

How important is parameter uncertainty around the UK EQ-5D-3L value set when estimating treatment effects?

Alastair Gray¹, Oliver Rivero-Arias^{1,3}, Jose Leal¹, Helen Dakin¹, Juan Manuel Ramos-Goñi^{2,3}

¹ Health Economics Research Centre, Department of Public Health, University of Oxford, UK

² Servicio de Evaluación y Planificación del Servicio Canario de Salud, Spain

³ CIBER Epidemiología y Salud Pública (CIBERESP), Spain

Email for correspondence: oliver.rivero@dph.ox.ac.uk

Abstract

Aims: The uncertainty around the EQ-5D-3L value set is commonly ignored in economic evaluation. This study evaluates the impact of including parameter uncertainty around the original UK EQ-5D-3L value set (or “tariff”) within standard errors as well as sampling uncertainty around the trial population.

Methods: First, we re-estimated the N3 model of the EQ-5D-3L value set with original data from the Measurement and Valuation of Health (MVH) study to replicate the published coefficients. Second, we estimated standard errors around the predicted utility of each EQ-5D-3L health state to evaluate the impact of parameter uncertainty on these estimated utilities. Third, we used a two-stage bootstrap approach to combine the resulting MVH parameter uncertainty with trial sampling uncertainty for a large randomised trial population. In the first step, we generated 1,000 sets of coefficients for the UK EQ-5D-3L tariff using bootstrap resampling from the original MVH sample. In the second step, we used bootstrap resampling from the clinical trial data 10,000 times for each of the 1,000 sets of EQ-5D tariff coefficients. The standard error including MVH parameter uncertainty was then estimated as the standard deviation across the resulting vector of results from the 10 million bootstrap replicates. This figure was compared against a one-stage bootstrap from the clinical trial sample to assess the impact of including parameter uncertainty.

Data: The EQ-5D N3 model was estimated using the original MVH data. The randomised control trial used as a case study was the International Subarachnoid Aneurysm Trial (ISAT), a large clinical trial comparing endovascular coiling and neurosurgery for the treatment of ruptured aneurysms. EQ-5D-3L data to calculate mean utilities at 2 and 12 months’ follow-up and the quality-adjusted life years (QALYs) accrued over the trial period were available for 1633 patients.

Findings: Including MVH parameter uncertainty around the original EQ-5D N3 model increased the standard errors around mean between-group differences in utility only very slightly

from 0.01542 (ignoring parameter uncertainty) to 0.01550 at 2 months follow-up, from 0.01490 to 0.01493 at 12 months follow-up and from 0.01149 to 0.01155 for QALYs. Sensitivity analysis suggests parameter uncertainty has a larger impact when the N3 model is re-estimated using much smaller valuation samples: estimates based on a 1% sub-sample of MVH respondents increases the variance by 5.60% at 2 months follow-up, 4.87% at 12 months follow-up and 4.59% for QALYs. The standard errors around with the original time trade-off valuations in the MVH study suggests a substantial amount of between respondent variation affecting most health states, which the original N3 model did not fully account for. Therefore, the standard errors around N3 model predictions are likely to be biased, and do not reflect the true uncertainty around the valuations.

Conclusions/implications: Our results suggest that parameter uncertainty around the EQ-5D value set has little impact on the estimated confidence intervals around utility differences in this example. This result is driven by the precision of the original estimated coefficients, given the large dataset used in the original MVH study. However, parameter uncertainty may have more influence on value sets estimated using different estimation methods or smaller valuation samples. Our results suggest that other types of uncertainty around health state valuations, such as model uncertainty, may have more important implications than parameter uncertainty: further research is required to clarify this.

Introduction

Since its development in the early 1990s, the EuroQol EQ-5D-3L has been used as a generic quality of life questionnaire in a wide range of disease areas. In addition, it has become the preferred quality of life instrument in the UK for health economists involved in the economic evaluation of health care technologies, influenced partly by current guidelines from the National Institute for Health and Clinical Excellence (NICE),¹ which describe the EQ-5D-3L as NICE's preferred method for measuring and valuing health-related quality of life (HRQL) in adults. The EQ-5D-3L permits the respondent to select from 243 possible health states, which can then be converted by means of a value set or "tariff" into a utility value intended to represent the preferences of a population sample for each state. Researchers have spent considerable time estimating value sets in different countries: a review in 2007 identified 8 complete value sets based on time trade-off (TTO) methods, and a further 9 using the visual analogue scale.² These studies have generally followed the methodology used to estimate the first value set in the UK, also known as the N3 model.³ The N3 model provides a simple equation with a set of coefficients that can easily be implemented in any spreadsheet or statistical software to predict utility for each of the 243 health states. The N3 model used data from the Measurement and Valuation of Health (MVH) study, a large exercise to collect preferences from the UK general population of a subset of health states from the EQ-5D-3L.⁴ The coefficients from the model were originally estimated using regression methods and therefore are subject to uncertainty, which is currently ignored when estimating utility values from any value set.

