Selection of TTO health states for valuation studies that map latent utilities from DCEs to TTO utilities

Menglu Che and Eleanor M Pullenayegum

June 21, 2022

Abstract

Background: Several valuation protocols for the EQ-family of instruments envision using both latent scale DCE and TTO tasks to create value sets. Given the costly nature of TTO data, design strategies that maximize value set precision per TTO response are important. We consider how to select the set of health states to be valued using TTO so as to maximize the precision of the resulting value set.

METHODS: Under simplifying assumptions, we derived a formula describing the mean square prediction error (MSE) of the final value set as a function of (a) the number J of health states to be valued using TTO, (b) the set $S(J)$ of health states to be valued using TTO and (c) the sample variance $V(S(J))$ of the latent utilities for the health states in $S(J)$. From this we formed the following hypotheses: (i) holding J fixed, increasing $V(S(J))$ reduces MSEs; (b) holding $V(S(J))$ fixed, increasing J decreases the MSE; (c) when TTO mean utilities are in a perfect linear relationship with latent utilities and each respondent values a single health state, the MSE is minimized when $J = 2$ and the two states are chosen as far apart on the latent utility scale as possible (i.e. when $V(S(2))$ is maximized). We used simulation to evaluate whether these hypotheses held when our simplifying assumptions were relaxed. The first set of simulations assumed an underlying linear relationship between TTO and DCE utilities but that TTO mean utilities were scattered around this line. The second set of simulations were parameterized using published results from the Dutch, US, and Indonesian EQ-5D-5L valuation studies.

RESULTS: The first set of simulations provided empirical support for all three of our hypotheses. In the second set of simulations, data from Indonesia suggested an underlying linear relationship between TTO and DCE utilities, whereas both the US and the Netherlands showed a non-linear relationship. Simulations parameterized based on the Indonesian valuation data continued to provide empirical support for the hypotheses, whereas simulations parameterized using US or Dutch valuation data refuted hypothesis (i) (i.e. for a fixed number of health states valued using TTO, reducing the latent utilities of the health states reduced rather than increase the MSE).

CONCLUSIONS: Given that the underlying relationship between TTO and DCE utilities may be non-linear, we suggest choosing health states to be valued through TTO evenly across the anticipated latent utility scale.

KEY WORDS: cost-utility analysis, EQ-5D-5L, health state preference, valuation

1 Introduction

Multi-attribute utility instruments (MAUIs) are widely used to facilitate reimbursement decisions [refs]. A MAUI consists of two parts: the descriptive system that categorizes responses into health states, and a value set that assigns a utility to each health state. Value sets are estimated through valuation studies, in which respondents use a task such as standard gamble (SG) [1], time trade-off (TTO) [2] or discrete choice experiments (DCE) [3] to provide their utilities for a subset of the instrument's health states. Statistical modelling is then used to estimate utilities for the remaining health states. Since utilities are country-specific [4], developing value sets for a new instrument is a costly undertaking.

Both SG and TTO tasks need to be interviewer-facilitated [5] and are thus expensive, however DCEs can be administered online without an interviewer [3]. The simplest form of DCEs asks respondents to choose between two health states, and produces utilities on a latent scale. This form of DCE identifies latent utilities only up to a linear transform and requires additional information to anchor utilities to the cardinal scale (such that full health has a utility of 1 and states of the same utility as immediate death have a utility of 0). Various methods for anchoring have been proposed: additional DCE tasks involving trade-offs with differing durations [3, 6], full health [7], or dead [8]; a Location of Dead exercise from within a Personal Utility Function approach [9, 10]; or administering some TTO tasks in addition to

the DCE, then using a transforming or hybrid modelling approach to anchor the DCE latent utilities to the cardinal scale [11]. In this paper we consider the administration of TTO tasks coupled with a transforming approach.

In a transforming approach, the DCE data are modelled to derive estimated latent utilities for all health states in the instrument. The sample mean TTO utilities for each TTO-valued health state are then regressed onto their estimated latent utilities in order to derive an equation relating latent utilities to mean TTO utilities for all health states captured by the instrument. Differing numbers of TTO health states have been used for this. For example, in valuing the EQ-5D-Y, the Slovenian valuation study used a single TTO health state (the worst health state) [12], whereas the Japanese valuation study used 26 TTO health states [13]. In the context of the AQL-5D [14], Rowen et al showed that differences between valuing 99 health states with TTO vs. 10 and 19 were small [11].

Even when the number of states to be valued with TTO has been decided, the questions of which states should be valued, and how many valuations per state are required still need to be considered. In this paper, we make some simplifying assumptions to derive a mathematical formula for how the value set's mean square error (MSE) varies as a function of (a) the number of health states valued with TTO; (b) the spread of the health states valued with TTO; (c) the number of valuations per health state. We use this formula to generate hypotheses regarding how the MSE changes as a function of each of these criteria when our simplifying assumptions do not hold, then test these hypotheses through simulation. We conclude with some recommendations on the selection of health states to be valued using TTO.

