Poster #: MSR64

Utility Analysis Methods: Can One Method Fit It All or Is a Case-by-Case Approach Required?

Angeliki Kaproulia,[1] Brittany Humphries,[2] Gabriel Tremblay,[2] Bart Heeg,[1] Andre Verhoek[1]
[1]Cytel, Rotterdam, Netherlands; [2]Cytel, Toronto, ON, Canada

Background

- Utility data scores are critical for the interpretation of clinical data and decision-making for resource allocation.
- Utility data have distributional properties (e.g., multimodal, bounded) that can make utility analysis statistically challenging. Improper statistical analysis may result in biased estimates in cost-effectiveness analyses.
- Many statistical methods have been used to analyze utility data in cost utility-analysis: the ordinary least squares (OLS), linear mixed effects, generalized linear (GL), Tobit, generalized estimating equations (GEE), adjusted limited dependent variable mixture (ALDVM), and repeated-measured mixed effects (RMME) models, among others.[1]
- Mixture models have been recently developed to address some of the distributional issues, such as the ALDVM model to address the multimodality of the utility distribution.[2] Mixed-effect models (e.g., linear mixed effects, RMME) are also expected to perform well in multimodal distributions, since their designs allow the prediction of latent sub-classes. The other regression approaches are expected to struggle in identifying those.
- Tobit, ALDVM,[2] and mixed effects models are anticipated to be able to handle bounded data, also known as ceiling effect. Based on the design of linear fixed effects models are not expected to predict the ceiling effect, since they do not have explicit cut-off points.

Objective

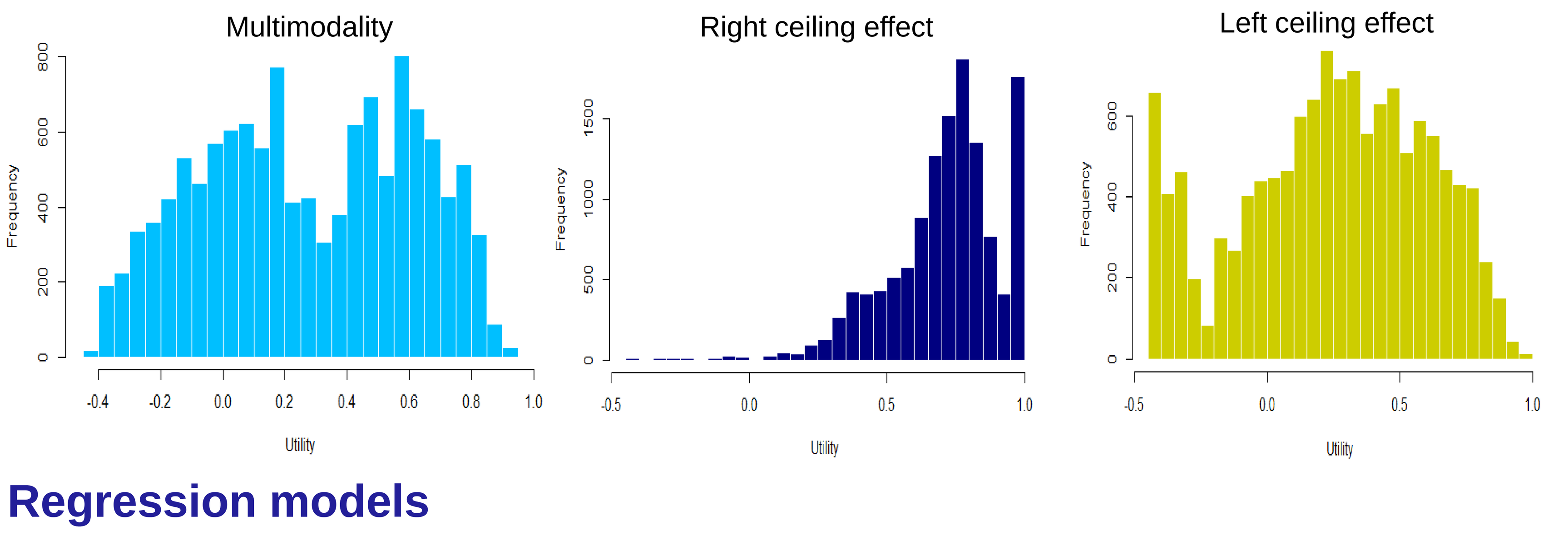
- The objective was to compare the performance of different predictive regression models in handling health utility data and to explore how these behave when the utility datasets show multimodality and (right and left) ceiling effect.