This type of parameter uncertainty is well-known to health economists, who have proposed methods to incorporate parameter uncertainty into economic models using probabilistic sensitivity analysis.⁵ Recently, a report from a NICE DSU⁶ clearly suggests that uncertainty around preference-based valuation weights such as those estimated from the N3 models should be incorporated in the probabilistic sensitivity analysis in a decision model-based context. When patient-level EQ-5D-3L data are available a similar scenario arises. Utility values are usually obtained across a sample of patients using the N3 model algorithm with a fixed set of coefficients. The MVH exercise has received some criticism in recent years, in particular the analytical approach followed in the original exercise.⁷ However, the impact of propagating parameter uncertainty from the original N3 model to the estimation of, for example, treatment effects in trial-based studies, has received little attention in the literature (although there has been recent interest in this issue in the context of mapping algorithms to predict utility from another quality of life instrument).^{8,9}

This paper explores the impact of incorporating parameter uncertainty from the N3 model coefficients into the standard errors and confidence intervals of mean differences in quality of life estimates using real trial data. We start from the premise that the true mean utility estimated using EQ-5D-3L is subject to three sources of uncertainty:¹⁰

- **Trial sampling variation:** uncertainty around the mean utility for the population in whom EQ-5D tariff is to be applied, which arises from not having sampled an infinitely large trial sample. This uncertainty is lower when the EQ-5D tariff is applied to larger, more homogeneous trial populations.
- **Parameter uncertainty:** uncertainty around the coefficients estimated in the N3 model. This uncertainty should reflect the imprecision in the valuations of different EQ-5D health states in the original valuation sample (e.g. in the MVH study). We call this “MVH parameter uncertainty” hereafter.
- **Model uncertainty:** uncertainty around the specification of the model used to estimate the N3 model.

We explore the first two sources of uncertainty and use a non-parametric bootstrapping method to account for both of these types of uncertainty simultaneously to estimate treatment effects in quality of life estimates in a randomised controlled trial. The paper is divided into four sections starting with this introduction. The methods section provides a description of the datasets used, the re-estimation of the N3 model using the MVH dataset and a step by step guide to the non-parametric approach used to propagate parameter uncertainty. Results are then presented, followed by discussion and closing comments.

Methods

The EQ-5D-3L questionnaire

The EQ-5D-3L questionnaire is a generic quality of life instrument widely used as a health outcome in the economic analysis of healthcare technologies in many different disease areas.¹¹ It contains two sections: a self-description part and a self-rating visual analogue scale or “thermometer”. The self-description section contains one question for each of five domains (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), with three levels (labelled 1 to 3) in each domain indicating no problems, some problems or extreme problem/unable. As a result the self-description can produce 243 (3^5) different possible health states. For example, the health state ‘21232’ represents some problems walking, no problems

washing or dressing, some problems performing usual activities, extreme pain or discomfort and moderate anxiety or depression.

Each of these descriptive health states from the EQ-5D-3L can be translated into a utilityⁱ using a tariffⁱⁱ (described in detailed in the following section) where 0 indicates “dead” and 1 indicates “full health”. Full health corresponds to the health state 11111 and negative values are possible, indicating health states worse than dead.

MVH dataset

In the UK valuation exercise, valuation of EQ-5D-3L health states was performed as part of the MVH study. The study was conducted in the early 1990s to obtain a set of values that represented preferences from the UK general population for all 243 theoretically possible health states within the EQ-5D-3L questionnaire.⁴ Forty-three health states were directly valued by 3,395 UK adults using TTO. During face-to-face interviews, each participant rank-ordered 12 health states plus full health, unconscious and “immediate death” from best to worst and identified which states he/she considered to be “worse than dead”. Different TTO approaches were used for health states “better than dead” and “worse than dead”. For health states better than dead, the respondent was asked to select a length of time (x) in the 11111 state considered equivalent to 10 years in the evaluated health state. In this case, the utility value was calculated as $x/10$. For health states worse than dead, the respondent was given the choice between dying immediately or spending a period of time (x) in the evaluated health state followed by (10-x) in the 11111 state. The utility value for these health states were calculated as $(x/10)-1$.ⁱⁱⁱ The transformation for both types of health states ensure that utilities are bounded between a value of 1 (full health) and -1. The full dataset from the study can be obtained from the UK Data Archive [<http://www.esds.ac.uk/findingData/snDescription.asp?sn=3444>].

Case study

The International Subarachnoid Aneurysm Trial (ISAT) was used to study the impact of parameter uncertainty around the EQ-5D value set. This large multinational randomised controlled trial evaluated the effectiveness and cost-effectiveness of endovascular coiling compared to neurosurgery for the treatment of aneurysmal subarachnoid haemorrhage.^{12 13} Detailed information on resource use and EQ-5D-3L quality of life data were collected at 2 months, 12 months and annually thereafter, although only data from 2 and 12 months were used here. Quality-adjusted life years (QALYs) were used as the health outcome measure in the

ⁱ The utility value may also be called the index, health state valuation or quality weight

ⁱⁱⁱ The tariff may also be called the “value set” or “algorithm”

ⁱⁱⁱ For health states worse than dead, the MVH dataset gives (10-x); x therefore needs to be calculated from this figure to obtain the utility.

economic evaluation and calculated as the average of utilities at 2 and 12 months, multiplied by 0.833 (10 /12 months) years. Our sample comprised 1,633 patients from the UK with complete data at 2 and 12 months follow-up, of which 843 were in the endovascular arm and 790 in the neurosurgical arm.