2 TTO health state selection: theoretical approach

2.1 Model for TTO utilities

Let μ_j be the population mean utility for health state j, and let l_j be the mean latent utility. We assume the relationship between mean TTO and DCE utilities can be captured through a linear model:

$$\mu_j = \beta_0 + \beta_1 l_j + \delta_j \tag{2.1}$$

with the δs measuring the degree to which the mean TTO utilities deviate from a perfect linear transform of the DCE utilities. As in [15, 16] these can be interpreted as random

effects and we take $\delta_j \sim N(0, \sigma^2_{\delta j})$. Larger values of $\sigma_{\delta j}$ correspond to greater departures from a linear relationship, while $\sigma_\delta = 0$ corresponds to a perfect linear relationship.

We do not observe μ_j but rather seek to estimate it given latent utilities l_j that for the same of simplicity we assume are known, and TTO valuations of a subset of J health states.

Letting Y_{ij} be the observed TTO utility for subject i valuing state j, we take

$$Y_{ij} = \mu_j + \epsilon_{ij} \tag{2.2}$$

with $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon ij})$.

2.2 Simplifying assumptions

Suppose now that a total of J health states are to be directly valued, and that we have N individuals each valuing K states, chosen so that each state receives the same number NK/J valuations. Let $\mathcal{S}_J = \{j(1), ..., j(J)\}$ denote the indices of the J states to be directly valued, and let $v(\mathcal{S}_J) = v\hat{a}r(l_{j(1)}, \ldots, l_{j(J)})$ be the sample variance of latent utilities of the states selected for TTO valuation.

Recalling that latent utilities are identified only up to a linear transform, we assume without loss of generality that the latent utilities of the J health states selected for inclusion in the TTO tasks have mean zero. We make three further assumptions. First, we assume that valuations from the same individual are independent. Second, we assume that random variation of subject-level TTO utilities around the true mean utilities is homoscedastic (i.e., $\sigma_{\epsilon_{ij}}$) depends neither on i nor j). Third, we assume that the variability of the true mean utilities μ_j around the latent utilities is homoscedastic, i.e. that $\sigma_{\delta j} = \sigma_\delta$.

With these simplifying assumptions in place we can quantify accuracy of the estimated value set as a function of the number and spread of TTO health states selected for valuation, and use this formula to form hypotheses about the relationships between health state selection and accuracy of the resulting predictions. We do not expect these assumptions to hold in practice and will examine whether our hypotheses continue to hold under departures from these assumptions through simulation.

2.3 Theoretical Result

In the appendix we show that the mean squared error (MSE) over all M health states captured by the MAUI is

$$MSE = \frac{1}{M}\sum_{k=1}^{M} E\big((\hat{\mu}_k - \mu_k)^2\big) = \sigma_\delta^2 + \Big(\frac{\sigma_\delta^2}{J} + \frac{\sigma_\epsilon^2}{NK}\Big)\Big(\frac{1}{\hat{v}(\mathcal{S}_J)} + \sum_{k=1}^{M} l_k^2\Big) \tag{2.3}$$

As expected, increasing the sample size N or the number K of health states valued per individual decreases the MSE. The number K is usually chosen based on the maximum number of TTO tasks a respondent can complete before becoming tired, and is typically in the range 10-17 [17, 18, 19]. There are time and financial costs to increasing N, and we also note that once $\frac{\sigma_\epsilon^2}{NK}$ is small compared to $\frac{\sigma_\delta^2}{J}$, there are minimal benefits to increasing N, a phenomenon noted both theoretically [20] and empirically [21]. Interview procedures that improve data quality and potentially reduce σ_ϵ have been developed in the context of the EQ-5D-5L [22].

We therefore focus on the remaining modifiable factors, namely how to select J and $j(1), \ldots, j(J)$ so as to minimize the MSE.

2.4 Hypotheses

From formula (2.3) it is apparent that:

1. For a given number J of health states to be valued using TTO, larger sample variances $v(\mathcal{S}_J)$ of the latent utilities of the J health states result in smaller MSEs;
2. If the sample variance $v(\mathcal{S}_J)$ of the valued health states' latent utilities is held fixed, increasing the total number J of health states to be valued will decrease the MSE;
3. When $\sigma_\delta = 0$, the MSE is minimized when $j(1), \ldots, j(J)$ are chosen to maximize $v(\mathcal{S}_J)$.

The formula in (2.3) is based on a number of simplifying assumptions. We hypothesize that the above hypotheses will hold when there is heteroscedastcity in both the respondent level utilities (i.e., $\sigma_{\epsilon j}$) and in the state-level deviations from the mean (i.e., $\sigma_{\delta j}$), and also when each respondent contributes multiple dependent responses. We test these hypotheses using simulation.

