Using Discrete Choice Experiments (DCE) based approaches to produce anchored values: comparative results from application to EQ-5D-Y-3L

Tianxin Pan[1,2], Nancy Devlin[1], Brendan Mulhern[3], Richard Norman[2]

1. Health Economics Unit, Centre for Health Policy, Melbourne School of Population and Global Health, The University of Melbourne, 207 Bouverie Street, Carlton, VIC, 3053, Melbourne, Australia.
2. School of Population Health, Curtin University, Perth, WA, Australia. tianxin.pan1@unimelb.edu.au.
3. Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, NSW, Australia.

Funding: This paper is part of a wider study to produce an EQ-5D-Y-3L value set in Australia using the standard protocol, funded by the EuroQol Research Foundation (EQ Project 83-2020VS). Views expressed in this paper are not necessarily those of the EuroQol Group.

Abstract

Objective:

In the international protocol for valuing EQ-5D-Y, both DCE and cTTO are included. However, the role of DCE is to determine the relative importance of dimensions and severity levels. Little methodological work has been done to explore the feasibility and appropriateness of using DCE to generate values and convert to QALY scale. The Australian EQ-5D-Y-3L valuation study followed the international protocol in collecting both latent scale DCE data and cTTO data. Alongside the latent scale DCE, we also included DCE tasks containing either duration or dead. The aim of this paper is to explore and compare different DCE-based approaches to valuing EQ-5D-Y, and anchoring values onto the utility scale.

Methods:

The choice experiment comprised three components: (1) latent scale DCE, following the Y protocol using 10 blocks and 15 choice sets per block, with each choice set including two health state options; (2) DCE + duration, using 15 blocks and 3 choice set per block, with each choice set including two EQ-5D-Y-3L health profile options, and a duration attribute (1, 4, 7 or 10 years); (3) DCE + dead – an unblocked design of 32 pairs, with each choice set involving comparisons of a 'severe' health state (all dimensions at least level 2) to being dead. Respondents were recruited by Survey Engine, and each answered 21 choice sets (15+3+3).

We explored four approaches (A-D). (A) uses DCE + duration as a 'stand-alone' approach, estimated using 1) an 11 parameter main effects model (i.e. duration and the 10 interactions between duration and level 2 and 3 of each of the five dimensions); and 2) a model introducing an additional interaction captured by N3 term; using the wtp Stata command to estimate QALY weights. We also analysed DCE latent scale data (using a mixed logit model allowing for unobservable random preference heterogeneity) and used three approaches to rescale these values onto a QALY scale: (B) a linear mapping model between latent scale coefficients and the utility decrements produced from Approach A; (C) anchoring on the pits state (33333) value obtained from Approach A; and (D) anchoring on dead, through a logit model to identify the relative position of health states and dead, and setting to zero the value of the health state with a 50% chance of being preferred to dead.

Results

A representative sample of 1002 adults completed DCE online between Dec 2021 and Jan 2022. DCE + duration data suggested PD was the most important dimension (with the largest overall decrement), followed by AD, MO, UA and SC. The N3 coefficient was negative but not statistically significant at the 10% level; the Likelihood-ratio test rejected the hypothesis that the interaction model outperformed the main effect model, supporting the conclusion of AIC and BIC. The estimated utilities from the main effect model range from -0.332 to 1. Results from the mixed logit model on latent DCE suggested different dimension ranking: PD, AD, UA, SC and MO. Utilities have different ranges when anchoring using approaches B (-0.319 to 1) and C (-0.332 to 1). Under approach D, the level decrements in all dimensions were smaller compared to previous approaches, resulting in a narrower range of utilities (0 to 1).

Conclusion

DCEs are feasible as a stand-alone approach to producing interpretable value sets for the EQ-5D-Y. However, different DCE approaches produce different results both with respect to the relative importance of dimensions and level decrements, and different value ranges. It is not clear what criteria should be used to choose between the alternative approaches.