Estimating the ‘tariff’ using the MVH dataset

We replicated the original N3 model using the MVH dataset to obtain highly precise tariffs, as well as the standard errors and the variance-covariance matrix. From an original sample of 3,395 respondents, 398 were excluded on the grounds of incompleteness or other criteria (e.g. all states valued the same), giving a total for analysis of 2,997 respondents.³ However, after careful reading of the MVH and modelling documentation available,^{3,4,14} it was not possible to reach exactly the same sample size as the authors. Therefore, we contacted the lead authors of a recent report on the use of health state utility values in decision models, who had recently estimated the variance-covariance matrix of the N3 model.⁶ The authors kindly provided an index variable that identified the final 2,997 patients included in the original estimation. As a result, the sample used in this study to re-estimate the N3 model comprised the 2,997 respondents included in the original N3 paper.³ The health state “unconscious” was not included in the model estimation as it is not part of the EQ-5D-3L.

The original N3 model was estimated using a random intercept model that allows the respondents’ responses to vary randomly around an overall mean across all respondents. In total, 35,964 valuations were available for the 42 health states elicited in the MVH study and included in the model:

$$\begin{aligned}
 Y_{ij} = & \alpha + \beta_1 MO2_{ij} + \beta_2 MO3_{ij} + \beta_3 SC2_{ij} + \beta_4 SC3_{ij} + \beta_5 UA2_{ij} \\
 & + \beta_6 UA3_{ij} + \beta_7 PA2_{ij} + \beta_8 PA3_{ij} + \beta_9 AD2_{ij} + \beta_{10} AD3_{ij} \\
 & + \beta_{11} N3_{ij} + e_{ij} + u_i
 \end{aligned} \tag{1}$$

for $i=1, \dots, 2997$ and $j=1, \dots, 12$

Where $Y_{ij} = 1 - s_{ij}$, and s_{ij} corresponds to value given by respondent i to health state j using the TTO method. The respondent-specific random effect, u_i , measures the difference between the average value for individual i and the average for the entire population of respondents. The observation specific error term, e_{ij} , measures the deviation of the j th valuation from the average for the i th respondent. Table 1 lists the explanatory variables used in the model.

The N3 model (equation [1]) was re-estimated from the original data, providing standard errors for each coefficient and a covariance matrix. The re-estimated N3 model was then used to

calculate mean utility values and respective standard errors for each of the 42 health states used in the valuation exercise. These were compared with the estimates published by Dolan, and with the mean values and associated standard errors yielded directly by the time trade-off exercise in the MVH study. The comparison ensured that the replication was accurate, and provided useful information on the performance of the N3 model against the original data.

The published N3 model was acknowledged to suffer from some mis-specification as a result of omitted variables and/or incorrect functional form.³ However, our main objective here was to explore the consequences of incorporating parameter uncertainty in the existing model rather than to propose a different model specification. Nonetheless, we report robust standard errors to explore how different they are from the conventional standard errors reported in the original analysis.

Methods to incorporate EQ-5D-3L MVH parameter uncertainty in clinical trials

Parameter uncertainty around the MVH value set tends to be ignored in economic evaluations, which typically calculate utilities for each patient based on the point estimates of utility for each health state. As a result, any measures of uncertainty produced in these evaluations, such as standard errors around the mean utility calculated by either parametric or non-parametric methods, include only the trial sampling variation, not parameter uncertainty. We explored several alternative methods to incorporate MVH parameter uncertainty as well as trial sampling variation into estimates of the standard error around the mean for a patient population and the difference between group means.

We initially explored analytical methods to combine MVH parameter uncertainty and trial sampling variation but these required assumptions that may not hold. Furthermore, it was unclear whether it was feasible to estimate standard errors analytically while allowing for the correlations between MVH coefficients and the explanatory variables in the trial sample.

We therefore used the following two-stage bootstrap procedure, illustrated in Figure 1, to combine MVH parameter uncertainty with trial sampling variation:

1. The first stage involved generating 1,000 sets of EQ-5D tariff coefficients by bootstrapping from the MVH sample and re-estimating the N3 model for each bootstrap sample. This produced a 12 x 1,000 matrix with 1,000 sampled mean values for each of the 12 tariff coefficients (Table 1), providing an empirical distribution of the mean values of each coefficient. An alternative (and equivalent) method is to use Monte Carlo simulation to sample values randomly from the Cholesky decomposition of the variance-covariance matrix for the EQ-5D tariff of coefficients. Although the Cholesky method would be easier to

implement in other applications, since it requires only the published variance-covariance matrix for the EQ-5D tariff⁶ rather than the entire MVH sample, we adopted bootstrapping to avoid making parametric assumptions (i.e. multivariate normality).

2. In the second stage, the sets of EQ-5D tariff coefficients obtained in the first stage were applied sequentially to bootstrap samples drawn from ISAT:
 - a. The first set of tariff coefficient values was drawn from the 12 x 1,000 matrix generated in Stage 1.
 - b. A bootstrap sample was drawn (with replacement) from each arm of the ISAT dataset.
 - c. The utility for each patient in the bootstrap sample was calculated using the coefficient values drawn in step (a).
 - d. The mean utility for each arm at 2 months and 1 year for this bootstrap sample were recorded.
 - e. Steps (b)-(d) were repeated 10,000 times using the same set of mean values drawn in step (a).
 - f. Steps (a)-(e) were repeated for the other 999 sets of tariff coefficients saved in Stage 1.
3. Results were analysed across all 10,000 x 1,000 (10 million) replicates.