3 Simulation

3.1 The EQ-5D-5L

For the purposes of illustration, we will focus on the EQ-5D-5L. This instrument captures five dimensions of health, namely mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension is described using 5 levels: "no", "slight", "moderate", "severe" and "extreme" problems in the corresponding dimension. The EQ-5D-5L valuation studies included direct TTO valuation of 86 out of the total 3125 (5^5) health states, organized into 10 blocks of 10 health states, as well as 196 DCE tasks, grouped into 28 blocks of 7 tasks [23].

Our simulations are parameterized using the three EQ-5D-5L valuation studies that, as of Dec 2019, reported both the sample mean and standard devation of TTO utilities for the 86 health states as well as the regression model for the latent scale DCE utilities. These countries were the Netherlands [24], the United States [25] and Indonesia [26], with respective sample sizes of 1003, 1134, and 1054.

3.1.1 Data generating mechanism

We simulated TTO data for 100 respondents, with each simulation scenario iterated 10,000 times. We considered four values of J, namely 2, 5, 10, 20. We bagan with each respondent valuing a single health state. For $J = 10, 20$, we also considered a scenario where each respondent valued 10 health states. For cases where each respondent valued 10 states, we induced within-respondent correlation in the TTO responses through a multivariate normal distribution. The respondent-level variances $\sigma^2_{\epsilon j}$ were the same as the case where each respondent valued a single health state, but with a within-respondent correlation of 0.5.

Note that the sample variance of the latent utilities of the selected health states for a given J is maximized by choosing states as far away from their sample mean as possible. For example, when $J = 2$, the sample variance will be maximized by choosing the state with the largest latent utility and the state with the lowest latent utility. For $J > 2$ preferentially choosing health states from the ends of the scale will lead to larger variances of the latent utilities than choosing states evenly distributed across the scale. We thus sampled health states as illustrated in the Appendix.

We considered two data generation procedures for the mean TTO utilities μ_j. In the first

we simulated μ_j from hypothetical distributions, while in the second we used the reported state-level mean from the three published valuation studies. We describe each of these in turn.

For the hypothetical distributions, latent utilities l_j were calculated using the published model for DCE latent utilities for Indonesia [26]. We then simulated true mean TTO utilities using equation (2.1) with $\beta_0 = 0, \beta_1 = 1$, and considered three choices for $\sigma_{\delta j}$: $\sigma^2_{\delta j} = 0.001$, representing a near perfect linear relationship between mean TTO and DCE utilities; $\sigma_{\delta j} = expit(-1.5\mu_j - 3.5)$, representing a mild scatter of mean TTO utilities around DCE latent utilities; and $\sigma_{\delta j} = expit(-1.25\mu_j - 3)$, representing a moderate scatter. The extent of the scatter for the three scenarios is depicted in Figure 1. Respondent-level TTO utilities for subject i valuing state j were generated using equation (2.2) with $var(\epsilon_{ij}) = \sigma^2_{Yj}$ taken to be the state-specific sample variance from the Indonesian valuation study.

For the simulations with the mean TTO utilities parameterized using data from published EQ-5D-5L valuation studies, we used the published model for the DCE data from each of the three countries to calculate the latent utilities l_j, and took the mean TTO utilities μ_j for each of the 86 health states as the reported sample means. Respondent-level TTO utilities for subject i valuing state j were generated using equation (2.2) with $var(\epsilon_{ij}) = \sigma^2_{Yj}$ taken to be the state-specific sample variance from each country's published valuation study.

3.1.2 Analytic procedure

We used ordinary least squares (OLS) to regress TTO responses onto latent utilities; note that OLS will give unbiased estimates of mean utilities even when error terms are dependent [27]. The fitted regression coefficients were used to capture the predicted mean utilities $\hat{\mu}_j$ for the 86 states captured in the valuation studies. The mean square error (MSE) and mean absolute error (MAE) of these predictions was calculated by comparing the predicted values $\hat{\mu}_j$ to the true values μ_j specified in the data generating mechanism. These were then averaged over all 86 health states.

4 Results

4.1 Hypothetical TTO distributions

Data generated from hypothetical TTO distributions provided empirical support for all three of our hypotheses (Table 1).

Firstly, weighted health state selection led to both larger sample variances of the latent utilities among selected health states and also to smaller RMSEs and MAEs. This held for every combination of number of states sampled, number of states valued by each respondent, and extent of scatter. The difference between weighted and even health state selection diminished as the number of health states assessed increased.

Secondly, the sample variances for $J = 5$ and $J = 10$ health states were similar (0.30 and 0.31) under even health state selection, but the RMSEs for $J = 10$ health states were smaller than those for $J = 5$ for all three levels of scatter, providing empirical support for our hypothesis that for a given sample variance, valuing more health states leads to smaller RMSEs.