1. Introduction

Both time trade-off (TTO) and Discrete choice experiments (DCEs) have been used to elicit preference for adult and child health-related quality of life (HRQoL) instruments. The time trade-off (TTO) has been widely used as a method for valuing EQ-5D instruments and has broad acceptability e.g., to HTA agencies such as NICE. However, TTO is costly and complex, and there remain unresolved issues with it e.g., by convention, duration has been fixed at 10 years, despite evidence showing violations of the assumption of constant proportionality which this rests on. In the space of valuing child health, the use of TTO has still further issues. For example, adults valuing EQ-5D-Y-3L health states appear to be being unwilling to trade child life years, resulting in relatively high health related quality of life (HRQoL) values and a narrower range of values (compared to values for adult EQ-5D instruments) (Devlin et al., 2022 ; Reckers-Droog et al., 2022). Further, given the potential interest in and relevance of adolescents' stated preferences, TTO is problematic because of the appropriateness and acceptability of its use to elicit preferences from adolescents. For all these reasons, there is continued interests in investigating alternatives to TTO which could provide potential future methods.

DCEs have been increasingly used in the development of preference weights for HRQoL. A recent review found 63 studies used DCEs for health state valuation as of 2018; more than half of which were for the EQ-5D instruments (Mulhern et al., 2019). Using DCEs as a stand-alone method of valuation have also been of long-standing interest to the EuroQol Group (Pullenayegum et al., 2020). The Australian value sets for EQ-5D-3L and EQ-5D-5L used a DCE + duration approach (Norman et al., 2013; Viney et al., 2014). Methodological research has explored different DCE-based approaches and anchoring methods to value adult instruments (Norman et al., 2016a; Pullenayegum et al., 2020; Stolk et al., 2010). However, less work has been done on using DCE to generate values on a quality-adjusted life-year (QALY) scale for child HRQoL instruments. Craig et al. (2016) elicited preference from US adult general population sample and summarized EQ-5D-Y-3L on QALY scale. The choice tasks were paired comparisons between a loss in child HRQoL and a loss in lifespan, and comparisons between two child HRQoL losses; where the description of HRQoL only included two EQ-5D-Y-3L dimensions rather than full profiles. Rowen et al. (2018) estimated a Dutch value set for CHU9D using online DCE + duration among adult general population sample.

In the international protocol for valuing EQ-5D-Y-3L (Ramos-Goñi et al., 2020), DCE is the principal method used to determine the relative importance of dimensions and severity levels; the role of TTO is 'scaled back' (compared to its usedin valuing adult EQ-5D states) and is restricted to anchoring the DCE data. Since the publication of the international protocol in 2020, four value set studies have been published: Slovenia (Prevolnik Rupel and Ogorevc, 2021), Japan (Shiroiwa et al., 2021), Spain (Ramos-Goñi et al., 2021) and Germany (Kreimeier et al., 2022). Seven other studies are underway or close to completion (for details, see (Devlin et al., 2022). Experience from completed and ongoing valuation studies suggest that, similar to the work in the adult space, the most appropriate means of anchoring the latent scale values produced by DCE remains unclear (Devlin et al., 2022 ; Rowen et al., 2014). Addressing methods for anchoring DCE data onto the QALY scale has been identified as a priority within the EuroQol Group. In particular, whether to continue with combining DCE and cTTO data (in keeping with the current protocol, but with an amended TTO design) or to introduce new methods, for example, variants of DCE which include comparisons of states with dead, or that include duration as one of the attributes (Devlin et al., 2022).

Given the role of DCE in the protocol for valuing both EQ-5D and EQ-5D-Y-3L instruments, and the increase in use of DCE for valuing child HRQoL generally (Bailey et al., 2022), there is particular interest in exploring whether extended forms of DCE have the potential to be used as 'stand-alone' approach, rather than as a method complementary to TTO in valuing EQ-5D-Y. The Australian EQ-5D-Y-3L valuation study extended the standard protocol by including additional DCE tasks containing either a duration attribute or dead. This paper aims to explore different approaches and compare values generated from different approaches and potential implication of the method chosen on value characteristics.

2. Method

2.1 Choice task and design

The choice experiment comprised three components: (1) latent scale DCE without duration; (2) DCE + durations; and (3) DCE + dead. All participants completed the three components.