Methods

Dataset simulation

- Three utility datasets were simulated to resemble randomized controlled trials. The health utility data were generated by using the Dutch EQ-5D-5L scores.[3] The three datasets resembled: 1) multimodality, 2) right, and 3) left ceiling effect in utility distributions. Each dataset contained 500 patients, with an average of 25 responses per patient. The simulation accounted for some missing data over time, due to loss of follow-up.
- All three datasets included covariates: sex and treatment variables were categorical and distributed equally across the arms; adverse events was a variable independent of the treatment arm; and baseline utility was a continuous variable. Since mixture models can be used to identify latent classes within a distribution,[2] disease severity was included as a latent sub-class of three levels (mild/medium/severe).
- A probability was generated to sample the EQ-5D-5L answers, where the probability was set by a mean answer and influenced by each covariate, latent sub-class, their interactions, and a time effect. Using this probability, EQ-5D-5L answers were generated by sampling a binomial distribution.
- The multimodality dataset simulated a multimodal distribution of the utility data, mainly caused by disease severity, the effect of not receiving treatment, and the presence of adverse events. The right ceiling effect dataset simulated a right-bounded distribution in which many scores reached the utility upper limit (i.e., 1) because of treatment and non-severe disease type. The last dataset showed a left-bounded distribution influenced by the covariates of sex, treatment, adverse events, and non-mild disease (Figure 1).

Figure 1. Histograms showing the actual utility distribution in the multimodality, right, and left ceiling effect datasets



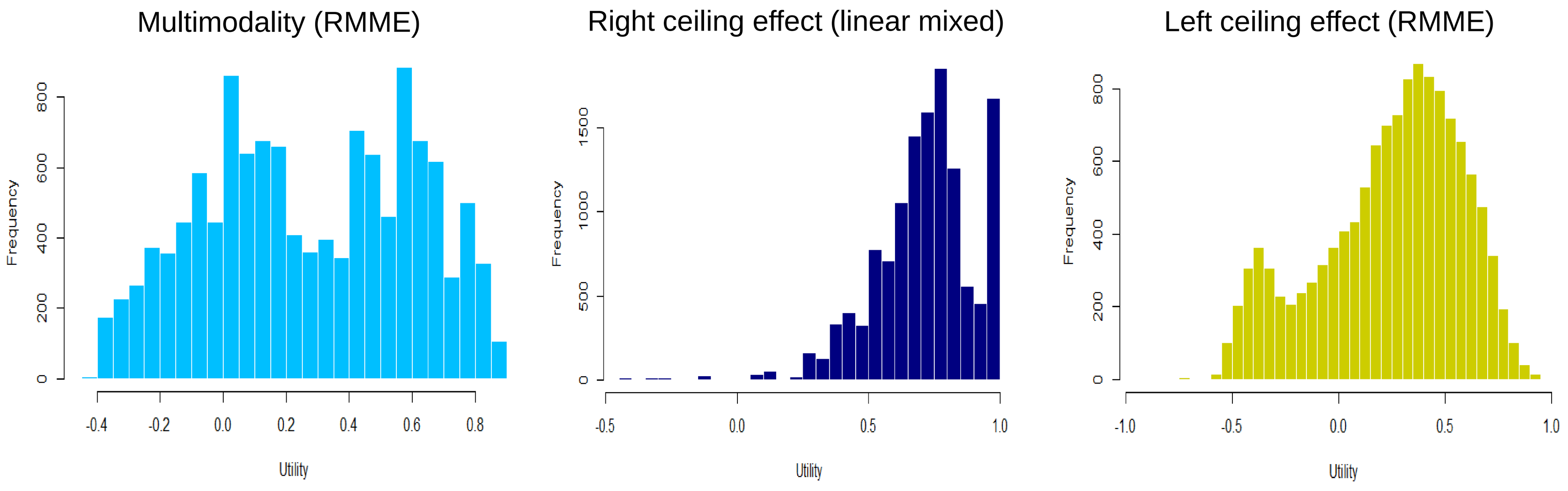
Regression models

- The performance of OLS, linear mixed effects GL, tobit, GEE, ALDVM, and RMME models was compared. The choice of regression models tested was based on a targeted literature review, part of this analysis, which showed that these were some of the most frequently used in utility analysis. Each regression model was tested in the simulated datasets and its performance was assessed based on the mean absolute error, root mean square error, and visual inspection. In addition, the true and predicted (via the regression models) mean utilities and confidence intervals were compared.