Thirdly, when the TTO mean utilities corresponded very closely to the latent utilities ("Minimal Scatter" scenario in Table 1) and each respondent valued a single health state, the smallest RMSE when occurred when just two health states were valued and weighted health state selection was used.

As expected, the greater the scatter, the larger the RMSEs and MAEs. Having each respondent value 10 health states rather than a single health state led to smaller MSEs and MAEs, regardless of the total number of health states captured using TTO or the extent of the scatter.

When each respondent valued 10 health states, using a design that captured 20 health states outperformed a design that captured 10 health states in each scenario, even though the sample variances of the valued health states were smaller.

4.2 TTO distributions drawn from published valuation studies

Turning now to the case where data was simulated using reported TTO mean utilities, note from Figure 2 that the relationship between mean TTO utilities and mean latent utilities shows two departures from linearity: a non-linear relationship between TTO and latent utilities and a scattering of points around the fitted curve. While for Indonesia the OLS-

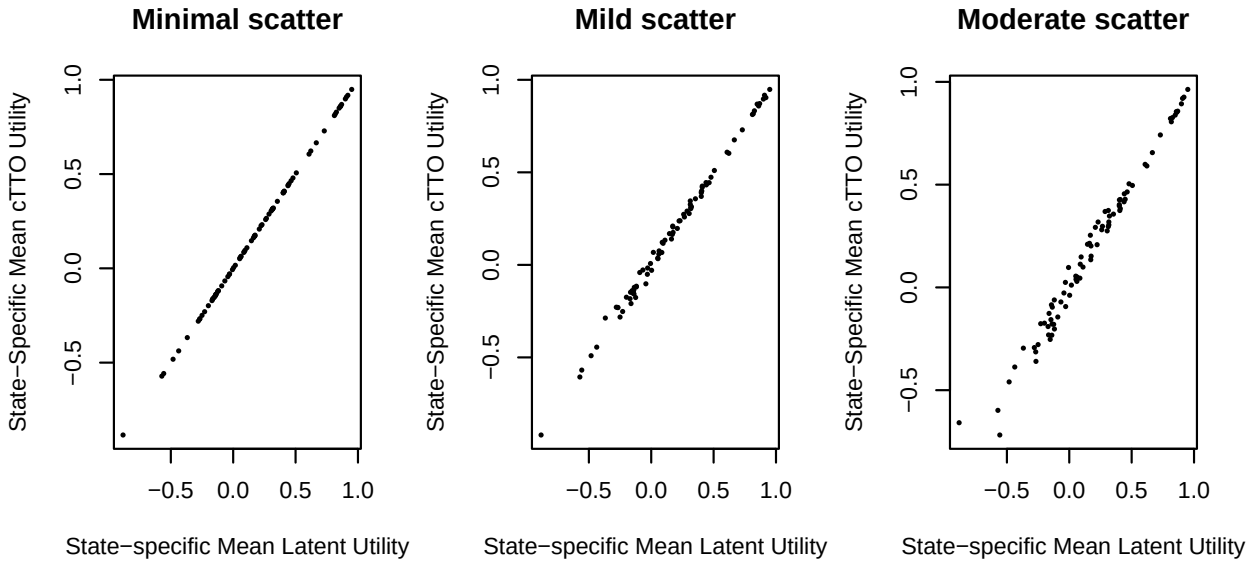

Figure 1: Scatterplots of the simulated TTO means vs latent DCE utitlities

	states per respondent	# states	Minimal scatter		Mild scatter		Moderate scatter	
			weighted	even	weighted	even	weighted	even
RMSE	1	2	0.024	0.087	0.068	0.099	0.052	0.109
	1	5	0.031	0.045	0.052	0.057	0.056	0.067
	1	10	0.033	0.042	0.045	0.056	0.055	0.062
	1	20	0.039	0.048	0.050	0.059	0.059	0.065
	10	10	0.020	0.025	0.042	0.045	0.051	0.056
	10	20	0.016	0.019	0.039	0.041	0.049	0.051
MAE	1	2	0.022	0.073	0.056	0.081	0.042	0.088
	1	5	0.029	0.040	0.041	0.046	0.044	0.054
	1	10	0.030	0.038	0.036	0.046	0.043	0.050
	1	20	0.035	0.043	0.040	0.048	0.046	0.053
	10	10	0.019	0.024	0.031	0.035	0.039	0.044
	10	20	0.016	0.018	0.029	0.031	0.037	0.040
Sample variance	1	2	1.68	0.10	1.68	0.10	1.68	0.10
	1	5	0.75	0.30	0.75	0.29	0.75	0.30
	1	10	0.56	0.31	0.56	0.31	0.56	0.31
	1	20	0.37	0.19	0.37	0.20	0.37	0.19
	10	10	0.56	0.31	0.56	0.31	0.56	0.31
	10	20	0.37	0.19	0.37	0.19	0.37	0.19

Table 1: Root mean square and mean absolute errors with simulated departures from a perfect linear relationship between TTO utilities and latent DCE utilities. We used 100 respondents and 10000 iterations

derived line and the loess fit are almost indistinguishable, in the US the two lines separate at both ends of the latent utility scale, and in the Netherlands there is a very clear non-linear relationship.