The first component was latent scale DCE without duration, following the EQ-5D-Y-3L international valuation protocol. The latent scale DCE task used pairwise comparisons, with each choice set including two health profiles defined by five dimensions of EQ-5D-Y-3L descriptive system. The respondent was asked to consider which of the two health states, A and B, they preferred for a 10 year old hypothetical child. The DCE design was D-efficient and consisted of 150 DCE pairs separated into ten blocks ensuring equal responses across choice pairs and approximate level balance within each person (Ramos-Goñi et al., 2020). A target sample size of 1000 participants suggested by the protocol was followed, which meant 100 observations per choice pair.

The second component was the DCE + duration task, with each choice set including two health profile options, defined by five dimensions of EQ-5D-Y-3L and a survival duration attribute. Four survival durations (1, 4, 7 or 10 years) were included in the experiment. The durations were selected to be spread across the range of those used in a conventional TTO task for the EQ-5D (which typically varies survival duration from 0 to 10 years). The construction of the choice sets was designed to be D-efficient and allow for estimation of the interaction between duration and each of the health state attributes, using multinomial logit models. All the coefficients were assumed to be zero other than a small positive coefficient on duration. A swapping algorithm produced a design of 45 pairs divided into 15 blocks (3 pairs per block). Each of the 1000 participants completed three DCE + duration tasks, and 66 observations per choice pair was ensured.

In the DCE + dead component, each choice set involved comparisons between a dead option, and living in an EQ-5D-Y-3L state for 10 years (Roudijk et al., 2020). The EQ-5D-Y-3L state was drawn from the 32 health states for which every dimension is at level 2 or worse (ranging from 22222 to 33333). We used an unblocked design of 32 pairs. Each participant completed 3 DCE + dead tasks.

2.2 Sample and recruitment

We aimed for a sample of 1000 respondents from the adult general population in Australia to complete DCE tasks online through self-completion. Participants were recruited through Survey Engine. Quota-based sampling, on gender, age groups, and whether the respondent had ever been a parent, was applied.

2.3 Quality control

We used the time taken to complete the survey as an indicator of data quality. To explore whether respondents were speeders, a proxy for non-engagement in task completion, we conducted the following checks: (1) we examined time taken on each page during the entire survey for the full sample and for the 10th percentile (i.e. fastest 10% of respondents) and 90th percentile respectively (Norman et al., 2016b). (2) Latent scale DCE data was divided into deciles based on the time they spent completing the 15 choice sets. We repeated the analysis using conditional logit models for each group and examined if preferences differ among these groups. We also presented pseudo R^2 for each group (Norman et al., 2021). (3) As a sensitivity analysis, we repeated the regression analysis by excluding the fastest 10% respondents across the entire DCE tasks (i.e. 21 choice tasks) and compared regression results with a full sample analysis. (4) We examined the proportion of the respondents who selected all 'option A' and all 'option B' in the three DCE components. In addition, structured feedback questions regarding difficulty and understanding of the tasks were asked to all respondents at the end of the survey.

We did not include repeated pairs or logically dominant pairs in the DCE tasks. Jonker et al. (2022) found that dominant and repeated choice tasks were unreliable screening tests and poor indicators of response quality. New evidence from the Dutch EQ-5D-Y-3L value set study investigated the effect of removing respondents based on speed of task completion and failing dominant choice tasks. These were correlated; but judging quality based on speed had the most impact on improving models (personal correspondence with B. Roudijk; paper yet to be published).

2.4 Analysis

We explored four approaches. Approach A used DCE + duration data as a stand-alone approach. Approaches B, C and D modelled latent DCE data as a base analysis and applied different anchoring strategies.

Approach A: The DCE + duration approach

This approach is similar to that taken in the published studies using DCE + duration data in valuing adult EQ-5D health states (Bansback et al., 2012; Norman et al., 2016a; Viney et al., 2014). Equation 1 sets out the broad approach to the utility specification to allow for estimation of QALY weights. The utility of alternative *j* in scenario *s* for individual *i* was:

$$U_{isj} = \alpha TIME + \beta X'_{isj} TIME + \varepsilon_{isj} \quad \text{(Equation 1)}$$

where X'_{isj} was a set of dummy relating to the levels of the EQ-5D-Y-3L health state presented in option *j*. The marginal rate of substitution for each of the levels of each of the dimensions of each of the utility instruments (other than level 1, which was set as the base in the regression) was estimated using TIME (the life expectancy variable) as the numeraire. The marginal utility of TIME was:

$$\frac{\delta U}{\delta TIME} = \alpha + \beta X' \quad \text{(Equation 2)}$$

To generate QALY weights for health states, we estimated the ratio of the marginal utility of TIME for the health state being valued and the marginal utility of TIME for full health. The $\beta X'$ term dropped out of the denominator as full health was the omitted level in each dimension and therefore each X term was zero, meaning the QALY weight for a health state is:

$$\frac{\alpha + \beta X'}{\alpha} \qquad \text{(Equation 3)}$$

In our study, DCE + duration data was estimated using two models: a) an 11 parameter main effects model (i.e. duration and the 10 interactions between duration and level 2 and 3 of each of the five dimensions); b) a model introducing an additional interaction captured by N3 term. Akaike information criteria, Bayesian information criteria and likelihood ratio test was performed to determine the preferred model for the DCE + duration data. All data were analysed using a conditional logit model. The wtp Stata command was used to estimate QALY weights.

Approaches B-D: Different strategies to anchor DCE latent scale value onto the QALY scale

The latent scale DCE data was analysed using choice models under a random utility framework with a linear, additive utility function, with 10 dummy variables representing two levels beyond level 1 (i.e. level 2 and level 3) for each of the five dimensions. A mixed logit model (mixlogit Stata command) (Hole, 2007) was chosen to identify unobserved heterogeneity. In this model, all parameters were modelled as random and uncorrelated using 5000 Halton draws. In addition, we estimated the standard multinomial logit (clogit Stata command) and the latent class models (lclogit2 Stata command). We evaluated different number of classes ranging from 2 to N classes and select the N−1 number of classes conditional on the BIC of the model with N classes being higher than of the model with N−1 classes.

Coefficients from the model were transformed into relative attribute importance (RAI) scores to aid interpretation, which were obtained by dividing the utility range for each attribute by the total utility range (Kreimeier et al., 2022).

The following approaches were used to anchor the coefficients from the mixed logit model onto the QALY scale.

Approach B: A linear mapping model between latent scale DCE coefficients and utility decrements from DCE + duration:

$$\beta_{DCEduration} = f(\beta_{latentDCE}) + \varepsilon \qquad \text{(Equation 4)}$$

This was first explored by plotting the coefficients from latent DCE model and utility decrements generated from the DCE + duration model; and fit into a specification assuming a linear relationship with an intercept. Estimation was undertaken using ordinary least squares and the rescaled coefficients were predicted from the model.

Approach C: Rescale on the 'pits' state (i.e., 33333) utility value from DCE + duration:

$$\beta'_{k\omega} = \left(\frac{\sum_{j=1}^{5} \beta_{j3(DCEduration)}}{\sum_{k=1}^{5} \beta_{k3(latent)}}\right) * \beta_{k\omega} \qquad \text{(Equation 5)}$$

where $\beta'_{k\omega}$ were the rescaled coefficients, $\beta_{k\omega}$ were coefficients for level w from dimension k from latent scale DCE regression model (mixed logit model), i.e., β_{k3} were to the level 3 coefficients. β_{j3} were the level 3 utility decrements generated under Approach A. This equation means that the value of the worst state in the DCE model was anchored at the value of the worst state from the DCE + duration results in Approach A.

Approach D: using dead preference estimated from DCE + dead

This analysis used DCE + dead data and undertook a logit regression using each comparison between a non-dead state and 'dead'. The dependent variable in this regression was whether an individual believed the health state to be better than dead, and the independent variables were the dimensions of the instrument. Because the DCE + dead tasks only involved 32 'severe' health states ranging from 22222 to 33333, 5 parameters (dummy variables for level 3) were included in the model. Using the regression results from logit model, we can identify the relative position of health states and 'dead', and set the utility of the health state with a 50% chance of being preferred to dead as zero. We rescaled the coefficients using the formula $\beta'_{k\omega}$ = $\beta_{k\omega}$/θDead; where θ was the coefficient for the health states set as "dead".