We drew 10,000 bootstraps from the ISAT dataset for each of 1,000 sets of coefficients to minimise Monte Carlo error (MCE) and ensure that the resulting statistics (mean and standard error) had sufficient precision for our purposes. MCE represents random noise between bootstrap samples and can be measured as the standard error of the mean statistic of interest across the bootstraps.¹⁵ A commonly used rule of thumb is to choose a number of bootstraps such that MCE is less than 5% of the statistic of interest.¹⁶ We estimated the MCE of the standard error of the mean utility differences at 2 months (ignoring tariff uncertainty)^{iv} using several bootstrap sample sizes and compared it with preliminary analyses that incorporated tariff uncertainty. Since preliminary analyses suggested that including MVH parameter uncertainty increased standard errors of the utility difference by around 0.0001, we therefore required an

^{iv} MCE was estimated by adding a second level of replication. Given X number of bootstrap estimates of mean utility difference at two months, U_{2mo} , we generated a bootstrap replicate of U_{2mo} (with replacement), which we call U_{2mo}^* , and estimated the mean across the replicate, $E(U_{2mo}^*)$. We repeated this process 1000 times to obtain $E(U_{2mo,1}^*), \dots, E(U_{2mo,1000}^*)$. The MCE was estimated as the standard deviation across the bootstrap statistics.

MCE $<5 \times 10^{-6}$ to accurately estimate the impact of tariff uncertainty, which we estimated (using the approach suggested by Koehler et al¹⁵) would require about 6 million bootstraps of the ISAT sample^v. Since incorporating tariff uncertainty may increase MCE, we based our results on a total of 10 million bootstrap replicates from ISAT (10,000 from each of 1,000 coefficient sets).

However, due to the computation time required, 2.5 million replicates were used for sensitivity analyses (2,500 bootstrap replicates drawn from ISAT for each of 1,000 sets of tariff coefficients).

The overall standard error was based on the standard deviation across all 10 million bootstrap replicates, and the proportion of uncertainty that is due to the tariff was quantified as [total variance-variance within each coefficient set]/total variance. No bias adjustment was made to the bootstrap results to simplify the analysis.

Like many economic variables, ISAT utilities were not normally distributed, with around 27% of patients having perfect health and 16% having utilities of 0.5 or less. As result, the standard error around the difference in mean utility ignoring tariff uncertainty differed slightly depending on whether the standard error was calculated using bootstrapping with 50,000 simulations (QALY difference: 0.0235; SE: 0.01149) or using parametric methods, such as a t-test (QALY difference: 0.0235; SE: 0.01155).

As previously recognised,¹⁷ the variance from bootstrapping tends to be $(n-1)/n$ times smaller than that estimated using the central limit theorem, but there is no clear indication of which method (or which estimate of the standard error) is correct. We therefore compared the standard errors obtained using the two-stage bootstrap including parameter uncertainty with those obtained using a one-stage bootstrap (rather than results of a t-test) to avoid any difference in standard errors between bootstrapping and parametric methods biasing the comparison between variances with and without tariff uncertainty.

Sensitivity analysis

We would expect the effect of incorporating parameter uncertainty to be larger for value sets estimated using smaller samples. We therefore assessed the impact of including tariff uncertainty on the standard error of mean utility differences for various sub-samples drawn at random from the full MVH sample and re-estimated the N3 model with 60%, 30%, 5% and 1% of the MVH study sample.

^v Briefly, we estimated the MCE for different bootstrap sizes X (1000, 5000, 10000 and 50000), without incorporating parameter uncertainty, and regressed these on $1/\sqrt{X}$ to obtain the slope $\hat{\beta}$ while omitting the intercept. The predicted number of bootstraps, X^* , to achieve MCE= 5×10^{-6} (target MCE) was determined as $X^* = \left(\frac{\hat{\beta}}{\text{target MCE}} \right)^2$.

Results

Table 1 shows the published N3 model coefficients alongside our re-estimated coefficients and standard errors. All coefficients match to the 3rd or 4th decimal place. The table also shows that the bootstrapped standard errors closely match the parametric results.

Uncertainty around valuations for individual health states

Table 2 presents the time trade-off valuations from the original MVH exercise for the 42 health states that were valued, and the utilities predicted by the N3 tariff. The absolute and squared differences between observed and predicted valuations were large for some health states, e.g. 21312, 23313 and 13322, but both the mean absolute difference (0.0043) and the mean square difference (0.0023) averaged across all health states are small, indicating that large differences in some health states are largely diluted by accurate predictions from other health states.

For most health states, valuations vary substantially between respondents, with standard deviations greater than predicted means in most cases. This is clear in Figure 1, which indicates the distributions of TTO utility valuations for selected health states. Individual responses are relatively homogenous for some very mild health states (e.g. 21111, 1211, 12111), although variability of individual responses and uncertainty around mean valuations is much greater for moderate and severe health states. Furthermore, the standard errors around predictions from the N3 model are much lower than those around the direct valuations for all but the very mildest and most severe health states. This indicates that the N3 model underestimates the observed uncertainty associated with nearly all health states valued in the MVH exercise. As Table 2 indicates, robust standard errors are larger than conventional standard errors in 26 of the 42 health states, but remain smaller than standard errors around the original time-trade-off valuations in all cases.