The results for Indonesia, where the only departure from linearity is a random scatter, continue to provide empirical support for our first two hypotheses: weighted health state selection led to larger sample variances and smaller RMSEs for all scenarios with respondents valuing a single health state (see Table 2). Moreover, the sample variances for $J = 5$ and $J = 10$ were similar under even health state selection (0.30 and 0.31 respectively) while the RMSE for 10 health states was smaller than that for 5. The Indonesian set-up was uninformative about the last hypothesis, which requires the random scatter around a straight line to be small.

For the Netherlands and the US, which both showed a non-linear relationship between latent and TTO utilities, weighted health state selection led to larger RMSEs than even health state selection, even though the sample variances for weighted selection were larger. The RMSEs were largest for the Netherlands, which had the most severe departure from linearity. In both countries valuing more health states resulted in smaller RMSEs and MAEs.

5 Discussion

We showed theoretically and empirically that when mean TTO utilities are a linear transform of latent utilities, MSEs and MAEs are minimized by selecting health states from the two ends of the scale. Moreover, if each subject is to value a single health state MSEs and MAEs are minimized by having a total of two health states valued: the one with the largest and the one with the smallest latent utility.

However, these results no longer hold with the relationships between TTO and latent utilities seen in practice. The valuation studies we analysed showed two departures from linearity: random variability around a straight line relationship, and an underlying non-linear relationship. We discuss each in turn.

In the Indonesian valuation study the plot of mean TTO utilites vs. latent utilities showed a random scatter around a straight line, and we observed that increasing the number of health states to be valued to 10 or 20 performed better than using 2 or 5 health states. Furthermore, in line with our finding that the benefits of weighted health state selection

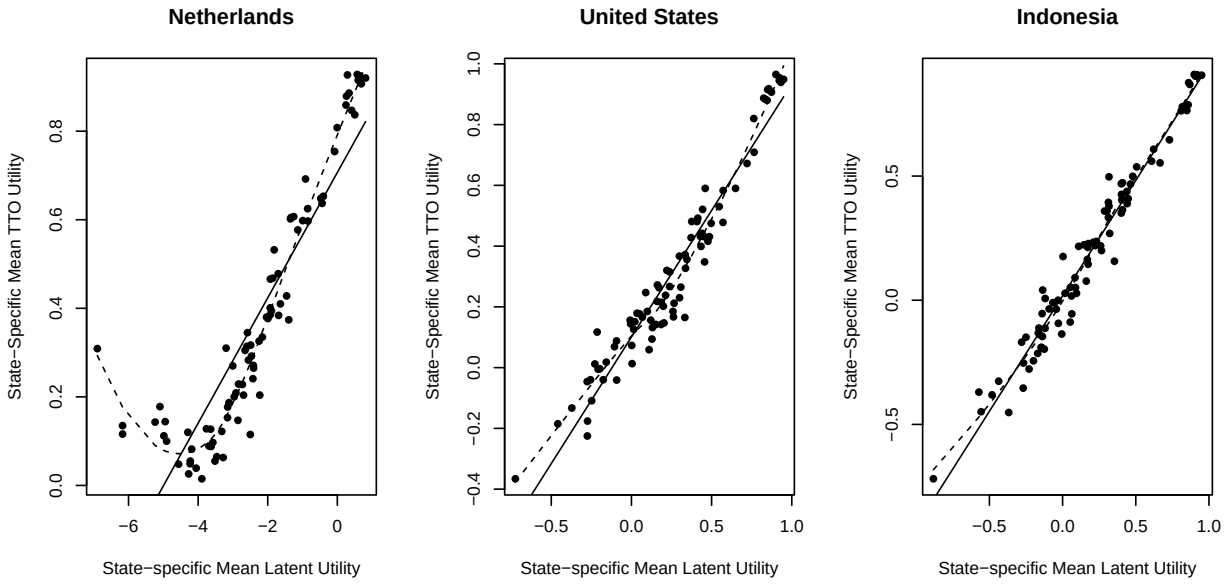

Figure 2: Scatterplots of mean TTO utilities vs. Estimated latent DCE utilities for 3 countries