2.5 Comparison of values generated from different approaches

To compare values, we produced descriptive statistics for values of the 243 health states defined by EQ-5D-Y, including the range of the values, the percentage of health states having values < 0 and the number of unique values. We plotted the overall distribution of the values using a kernel density histogram, and displayed the values generated by approaches (A-D) ordered by ascending latent scale value from latent scale DCE data produced using a mixed logit model.

2.6 Subgroup analysis – Preferences of parents and non-parents

Published EQ-5D-Y-3L valuation studies have found differences in the values given to child health states among parents and non-parents. Such differences are found mostly in TTO tasks where parents were less willing to trade HRQOL and length of life (Kreimeier et al., 2022; Matza et al., 2014). Our study also examined whether preferences and values differ by parental status in the three DCE tasks. We split the two samples based on responses to the question "Have you ever been a parent?" Respondents who answered "yes" were classified as having parental status; those answering "no" or "prefer not to say" were classified as "non-parents". We generated and reported values using Approaches A and D for the two groups respectively. This is because the two approaches used mutually exclusive data (i.e., Approach A used DCE + duration as "stand-alone" method; Approach D used latent DCE and anchored on DCE + dead). We ran mixed logit model (using 200 Halton draws) for latent scale DCE tasks, and conditional logit model results for DCE + duration tasks, and logit model for DCE + dead tasks.

3. Results

3.1 Sample Characteristics

In total, data were provided by 1002 respondents completed the DCE survey. Overall, the DCE sample is similar in the distribution of age, gender, education to Australian general population (Table 1).

Table 1 Sample characteristics

	N	%	Population*
Age			
18-29 years	215	21.5%	21.7%
30-39 years	185	18.5%	18.6%
40-49 years	174	17.4%	16.6%
50-59 years	155	15.5%	15.6%
60-69 years	151	15.1%	13.3%
70 + years	122	12.2%	14.2%
Gender			
male	481	48.0%	49.1%
female	516	51.5%	50.9%
diverse	3	0.3%	
not say	2	0.2%	
Ever been a parent			N/A
yes	628	62.7%	
no	367	36.6%	
not say	7	0.7%	
Country of Birth			
Australia	811	80.9%	64.3%
other English speaking countries	103	10.3%	35.7%
non- English speaking countries	88	8.8%	
Education			
Years 11 or below	165	16.5%	24.2%
Year 12	209	20.9%	15.2%
Trade certificate	136	13.6%	5.9%
Diploma	136	13.6%	15.5%
Bachelor's degree	264	26.3%	15.3%
Higher degree	92	9.2%	6.3%
Self-rated health			
Excellent	113	11.3%	10.4%
Very good	293	29.2%	34.6%
Good	374	37.3%	36.5%
Fair	168	16.8%	14.9%
Poor	54	5.4%	3.6%
Chronic condition			N/A
Yes	312	31.1%	
No	690	68.9%	

* Population norms sourced from Australian Bureau of Statistics (ABS) and Norman et al. (2021)

3.2 Modelling DCE +duration data (Approach A)

Table 2 presents the coefficients from the main effect model and the model including the N3 term. In both models, the results reflected the monotonic nature of the EQ-5D-Y. Specifically, all coefficients were negative, and the level 3 coefficients were absolutely larger than the level 2 ones. Both models suggested similar importance ranking: PD had the largest overall decrement (a proxy for importance), followed by AD, MO, UA and SC. The N3 coefficient was negative but not statistically significant at the 10% level; the Likelihood-ratio test rejected the hypothesis that the N3 model outperformed the main effect model, supporting the conclusion of AIC and BIC. The estimated values from the main effect model range from -0.332 to 1.

