Descriptives vs Statistics in Utility Analyses: A Simulation

Angeliki Kaproulia, Andre Verhoek*, Bart Heeg
Cytel, Rotterdam, Netherlands

Background

Guidance from the National Institute of Health and Care Excellence (NICE), Canada's Drug Agency (CDA) and Zorginstituut Nederland (ZIN) on how to appropriately conduct utility analyses is limited.[1,2,3]

A frequently used model for Health Technology Assessment (HTA) submissions is the repeated-measured mixed effects (RMME).[4] Despite its theoretical limitations, like inability to handle ceiling effects and multimodality, previous analyses showed that RMME often outperformed other approaches.[4,5]

Its strength lies in accounting for subject-specific effects across repeated observations.[6] This characteristic may become a weakness in the presence of latent variables (e.g., disease severity).

RMME is expected to perform poorly in cases where patients have low baseline utility, often due to severe disease. These patients may experience disease progression, but since they already have low baseline utility, the decline in post-progression utility may be minimal, making it difficult for RMME to accurately capture the impact of progression on utility.

Objective

The objective of this study was to compare RMME with descriptive statistics, in a simulated dataset where low baseline utility, reflecting severe disease as latent variable, is predictive of early progression and minimal utility decline post-progression.

Methods

Data simulation

A dataset was simulated to resemble a randomised controlled trial in early-stage cancer, to generate utility scores. A descriptive summary of the EQ-5D utilities was included in the analysis.

The statistical relationship between utilities was assessed using regression analysis. To account for the repeated measurements in the study, the RMME model was used.

Sampling individual patient/observation utilities

The simulated dataset included the following features: 2000 patients with early-stage cancer (large sample to avoid noise), utility observations received monthly for approximately 1.5 years, progression, death, and treatment. The Dutch-specific dataset for EQ-5D-5L was used to generate health utility data.[7]

The data set included three co-variates: treatment, progression, and disease severity (latent variable in utility analysis, i.e., indirectly observed). To determine treatment status and severity, sampling from a normal distribution was used.

All patients started in a relatively healthy EQ-5D state (22222), with a utility base of 0.705 (as per Dutch tariffs).[7] Baseline utility was sampled from a normal distribution using score increment/decrement operators associated with population base utility, treatment, severity and progression effect. The utility score increment/decrement operators were determined as presented in Table 1.

Table 1. Utility scores for the different combinations of the effect of co-variates

Components	Utility score
Population utility	0.705
Population utility + disease severity	0.635
Population utility + treatment	0.761
Population utility + progression	0.612

Patients were assigned to a baseline utility drawing numbers from a normal distribution. Severe patients had lower baseline utility than non-severe. Progression-free (PFS) and overall survival (OS) were defined using numbers from Weibull and Binomial distributions, accordingly.

These were simulated to be correlated with disease severity and treatment and to be representative of early-stage cancer (Figure 1; Figure 2). Patients with severe disease (and low baseline utility) progressed earlier. Patient utility observations over time were simulated, accounting for progression and treatment effect.

In total, around 31,000 utility observations were simulated. The simulated dataset included 52% treated patients, 29% with disease severity and 24% with progression. Approximately 2,500 utility observations in post-progression were included in the analysis, 66% of which were of severe patients. Severe patients had lower drop in utility because they were already at a worse health state, with lower baseline utility, than non-severe patients.