	states per respondent	# states	Netherlands		US		Indonesia	
			weighted	even	weighted	even	weighted	even
RMSE	1	2	0.338	0.222	0.125	0.192	0.086	0.168
	1	5	0.271	0.153	0.123	0.110	0.088	0.093
	1	10	0.220	0.164	0.116	0.106	0.085	0.088
	1	20	0.162	0.147	0.111	0.109	0.084	0.090
	10	10	0.217	0.157	0.109	0.093	0.080	0.081
	10	20	0.155	0.133	0.098	0.087	0.075	0.077
MAE	1	2	0.300	0.156	0.105	0.153	0.068	0.136
	1	5	0.238	0.119	0.102	0.090	0.070	0.074
	1	10	0.190	0.139	0.096	0.087	0.067	0.071
	1	20	0.135	0.119	0.092	0.089	0.067	0.072
	10	10	0.188	0.135	0.088	0.076	0.063	0.064
	10	20	0.130	0.111	0.079	0.071	0.059	0.061
Sample variance	1	2	29.71	1.65	1.40	0.07	1.68	0.10
	1	5	13.82	5.67	0.63	0.24	0.75	0.29
	1	10	10.90	5.79	0.48	0.27	0.56	0.31
	1	20	7.27	3.66	0.33	0.17	0.37	0.19
	10	10	10.91	5.79	0.48	0.27	0.56	0.31
	10	20	7.27	3.66	0.33	0.17	0.37	0.20

Table 2: Root mean square and mean absolute errors using data generating mechanisms based on the US, Dutch and Indonesian valuation studies. We used 100 respondents and 10000 iterations

diminish as the scatter of mean TTO utilities around the latent utilities increases, weighted health state selection was only marginally superior to even health state selection.

The US valuation study showed a mild departure from an underlying linear relationship between TTO and latent utilities, and as in the Indonesian valuation study valuing 10 or 20 health states performed better than using 2 or 5. Furthermore, once at least 10 health states were valued, even health state selection performed better than weighted health state selection.

We studied both the case where each respondent values a single health state and where each respondent values 10 health states using TTO. Studies have shown respondents' utilities for a given health state depend on which other health states they have been presented with. Due to this framing effect, the valuation protocol for the EQ-5D-5L has each respondent seeing both a mild and the most severe health state.

Note that there is no theoretical reason to expect latent utilities and TTO utilities to follow an underlying linear relationship [28, 29]. When the underlying relationship is non-linear, using non-linear transforms (e.g. splines or fractional polynomials) may lead to better predictive precision. These transforms would require TTO means over the full range of latent utilities.

There are several limitations to this work. While we considered up to 20 health states to be valued using TTO, we did not consider larger numbers. We thus cannot comment on whether having 100 TTO respondents valuing 10 health states each, with those health states chosen from among 50 or 100 health states would do better than valuing a total of 20 health states. Secondly, we did not consider using non-linear transforms; this is the subject of ongoing work. Thirdly, we did not consider sample sizes larger than 100; the rationale for this was that the motivation for the study design was to have a relatively small sample of respondents completing the TTO so as to reduce the costs of the study. Fourthly, the Dutch valuation study was in the first wave of EQ-5D-5L valuation studies; subsequent studies used an updated protocol with quality control criteria and showed less departure from linearity. Consequently, the extent of departure from linearity in the Dutch valuation study may not represent the true nature of the TTO-latent utility relationship. However, the US valuation study used the updated protocol yet still shows small departures from linearity, so we believe this remains an important feature to account for when designing the TTO portion of the valuation study.

We thus recommend that (a) each respondent values xx to 10 health states using TTO; (b) 20 or more health states be directly valued using TTO; (c) these health states be spread evenly over the range of severities.

A Mathematical Derivation

Let $X_k = (1, l_k)$ and note that $\hat{\mu}_k = X_k\hat{\beta}$, where $\hat{\beta}$ is the estimated value of β. Furthermore, since we have assumed homosedasticity, equal numbers of valuations for each health state, and independence of valuations within subjects, ordinary least squares (OLS) regression on the responses for each subject for each state is the same as weighted least squares regression on the state-specific means [15], which in turn is the same as ordinary least squares regression on the state-specific means (due to homoscedasticity). Letting $\bar{Y}_k$ denote the mean TTO valuation for health state k, observe that

$$\bar{Y}_k = X_k\beta + \delta_k + \bar{\epsilon}_k,$$

where $\bar{\epsilon}_k$ is the sample mean of the subject-specific errors for health state k. Since each state is valued by NK/J subjects and each ϵ has variance σ^2_ϵ, $\bar{\epsilon_k}$ has variance $\frac{J\sigma^2_\epsilon}{NK}$. We therefore have that the

$$\bar{Y}_k = X_k\beta + \epsilon^*_k, \text{ with } \epsilon^*_k = \delta_k + \bar{\epsilon}_k$$

so that $var(\epsilon^*_k) = \sigma^2_\delta + \frac{J\sigma^2_\epsilon}{NK}$. Since we have assumed that the sample mean of the latent utilities among valued health states is zero, it follows from standard results on OLS that

$$\begin{aligned} E(\hat{\beta}) &= \beta \\ var(\hat{\beta}) &= \Big(\frac{\sigma^2_\delta}{J} + \frac{\sigma^2_\epsilon}{NK}\Big)\begin{pmatrix} \frac{1}{\hat{v}(\mathcal{S}_J)} & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