Table 2 Regression and utility weights using DCE + duration as a stand-alone approach

Model	Main effect			N3 model		
	coefficients	SE	utility decrement	coefficients	SE	utility decrement
Duration	0.911***	0.065		0.950***	0.075	
MO2 * DUR	-0.104***	0.026	-0.114	-0.114***	0.027	-0.120
MO3 * DUR	-0.171***	0.027	-0.187	-0.166***	0.027	-0.175
SC2 * DUR	-0.078***	0.026	-0.085	-0.087***	0.027	-0.091
SC3 * DUR	-0.100***	0.029	-0.110	-0.095***	0.029	-0.100
UA2 * DUR	-0.043*	0.026	-0.047	-0.042	0.026	-0.045
UA3 * DUR	-0.121***	0.026	-0.132	-0.106***	0.029	-0.111
PD2 * DUR	-0.160***	0.029	-0.175	-0.166***	0.029	-0.175
PD3 * DUR	-0.421***	0.03	-0.463	-0.410***	0.032	-0.432
AD2 * DUR	-0.142***	0.025	-0.156	-0.148***	0.026	-0.155
AD3 * DUR	-0.400***	0.033	-0.439	-0.391***	0.034	-0.412
N3 *DUR				-0.053	0.053	-0.056
Observations	6,012			6,012		
Log pseudolikelihood	-1714			-1714		
Wald Chi-square	340.9			341.1		
Pseudo R2	0.177			0.177		
Degree of freedom	11			12		
AIC	3450.826			3451.811		
BIC	3524.543			3532.229		

3.3 Modelling latent scale DCE data

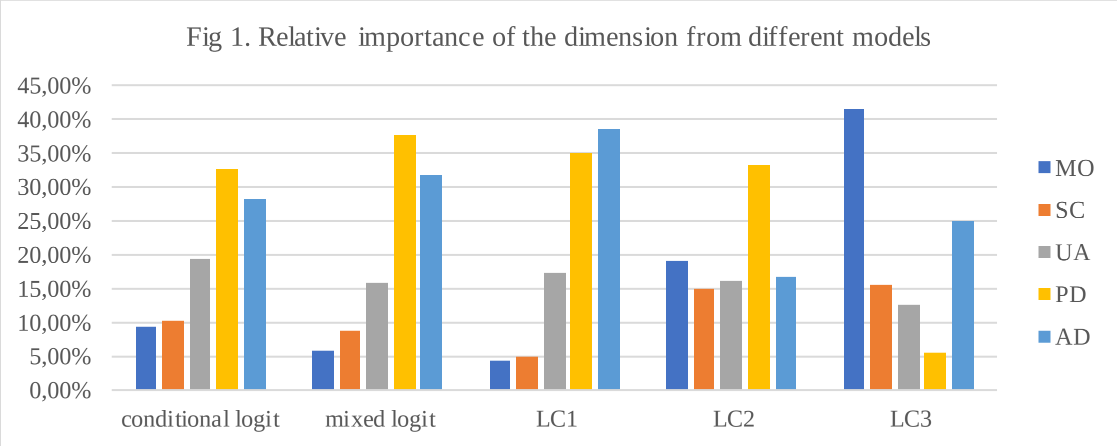
Results from conditional logit model, mixed logit model and latent class analysis are presented in Table 3. In the conditional logit and mixed logit model, all the coefficients were negative and monotonic. Results from the two models suggested similar importance (ordered from most to least important): PD, AD, UA, SC, and MO. However, the results from the mixed logit model suggested that there was considerable heterogeneity in the preferences of respondents with respect to the more severe dimension levels. Of the 10 standard deviations estimated, all the level 3 standard deviations were statistically significant at 5% level, suggesting that the heterogeneity related to level 3 choices. Such results indicated that it was suitable to use the mixed logit model over the conditional logit model in this study, which was also captured by better fit of the data with lower log-likelihood, AIC and BIC.

The latent class analysis confirmed the heterogeneity observed in the mixed logit model. A latent class model with 3 classes exhibited the lowest BIC. Looking at within-class model structure, the first class constituted 42.5% of the sample. They gave most importance (or highest preference to avoid) (38.5%) to AD followed by PD (35.0%), UA (17.2%), SC (5.1%) and MO (4.2%); the second class (23.9% of the sample) focused on the importance in PD (33.1%) and MO (19.0%). Respondents from the third class (33.6% of the sample) did not show strong preferences. The overall relative importance of the five dimensions from different models are presented in Figure 1.