Impact of parameter uncertainty from EQ-5D-3L value set in trial-based economic evaluation

Table 3 presents the results of the quality of life estimates of patients in the ISAT study estimated using the EQ-5D-3L value set with and without parameter uncertainty. Ignoring parameter uncertainty, the difference in utility at two months is 0.0363 (95% non-parametric confidence interval, CI: 0.0060 to 0.0665). The difference is no longer significant at 12 months but the mean QALY difference over months 2 to 12 shows a significant difference of 0.0235 with associated confidence interval of 0.0010 to 0.0460. Allowing for parameter uncertainty, the standard error around the mean difference increases from 0.0154 to 0.0155 at two months follow-up, from 0.01490 to 0.01493 at 12 months follow-up and from 0.0115 to 0.0116 for QALYs. In all cases,

the 95% CI of these mean differences is slightly wider than the CIs without parameter uncertainty. Similarly, a small increase in the standard errors of mean estimates in each treatment arm is observed when compared with standard errors ignoring parameter uncertainty.

Table 4 presents the results of the sensitivity analysis that evaluates the effect of varying the sample size of the original valuation study, achieved by replicating bootstrapping on the N3 model using different proportions of respondents from the original MVH sample, ranging from 60% to 1%. When using 60% of the MVH sample, 1,798 participants and 21,576 valuations were included in the estimation. This figure decreases to 30 participants and 360 valuations when the N3 model is estimated using 1% of the MVH sample. As expected, the impact of including tariff uncertainty (and the percentage increase in the variance from accounting for this) increases as the size of the sample used in the valuation study decreases, due to the increased uncertainty around the N3 estimates. When valuations are based on the full MVH sample, the variance around the difference at 2 months increases by just 0.55% when parameter uncertainty is taken into account, but this variance increases by 5.60% when parameter uncertainty is included and when just 1% of the MVH sample is used to estimate the parameters.

Discussion

Analysts undertaking economic evaluations have become used to the idea that uncertainty should be fully represented when reporting results. In decision analytic modelling this usually includes a recommendation to undertake probabilistic sensitivity analysis to address uncertainty in all important model parameters.^{1 18} However, although trial-based economic evaluations using EQ-5D utilities to measure health outcomes use a set of externally estimated parameter values – the value set with which EQ-5D-3L responses are converted to utilities – the uncertainty around these parameters is typically ignored. In this paper we address the question of whether this matters. We re-estimate the original N3 model and also the uncertainty around the parameters, and then propagate that uncertainty through the quality of life data generated by a fairly typical trial to see what effect this has.

Our general conclusion is that the effect is small, and at least in the example used would not alter any conclusions drawn from the trial results. It is possible that the impact of such value set parameter uncertainty could be more noticeable if the QALY difference was very small, but typically such parameter uncertainty is likely to be small compared with the sampling uncertainty for the trial.

This finding is partly attributable to the large size of the original MVH study, with almost 3,000 respondents and almost 36,000 health state valuations. We demonstrate that the effect of value set parameter uncertainty does become more pronounced as the estimation sample is reduced. It is possible that such uncertainty is more important where the value set is based on smaller sample sizes: the Dutch EQ-5D-3L value set, for example, is based on 298 respondents,^{19,20} the German EQ-5D-3L value set is based on 339 respondents,²¹ and the value set for the HUI3 is based on 256 participants valuing 22-24 states using VAS valuations and 4 states using standard gamble.²² The analyses presented here may help researchers to choose sample sizes for future valuation exercises.

The question addressed here was fairly tightly constrained to the issue of parameter uncertainty in the *existing* N3 model. This is a reasonable question as it is precisely that value set which is in effect recommended at present by NICE in the UK.¹ However, it is clear that even the original MVH study could have been analysed in different ways giving potentially quite different results. For example, Craig and colleagues have demonstrated that transforming the worse than dead states in a different way to that chosen by Dolan can result in substantially different utility values for many health states, adding 0.25 to all states originally valued at less than 0.5.⁷ Similarly, Shaw and colleagues used a different transformation method for worse than dead states (in addition to some other differences such as sample weights) when estimating a US value set, again giving rise to some marked differences.²³ Dolan has himself suggested an alternative model for estimating a tariff for the EQ-5D, based on differences in valuations between worst and other states.²⁴ The results of these studies suggest that model specification is likely to be a much more important source of uncertainty than parameter uncertainty. Such model uncertainty could of course be considerably widened to encompass different instruments and elicitation methods, and may shortly become a more pressing issue for analysts when value sets for the EQ-5D-5L begin to appear. We conclude that ways of incorporating value set parameter uncertainty in economic evaluations, and the implications of quality of life differences arising from value set model uncertainty, merit closer attention.

Acknowledgements

We are indebted to the International Subarachnoid Aneurysm Trial (ISAT) Collaborative Group for providing the data for this methodological work. We are grateful to Alan Wailoo for access to the index variable that identified the 2,997 included in the N3 model. HERC is partly funded by the NIHR.