Figure 1. Progression-free survival Kaplan-Meier

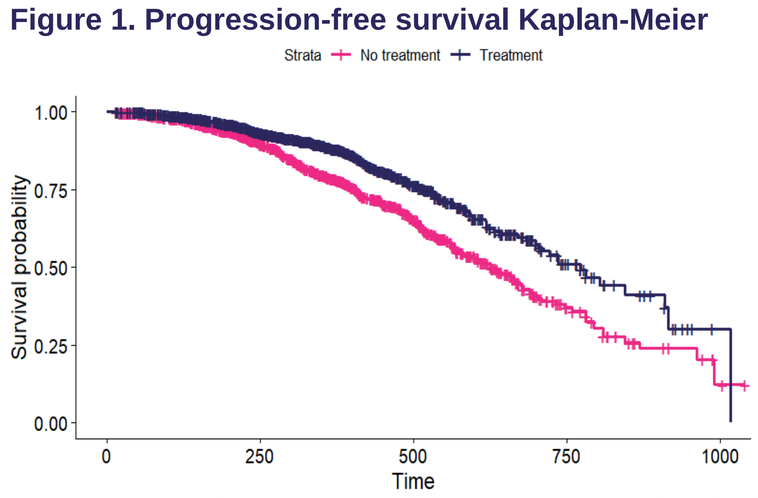
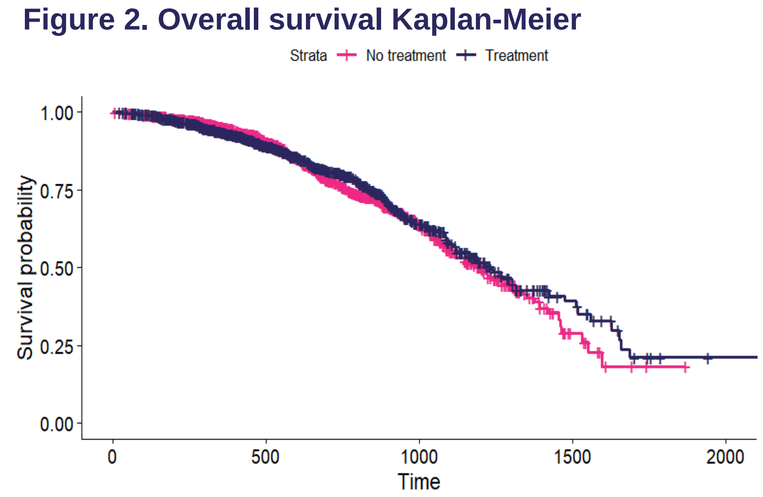


Figure 2. Overall survival Kaplan-Meier



In scenario analysis, a second dataset was simulated, similar to the base case one. The only difference was that the drop in utility after progression was set to be independent of severity, i.e., no matter the baseline utility, the drop in utility due to progression, would be similar for severe and non-severe patients.

Results

The true, simulated, utility values were compared against the predicted ones from the regression analysis, using RMME and descriptive statistics, in pre- and post-progression, stratified by treatment group (Figure 3).

In the base case, pre-progression, the true values aligned with the mean RMME utilities and descriptive statistics, across both treatment groups.

Post-progression, the true utility dropped by 0.074 (*0.714-0.640*) and 0.088 (*0.683-0.595*) from pre-progression, for treated and non-treated patients, respectively.

Descriptive statistics captured 100% [mean = 0.637; 95% Confidence intervals (CIs) (0.633, 6.641)] and 88% [mean = 0.606 95% CIs (0.602, 0.609)] of the true drop, for patients with and without treatment, respectively.

RMME estimated a smaller decrease, 71% (mean 0.656; 95% CIs (0.644, 0.669) and 56% (mean = 0.637; 95% CIs (0.624, 0.696) of the true drop for treated and non-treated, respectively.

In the scenario analysis, true utilities closely matched with descriptive statistics and RMME results, before and after progression across both treatment arms (Figure 4).

Figure 3. Utility estimates across the different analysis methods (Base-case)