Table 3 Modelling results on latent scale DCE data using different models

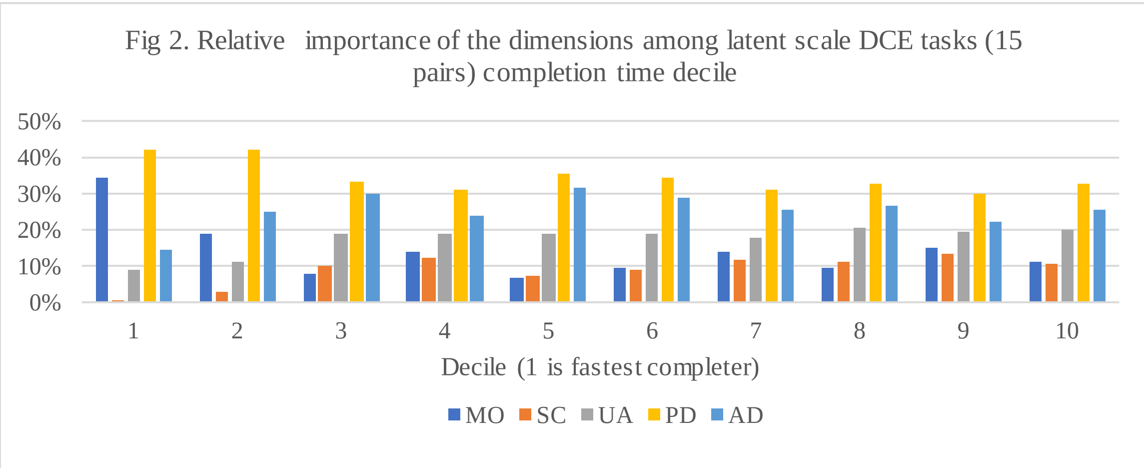
Level and dimensions	conditional logit model	mixed logit model		Latent class analysis		
	coefficients	coefficients	SD	class 1	class 2	class 3
MO2	-0.206***	-0.258***	0.021	0.021	-0.961***	0.001
	(0.05)	(0.07)	(0.03)	(0.28)	(0.16)	(0.08)
MO3	-0.426***	-0.370***	1.214***	-0.57	-1.913***	-0.237**
	(0.07)	(0.11)	(0.09)	(0.44)	(0.25)	(0.12)
SC2	-0.143***	-0.183***	0.01	-0.13	-0.590***	0.014
	(0.04)	(0.05)	(0.01)	(0.16)	(0.14)	(0.07)
SC3	-0.462***	-0.557***	0.320**	-0.678**	-1.492***	-0.089
	(0.06)	(0.08)	(0.13)	(0.27)	(0.22)	(0.09)
UA2	-0.415***	-0.430***	0.006	-0.807***	-0.679***	-0.028
	(0.04)	(0.05)	(0.02)	(0.16)	(0.13)	(0.06)
UA3	-0.869***	-0.998***	0.621***	-2.295***	-1.621***	-0.072
	(0.06)	(0.07)	(0.07)	(0.26)	(0.21)	(0.08)
PD2	-0.682***	-0.796***	0.186	-1.931***	-1.396***	0.02
	(0.04)	(0.06)	(0.15)	(0.19)	(0.18)	(0.06)
PD3	-1.467***	-2.379***	1.768***	-4.672***	-3.329***	-0.031
	(0.08)	(0.13)	(0.10)	(0.36)	(0.34)	(0.09)
AD2	-0.463***	-0.559***	-0.005	-1.697***	-0.670***	0.003
	(0.04)	(0.05)	(0.03)	(0.19)	(0.12)	(0.06)
AD3	-1.264***	-1.997***	1.661***	-5.131***	-1.692***	-0.142*
	(0.06)	(0.10)	(0.08)	(0.45)	(0.19)	(0.08)
class share				0.425	0.239	0.336
Observations	30,060	30,060		30,060		
Log-Likelihood	-8036	-7138		-6633		
Chi-square test	773.2	584.3				
Pseudo R2	0.229					
AIC	16091.55	14316.85		13330		
BIC	16174.66	14483.07		13596		