References

1. Nice. *Guide to the methods of technology appraisal*. London: National Institute for Clinical Excellence, 2008.
2. Szende A, Oppe M, Devlin NJ. *EQ-5D value sets: inventory, comparative review and user guide*. Dordrecht ; [London]: Springer, 2007.
3. Dolan P. Modeling valuations for EuroQol health states. *Medical Care* 1997;35(11):1095-108.
4. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. *Health Econ*. 1996;5(2):141-54.
5. Briggs AH, Claxton K, Sculpher MJ. *Decision modelling for health economic evaluation*. Oxford: Oxford University Press, 2006.
6. Ara RW, A. The Use of Health State Utility Values in Decision Models. *NICE DSU Technical Support Document*. Sheffield, 2011.
7. Craig BM, Busschbach JJ. Toward a more universal approach in health valuation. *Health Econ* 2011;20(7):864-75.
8. Siani C, de Peretti C, Castelli C, Phung T, Daures J, Duru G. Uncertainty around the Mean Utility Assessment Accounting for Mapping Extrapolation: Application to Prostate Cancer. *Collège des Économistes de la Santé*, 2010.
9. Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Med Decis Making* 2006;26(1):18-29.
10. Drummond MF, McGuire A. *Economic evaluation in health care : merging theory with practice*. Oxford ; New York: Oxford University Press, 2001.
11. EuroQol G. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16:199-208.
12. Molyneux A, Kerr R, Stratton I, Sandercock P, Clarke M, Shrimpton J, et al. International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised trial. *Lancet* 2002;360(9342):1267-74.
13. Molyneux AJ, Kerr RS, Yu LM, Clarke M, Sneade M, Yarnold JA, et al. International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion. *Lancet* 2005;366(9488):809-17.
14. Williams A. *The Measurement and Valuation of Health: a Chronicle*. York, 1995.
15. Koehler E, Brown E, Haneuse SJPA. On the Assessment of Monte Carlo Error in Simulation-Based Statistical Analyses. *Am Stat* 2009;63(2):155-62.
16. Efron B. Better Bootstrap Confidence-Intervals. *Journal of the American Statistical Association* 1987;82(397):171-85.
17. Nixon RM, Wonderling D, Grieve RD. Non-Parametric Methods for Cost-Effectiveness Analysis: The Central Limit Theorem and the Bootstrap Compared. *Health Economics* 2010;19(3):316-33.
18. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment - A review and consolidation of quality assessment. *Pharmacoeconomics* 2006;24(4):355-71.

19. Lamers LM, Stalmeier PFM, McDonnell J, Krabbe PFM, van Busschbach JJ. Measuring the quality of life in cost-utility analyses: the Dutch EQ-5D tariff. *Ned Tijdschr Genees* 2005;149(28):1574-78.
20. Lamers LM, McDonnell J, Stalmeier PFM, Krabbe PFM, Busschbach JJV. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Economics* 2006;15(10):1121-32.
21. Greiner W, Claes C, Busschbach JJ, von der Schulenburg JM. Validating the EQ-5D with time trade off for the German population. *Eur J Health Econ* 2005;6(2):124-30.
22. Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Medical Care* 2002;40(2):113-28.
23. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical Care* 2005;43(3):203-20.
24. Dolan P, Roberts J. Modelling valuations for Eq-5d health states: an alternative model using differences in valuations. *Med Care* 2002;40(5):442-46.

Table 1: Explanatory variables in N3 model, published model coefficients, and re-estimated coefficients and standard errors

Variable	Definition	N3	Our estimation		Empirical	
		published model coefficients	of the N3 model		distribution using bootstrapping	
		Mean value	Mean value	SE	Mean value	SE
MO2	1 if mobility is level 2; 0 otherwise	0.069	0.0686	0.0051	0.0687	0.0051
MO3	1 if mobility is level 3; 0 otherwise	0.314	0.3133	0.0065	0.3133	0.0068
SC2	1 if self-care is level 2; 0 otherwise	0.104	0.1035	0.0054	0.1035	0.0052
SC3	1 if self-care is level 3; 0 otherwise	0.214	0.2133	0.0066	0.2132	0.0068
UA2	1 if usual activities is level 2; 0 otherwise	0.036	0.0360	0.0062	0.0361	0.0057
UA3	1 if usual activities is level 3; 0 otherwise	0.094	0.0944	0.0073	0.0947	0.0075
PA2	1 if pain/discomfort is level 2; 0 otherwise	0.123	0.1226	0.0051	0.1227	0.0049
PA3	1 if pain/discomfort is level 3; 0 otherwise	0.386	0.3847	0.0057	0.3847	0.0061
AD2	1 if anxiety/depression is level 2; 0 otherwise	0.071	0.0711	0.0053	0.0710	0.0052
AD3	1 if anxiety/depression is level 3; 0 otherwise	0.236	0.2365	0.0059	0.2363	0.0063
N3	1 if any dimension is level 3; 0 otherwise	0.269	0.2693	0.0071	0.2693	0.0071
α	Constant	0.081	0.0806	0.0078	0.0805	0.0048

Table 2: Observed and predicted trade-off valuations for the 42[†] EQ-5D health states included in the N3 model ordered by level of severity (total n=2997)