It follows that

$$\begin{aligned} E\big((\hat{\mu}_k - \mu_k)^2\big) &= E\Big((X_k(\hat{\beta} - \beta) + \delta_k)^2\big) \\ &= X_k var(\hat{\beta} - \beta) X_k' + \sigma_\delta^2 \\ &= \Big(\frac{\sigma_\delta^2}{J} + \frac{\sigma_\epsilon^2}{NK}\Big)\left(\frac{1}{\hat{v}(\mathcal{S}_J)} + l_k^2\right) + \sigma_\delta^2. \end{aligned}$$

Consequently, computing the mean of the square predictions over all health states, we have

$$MSE = \frac{1}{M}\sum_{k=1}^{M} E\big((\hat{\mu}_k - \mu_k)^2\big) = \sigma_\delta^2 + \Big(\frac{\sigma_\delta^2}{J} + \frac{\sigma_\epsilon^2}{NK}\Big)\Big(\frac{1}{\hat{v}(\mathcal{S}_J)} + \sum_{k=1}^{M} l_k^2\Big)$$

B Health State Selection

Number of health states valued →	2		5		10		20	
Health State Number ↓	Weighted	Even	Weighted	Even	Weighted	Even	Weighted	Even
1	1		1		1	1	1	1
2								
3								
4								
5			2	1	4		4	1
6								
7								
8								
9								
10								
11								
12						2		
13								
14							5	2
15								
16								
17								
18								
19								
20								
21								
22								2
23								
24								
25		1						
26								
27				1				
28								
29								
30								
31								2
32						2		
33								
34								
35								
36								2
37								
38								
39								
40								
41								
42								
43								
44								
45								
46								
47				1				2
48								
49								
50								
51								
52								
53								
54						2		
55								
56								
57								2
58								
59								
60		1						
61								
62								
63								
64								
65				1				2
66								
67								
68								
69								
70								
71								
72								
73								
74						2	5	2
75								
76								
77								
78								
79								
80								
81			1		5			
82								
83							5	2
84								
85								
86	1		1	1		1		

References

[1] Gafni A. The standard gamble method: what is being measured and how it is interpreted. *Health services research* 1994; 29(2): 207.

[2] Wee HL, Li SC, Xie F et al. Validity, feasibility and acceptability of time trade-off and standard gamble assessments in health valuation studies: a study in a multiethnic asian population in singapore. *Value in health* 2008; 11: S3–S10.

[3] Bansback N, Brazier J, Tsuchiya A et al. Using a discrete choice experiment to estimate health state utility values. *Journal of health economics* 2012; 31(1): 306–318.

[4] Norman R, Cronin P, Viney R et al. International comparisons in valuing eq-5d health states: a review and analysis. *Value in Health* 2009; 12(8): 1194–1200.

[5] Stolk E, Ludwig K, Rand K et al. Overview, update, and lessons learned from the international eq-5d-5l valuation work: version 2 of the eq-5d-5l valuation protocol. *Value in Health* 2019; 22(1): 23–30.

[6] Norman R, Viney R, Brazier J et al. Valuing sf-6d health states using a discrete choice experiment. *Medical Decision Making* 2014; 34(6): 773–786.

[7] Viney R, Norman R, Brazier J et al. An australian discrete choice experiment to value eq-5d health states. *Health Economics* 2014; 23(6): 729–742.

[8] Brazier J, Rowen D, Yang Y et al. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. *The European journal of health economics* 2012; 13(5): 575–587.

[9] Devlin NJ, Shah KK, Mulhern BJ et al. A new method for valuing health: directly eliciting personal utility functions. *Eur J Health Econ* 2019; 20(2): 257–270. DOI: 10.1007/s10198-018-0993-z. URL `https://www.ncbi.nlm.nih.gov/pubmed/30030647`. Devlin, Nancy J Shah, Koonal K Mulhern, Brendan J Pantiri, Krystallia van Hout, Ben 2018/7/22.

[10] Shah KK, Ramos-Goñi JM, Kreimeier S et al. An exploration of methods for obtaining 0=dead anchors for latent scale eq-5d-y values. *Eur J Health Econ* 2020; 21(7):

1091–1103. DOI:10.1007/s10198-020-01205-9. URL https://www.ncbi.nlm.nih.gov/pubmed/32506281. Shah, Koonal K Ramos-Goñi, Juan Manuel Kreimeier, Simone Devlin, Nancy J 2020/6/9.

[11] Rowen D, Brazier J and Van Hout B. A comparison of methods for converting dce values onto the full health-dead qaly scale. *Medical Decision Making* 2015; 35(3): 328–340.

