focused on demonstrating economEST's ability to systematically explore modelling approach uncertainty, rather than identifying the most clinically plausible assumptions or definitive cost-effectiveness estimates.

Conclusions

By overcoming resource constraints associated with traditional evaluation of alternative modelling approaches, generative AI may enable significantly more comprehensive and systematic assessment of modelling approach uncertainty. Unlike deterministic and probabilistic sensitivity analyses, which primarily assess parameter uncertainty, this study systematically explores combinations of alternative modelling approaches rather than isolated changes relative to a reference base case. This could provide decision-makers with a more robust understanding of the impact of modelling choices on estimates of intervention value.

References

1. Polly DW, Larson AN, Samdani AF, Rawlinson W, Brechka H, Porteous A, Marsh W, Ditto R. Cost-Utility Analysis of Anterior Vertebral Body Tethering versus Spinal Fusion in Idiopathic Scoliosis from a US Integrated Healthcare Delivery System Perspective. Clinicoecon Outcomes Res. 2021 Mar 15;13:175-190. doi: 10.2147/CEOR.S289459
2. Sullivan, W., Hirst, M., Beard, S. et al. Economic evaluation in chronic pain: a systematic review and de novo flexible economic model. Eur J Health Econ 17, 755–770 (2016). https://doi.org/10.1007/s10198-015-0720-y