State	Original MVH exercise				Predictions from N3 model			AD	SQD
	n	Actual Mean	SD	SE	Predicted Mean	SE	Robust SE		
21111	1177	0.878	0.226	0.007	0.851	0.009	0.007	0.027	0.001
11211	1211	0.869	0.222	0.006	0.883	0.010	0.006	0.014	0.000
11121	1205	0.850	0.242	0.007	0.797	0.008	0.006	0.053	0.003
12111	1194	0.834	0.287	0.008	0.816	0.008	0.006	0.018	0.000
11112	1207	0.829	0.286	0.008	0.848	0.008	0.007	0.019	0.000
12211	750	0.767	0.321	0.012	0.780	0.009	0.007	0.013	0.000
12121	772	0.742	0.315	0.011	0.693	0.008	0.007	0.049	0.002
11122	738	0.722	0.365	0.013	0.726	0.009	0.007	0.004	0.000
22112	762	0.665	0.372	0.013	0.676	0.009	0.009	0.011	0.000
22121	761	0.642	0.421	0.015	0.625	0.009	0.008	0.017	0.000
21222	741	0.553	0.454	0.017	0.621	0.009	0.008	0.068	0.005
11312	749	0.552	0.466	0.017	0.485	0.010	0.010	0.067	0.004
12222	747	0.551	0.458	0.017	0.586	0.009	0.008	0.035	0.001
22122	743	0.540	0.467	0.017	0.554	0.009	0.009	0.014	0.000
21312	723	0.536	0.464	0.017	0.416	0.010	0.011	0.12	0.014
22222	770	0.500	0.478	0.017	0.518	0.009	0.009	0.018	0.000
11113	753	0.392	0.553	0.020	0.414	0.009	0.011	0.022	0.000
13212	745	0.389	0.532	0.019	0.330	0.009	0.011	0.059	0.003
13311	740	0.346	0.555	0.020	0.342	0.009	0.010	0.004	0.000
12223	755	0.216	0.560	0.020	0.151	0.010	0.012	0.065	0.004
11131	731	0.200	0.604	0.022	0.265	0.009	0.011	0.065	0.004
21323	747	0.160	0.588	0.022	0.128	0.009	0.010	0.032	0.001
32211	745	0.152	0.593	0.022	0.197	0.010	0.011	0.045	0.002
23321	749	0.147	0.607	0.022	0.151	0.010	0.012	0.004	0.000
21232	764	0.064	0.602	0.022	0.090	0.009	0.011	0.026	0.001
22323	743	0.042	0.583	0.021	0.024	0.009	0.011	0.018	0.000
22331	740	-0.011	0.597	0.022	-0.001	0.010	0.012	0.01	0.000
33212	767	-0.022	0.593	0.021	0.016	0.009	0.011	0.038	0.001
11133	755	-0.049	0.607	0.022	0.029	0.009	0.011	0.078	0.006
21133	752	-0.063	0.594	0.022	-0.040	0.009	0.011	0.023	0.001
23313	751	-0.070	0.586	0.021	0.037	0.009	0.011	0.107	0.011
23232	726	-0.084	0.583	0.022	-0.124	0.009	0.012	0.04	0.002
33321	741	-0.120	0.566	0.021	-0.094	0.009	0.011	0.026	0.001
22233	762	-0.142	0.568	0.021	-0.179	0.009	0.012	0.037	0.001
32313	759	-0.152	0.563	0.020	-0.098	0.009	0.010	0.054	0.003
32223	749	-0.174	0.563	0.021	-0.162	0.009	0.011	0.012	0.000
32232	749	-0.223	0.572	0.021	-0.259	0.009	0.011	0.036	0.001
13332	737	-0.228	0.551	0.020	-0.113	0.010	0.011	0.115	0.013
32331	739	-0.276	0.549	0.020	-0.246	0.010	0.011	0.03	0.001
33232	757	-0.332	0.509	0.019	-0.368	0.009	0.010	0.036	0.001
33323	761	-0.386	0.492	0.018	-0.330	0.009	0.010	0.056	0.003
33333	2997	-0.543	0.412	0.008	-0.592	0.008	0.008	0.049	0.002
MAD	0.0043								
MSD	0.0023								

[†] Respondents were also asked to value the state “unconscious”, giving a total of 43 states.

Abbreviations: AD, absolute difference [between observed predicted]; SQD, square difference [between observed predicted]; MAD, mean absolute difference; MSD, mean squared difference; MVH, measurement and valuation of health study; SD, standard deviation; SE, standard error.

Table 3: Impact of parameter uncertainty on EQ-5D utility values at 2 and 12 months, and QALYs over that period in the case study

n = 1,633	Ignoring parameter uncertainty				Allowing for parameter uncertainty			
	Endovascular Mean (SE)	Neurosurgery Mean (SE)	Mean diff. (SE) [†]	Non-parametric 95% CI	Endovascular Mean (SE)	Neurosurgery Mean (SE)	Mean diff. (SE) [‡]	Non-parametric 95% CI*
	n ₁ = 843	n ₂ = 790			n ₁ = 843	n ₂ = 790		
2 months	0.6957 (0.0107)	0.6595 (0.0112)	0.0363 (0.01542)	0.0060 to 0.0665*	0.6957 (0.0111)	0.6595 (0.0115)	0.0363 (0.01550)	0.0059 to 0.0666*
12 months	0.7337 (0.0105)	0.7136 (0.0106)	0.0201 (0.01490)	-0.0091 to 0.0493	0.7337 (0.0108)	0.7136 (0.0108)	0.0201 (0.01493)	-0.0092 to 0.0493
QALYs over 2-12 months	0.5956 (0.0081)	0.5721 (0.0082)	0.0235 (0.01149)	0.0010 to 0.0460*	0.5956 (0.0084)	0.5721 (0.0085)	0.0235 (0.01155)	0.0008 to 0.0461*