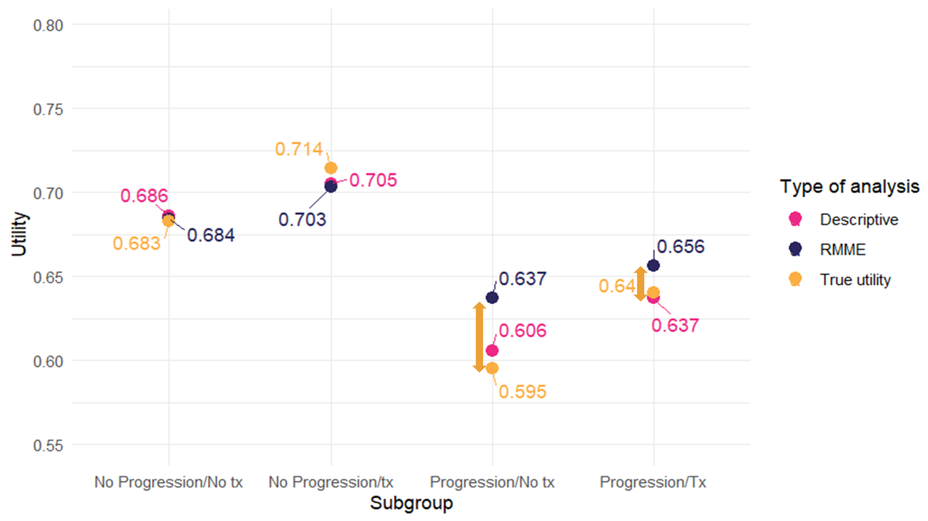
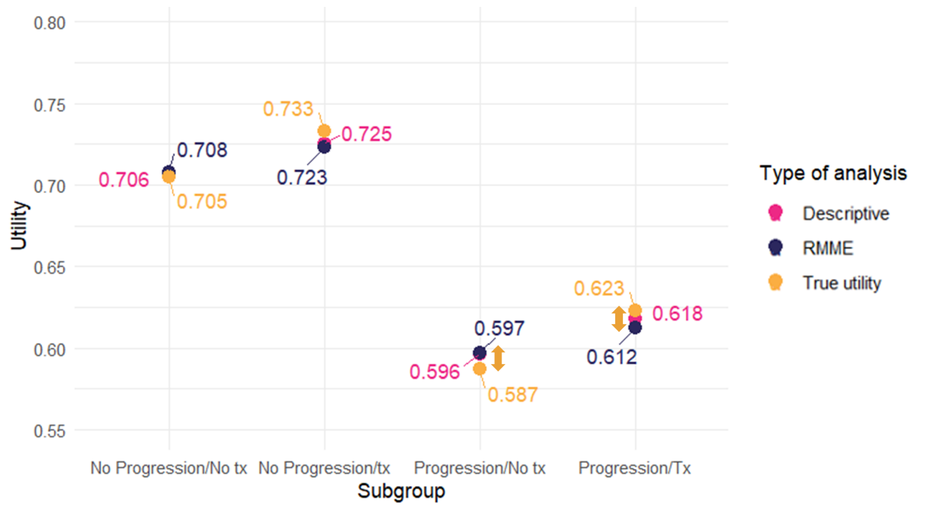


Figure 4. Utility estimates across the different analysis methods (Scenario)



Abbreviations: RMME, repeated-measured mixed effects; tx, treatment

Conclusions

RMME is commonly used,[4] shows advantages over other regression methods in certain aspects,[5] and is accepted by HTA bodies, but is not refrained from bias. This analysis showed that in case low baseline utility is prognostic for progression and low utility drop in post-progression, RMME might return biased results in post-progression utility, especially if data is immature.

Numerous statistical methods are available for analyzing utilities, which have pros and cons depending on the underlying dataset. HTA guidance is needed to provide a framework for the appropriate use of statistical models, considering the underlying data.

In the meantime, descriptive statistics should not be overlooked, as thorough descriptive analyses can help inform the selection of the appropriate statistical method.

References

1. Wailoo, A., Hernandez Alava, M., Pudney, S. NICE DSU Technical Support Document 22 Mapping to estimate health state utilities. 2023; 2. Guidelines for the economic evaluation of health technologies: Canada. 4th ed. Ottawa: CADTH; 2017 Mar; 3. Zorginstituut Nederland. Jan van Busschbach. Verdiepingsmodule QALY en kwaliteitvan-leven-metingen. Herziene versie 2024; 4. Humphries, B et al. Safe to Assume? Unpacking Regression Assumptions for Utility Data Value in Health, Volume 25, Issue 12, S376. 2022; 5. Kaproulia, A et al. Can One Method Fit It All or Is a Case-By-Case Approach Required?. Value in Health, Volume 25, Issue 12, S362. 2022; 6. Detry MA, Ma Y. Analyzing Repeated Measurements Using Mixed Models. JAMA. 2016 Jan 26;315(4):407-8. doi: 10.1001/jama.2015.19394; 7. M Versteegh M, M Vermeulen K, M A A Evers S, de Wit GA, Prenger R, A Stolk E. Dutch Tariff for the Five-Level Version of EQ-5D. Value Health. 2016 Jun;19(4):343-52. doi: 10.1016/j.jval.2016.01.003.

Disclosures

This study was investigator-initiated and received no funding.

* Employed at Cytel during the development of the abstract for this research.