Results

- In the multimodality dataset, the RMME model achieved the best performance on all three criteria and based on the comparison of the true and predicted mean utility (Figure 2; Table 1). The ALDVM model was able to identify the latent sub-classes but returned very high errors and showed poor visual fit (Table 1), likely by overfitting latent classes.
- In the right ceiling effect dataset, the linear mixed effects model was the best-performing model, re-creating a good ceiling effect (Figure 2). The RMME model also performed well (Table 1). The Tobit model had a satisfactory performance in terms of visual fit, although it returned higher errors than expected; linear fixed effects models performed poorly (Table 1).
- In the left ceiling effect dataset, all models struggled to predict the strong cut-off, with the RMME still performing reasonably (Figure 2; Table 1). Same as in the ceiling effect dataset, the linear fixed effects models underperformed, as anticipated (Table 1).

Figure 2. Histograms showing the predicted utility distribution in the datasets: multimodality (RMME), right (linear mixed) and left ceiling effect (RMME)



Abbreviations: RMME, repeated measured mixed effects

Table 1. Regression models performance criteria and mean (true and predicted) utility values

Criteria	OLS*	LM	GL*	Tobit	GEE*	ALDVM	RMME
Multimodality							
MAE	0.190	0.044	0.190	0.190	0.190	0.203	**0.042**
RMSE	0.247	0.072	0.247	0.247	0.248	0.261	**0.067**
Right ceiling effect							
MAE	0.075	**0.044**	0.075	0.073	0.075	0.164	0.047
RMSE	0.104	**0.072**	0.104	0.105	0.104	0.217	0.072
Left ceiling effect							
MAE	0.155	0.093	0.155	0.155	0.155	0.188	**0.073**
RMSE	0.198	0.136	0.198	0.198	0.198	0.230	**0.105**
Predicted mean utility (95% CI)							
	OLS*	**LM**	**GL***	**Tobit**	**GEE***	**ALDVM**	**RMME**
Multimodality (true mean utility [95% CI]: 0.274 [-0.344, 0.817])							
	0.274 (-0.179, 0.659)	0.274 (-0.325, 0.812)	0.274 (-0.179, 0.659)	0.274 (-0.179, 0.659)	0.274 (-0.178, 0.631)	0.281 (-0.043, 0.575)	**0.274 (-0.328, 0.815)**
Right ceiling effect (true mean utility [95% CI]: 0.710 [0.243, 1.000])							
	0.710 (0.303, 0.979)	**0.710 (0.291, 0.997)**	0.710 (0.303, 0.979)	0.719 (0.303, 1.000)	0.710 (0.303, 0.979)	0.617 (0.268, 0.918)	0.709 (0.295, 1.000)
Left ceiling effect (true mean utility [95% CI]: 0.243 [-0.446, 0.826])							
	0.243 (-0.328, 0.749)	0.243 (-0.440, 0.725)	0.243 (-0.328, 0.749)	0.243 (-0.328, 0.749)	0.243 (-0.340, 0.737)	0.195 (-0.238, 0.624)	**0.243 (-0.439, 0.815)**

* Based on this table, the results across the models were the same; however, at a higher level of precision, differences were found.
Abbreviations: ALDVM, adjusted-limited dependent variable mixture; CI, confidence interval; GEE, generalized estimating equations; GL, generalized linear; LM, linear mixed; MAE, mean absolute error; OLS, ordinary least squares; RMME, repeated measured mixed effects; RMSE, root mean square error

Conclusions

- This analysis evaluated the performance of different regression approaches in simulated utility datasets. The mixed effects design of the linear mixed and RMME models, accounting for both fixed and random effects, led to their superior performances across the different datasets, in terms of errors, visual fit, and precise reflection of the true utility data. Contrary to expectations, the ALDVM underperformed and was not able to deal with multimodality.

Disclosures

This study was investigator-initiated and received no funding.

References

1. Wen J. (2020). Mapping for the Eq-5D-5L for Use in Cost-Utility Analysis (thesis); 2. Hernández Alava M, Wailoo AJ, Ara, R. (2012). Tails from the Peak District: Adjusted limited dependent variable mixture models of EQ-5D Questionnaire Health State Utility Values. Value in Health, 15(3), 550–561;. 3. Versteegh MM, Vermeulen KM, Evers SMAA, de Wit, GA, Prenger R, Stolk EA. (2016). Dutch tariff for the five-level version of EQ-5D. Value in Health, 19(4), 343–352.