† Standard error calculated non-parametrically using bootstrapping with 50,000 replications

‡ Standard error calculated non-parametrically using 1,000 sets of EQ-5D tariff coefficients and 10,000 bootstrap samples of ISAT

* Significant difference at a p-value of 0.05

Table 4: Sensitivity analysis on of varying valuation sample size to evaluate parameter uncertainty on baseline mean differences at follow-up points and QALYs over 2-12 months

Proportion of MVH sample	Participants (number of valuations)	Ignoring parameter uncertainty	Allowing parameter uncertainty	% increase in variance from considering parameter uncertainty
		SE of mean difference [†]	SE of mean difference [‡]	
2 months				
Full sample	2,997 (35,964)	0.01542	0.01550	0.55%
60%	1,798 (21,576)	0.01547	0.01556	0.63%
30%	899 (10,788)	0.01549	0.01559	0.69%
5%	150 (1,800)	0.01572	0.01593	1.34%
1%	30 (360)	0.01475	0.01558	5.60%
12 months				
Full sample	2,997 (35,964)	0.01490	0.01493	0.20%
60%	1,798 (21,576)	0.01493	0.01497	0.27%
30%	899 (10,788)	0.01493	0.01498	0.30%
5%	150 (1,800)	0.01522	0.01535	0.83%
1%	30 (360)	0.01380	0.01447	4.87%
QALYs over 2-12 months				
Full sample	2,997 (35,964)	0.01150	0.01155	0.43%
60%	1,798 (21,576)	0.01152	0.01159	0.59%
30%	899 (10,788)	0.01153	0.01161	0.63%
5%	150 (1,800)	0.01180	0.01193	1.11%
1%	30 (360)	0.01085	0.01134	4.59%

MVH, measurement and valuation of health study; SE, standard error.

[†] Standard error calculated non-parametrically using bootstrapping with 50,000 replications

[‡] Standard error calculated non-parametrically using 1,000 sets of EQ-5D tariff coefficients and 10,000 bootstrap samples from ISAT

Figure 1: Propagating MVH parameter uncertainty

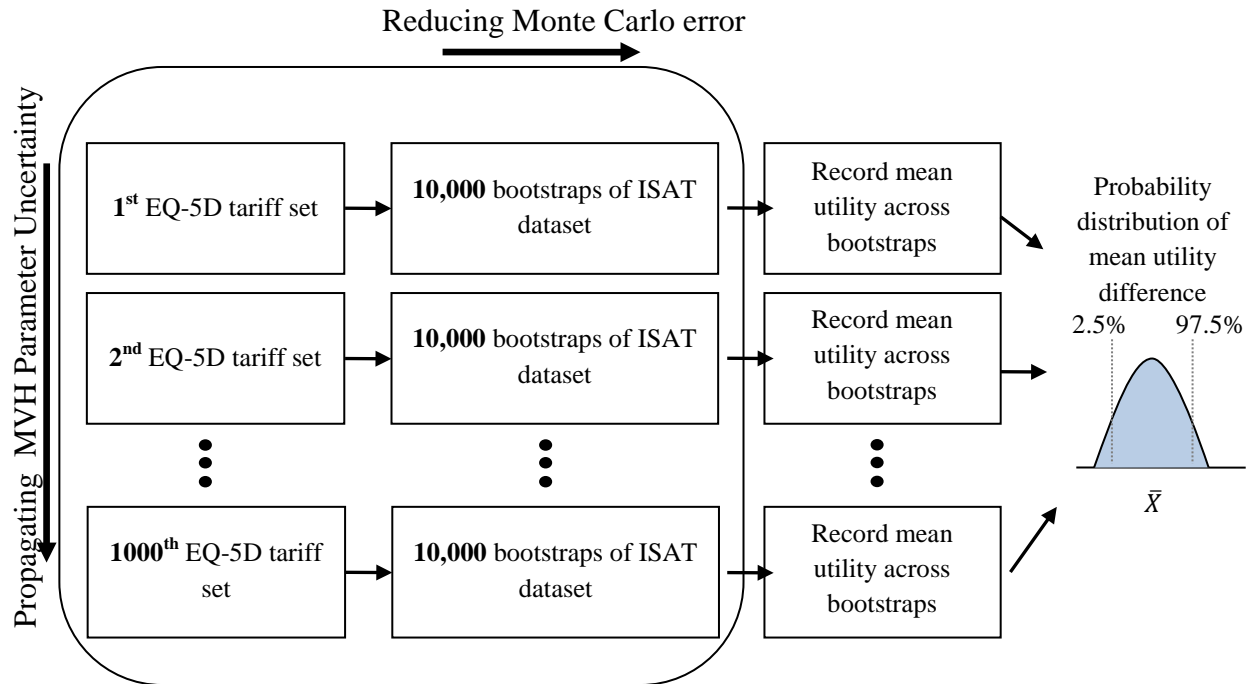


Figure 2: Distribution of TTO utility valuations among selected health states included in the original MVH exercise