[12] Prevolnik Rupel V, Ogorevc M and Group IHH. Eq-5d-y value set for slovenia. *Pharmacoeconomics* 2021; 39(4): 463–471. DOI:10.1007/s40273-020-00994-4. URL https://www.ncbi.nlm.nih.gov/pubmed/33565048. Prevolnik Rupel, Valentina Ogorevc, Marko 2021/2/11.

[13] Shiroiwa T, Ikeda S, Noto S et al. Valuation survey of eq-5d-y based on the international common protocol: Development of a value set in japan. *Med Decis Making* 2021; 41(5): 597–606. DOI:10.1177/0272989X211001859. URL https://www.ncbi.nlm.nih.gov/pubmed/33754886. Shiroiwa, Takeru Ikeda, Shunya Noto, Shinichi Fukuda, Takashi Stolk, Elly 2021/3/24.

[14] Yang Y, Brazier JE, Tsuchiya A et al. Estimating a preference-based index for a 5-dimensional health state classification for asthma derived from the asthma quality of life questionnaire. *Med Decis Making* 2011; 31(2): 281–91. DOI:10.1177/0272989X10379646. URL https://www.ncbi.nlm.nih.gov/pubmed/20924045. Yang, Yaling Brazier, John E Tsuchiya, Aki Young, Tracey A 2010/10/7.

[15] Pullenayegum EM, Chan KK and Xie F. Quantifying parameter uncertainty in eq-5d-3l value sets and its impact on studies that use the eq-5d-3l to measure health utility: A bayesian approach. *Med Decis Making* 2016; 36(2): 223–33. DOI:10.1177/0272989X15591966. URL https://www.ncbi.nlm.nih.gov/pubmed/26139449.

[16] Shams S and Pullenayegum E. Reducing uncertainty in eq-5d value sets: The role of spatial correlation. *Medical Decision Making* 2019; 39(2): 91–99.

[17] Tsuchiya A, Ikeda S, Ikegami N et al. Estimating an eq-5d population value set: the case of japan. *Health economics* 2002; 11(4): 341–353.

[18] Lamers LM, McDonnell J, Stalmeier PF et al. The dutch tariff: results and arguments for an effective design for national eq-5d valuation studies. *Health economics* 2006; 15(10): 1121–1132.

[19] Devlin N, Roudijk B and Ludwig K. Value sets for eq-5d-5l: A compendium, comparative review & user guide 2022; .

[20] Chan KKW and Pullenayegum EM. The theoretical relationship between sample size and expected predictive precision for eq-5d valuation studies: A mathematical exploration and simulation study. *Med Decis Making* 2020; 40(3): 339–347. DOI:10.1177/0272989X20915452. URL `https://www.ncbi.nlm.nih.gov/pubmed/32428427`. Chan, Kelvin K W Pullenayegum, Eleanor M 2020/5/20.

[21] Hansen TM, Stavem K and Rand K. Sample size and model prediction accuracy in eq-5d-5l valuations studies: Expected out-of-sample accuracy based on resampling with different sample sizes and alternative model specifications. *MDM policy & practice* 2022; 7(1): 23814683221083839.

[22] Ramos-Goñi JM, Oppe M, Slaap B et al. Quality control process for eq-5d-5l valuation studies. *Value Health* 2017; 20(3): 466–473. DOI:10.1016/j.jval.2016.10.012. URL `https://www.ncbi.nlm.nih.gov/pubmed/28292492`. Ramos-Goñi, Juan M Oppe, Mark Slaap, Bernhard Busschbach, Jan J V Stolk, Elly 2017/3/16.

[23] Oppe M and Van Hout B. The "power" of eliciting eq-5d-5l values: the experimental design of the eq-vt. *EuroQolWorking Paper Series* 2017; 17003.

[24] Versteegh MM, Vermeulen KM, Evers SM et al. Dutch tariff for the five-level version of eq-5d. *Value in health* 2016; 19(4): 343–352.

[25] Pickard AS, Law EH, Jiang R et al. United states valuation of eq-5d-5l health states using an international protocol. *Value in Health* 2019; 22(8): 931–941.

[26] Purba FD, Hunfeld JA, Iskandarsyah A et al. The indonesian eq-5d-5l value set. *PharmacoEconomics* 2017; 35(11): 1153–1165.

[27] Wooldridge JM. *Introductory econometrics: A modern approach*. Cengage learning, 2015.

[28] Pullenayegum E and Xie F. Scoring the 5-level eq-5d: can latent utilities derived from a discrete choice model be transformed to health utilities derived from time tradeoff tasks? *Medical Decision Making* 2013; 33(4): 567–578.

[29] Xie F, Pullenayegum E, Pickard AS et al. Transforming latent utilities to health utilities: East does not meet west. *Health economics* 2017; 26(12): 1524–1533.