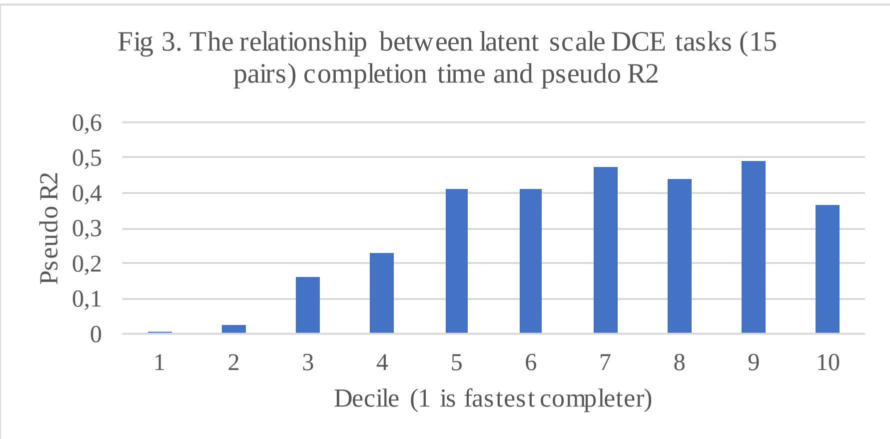
Numbers in parentheses represent standard errors



3.4 Quality assessment

As a quality check, we repeated the analysis using conditional logit models in deciles of participants classified based on latent scale DCE task completion time. The preference for dimensions was different between the quickest 10% of responders and those who spent more time on the tasks (Fig 2). The relationship between completion time decile and pseudo R2 (Fig 3) also showed that those who finished quickest were not engaged with the task, and the predictive value of the model in that sub-group was almost zero. However, in progressively slower deciles, the R2 increased to more than 0.4 for the fifth decile and remained for the following deciles.





3.5 Anchoring methods

Approach B & C: Anchoring using DCE + duration data

The rescaled coefficients (utility decrements) under Approach B and Approach C are presented in Table 4. Under Approach B, plots of latent DCE coefficients from mixed logit model and utility decrements from DCE +duration approach indicated a linear relationship. We regressed DCE +duration utility decrements on latent scale coefficients, and the rescaled coefficients were predicted from the model. Under this approach, the value for the 'pits' state was lower at -0.319. Approach C used the value of 'pits' state generated from Approach A to anchor, therefore, its minimal value was the same as Approach A, at -0.332. It should be noted that the difference between approaches B and C was on the scale length only; the relative importance of dimensions was almost identical.

Approach D: Rescale using DCE + dead

In the DCE + dead tasks, participants considered a health state to be worse than dead (WTD) in about 10% of the cases. The logit model regression results are reported in Table S1 in the Appendix. Because the DCE dead tasks only included 32 severe health states from 22222 to 33333; only level 3 coefficients were estimated, and the probability of a health state being considered as WTD was predicted for the 32 health states. Health state 33333 had an estimated 56.7% chance of considered being WTD, followed by 33233 (52.3%), 32333 (51.5%), and 32233 (47.1%) (Table S2 in the Appendix). If we set 32233 as zero (or equivalent to dead) to anchor, the rescaled value of 33333 was -0.175, and rescaled utility decrements were in Table 4 Column D1. If we set 32333 as zero, then the rescaled value of 33333 was -0.063.

There were some inconsistences of preferences between the DCE + dead tasks and latent DCE tasks. Some health states (e.g. 23333) had values > 0 in the DCE + dead tasks, but had a lower value in the latent scale DCE than a health state (e.g. 32333) which had a value < 0 DCE + dead data. We further regressed latent scale DCE value for each of health state on the predicted probability of being worse than dead for the 32 health states in the DCE + dead task, and then predicted the probability of dead preference for all 243 health states and determined the health state with a 50% chance. In the predicted probability, the 'pits' state has a 50.5% chance of being WTD which was the nearest position to dead. Column D3 in Table 4 shows the rescaled utility decrements when anchoring on dead using the 'pits' state.

3.6 Comparison of values generated from different approaches

Table 4 also presents the characteristics of values generated from different approaches. Values generated from Approaches A and C had the widest range. The minimum value increased when using approach D anchoring on dead preference; and the more 'severe' health states used to anchor, the greater the minimum value. However, it is worth noting that although the overall length of the value scale reduced, the number of unique values was similar across different approaches. The differences in values across approaches can be seen in Figure 4 and Figure 5. Approaches C and D show a similar pattern and follow that of the latent scale values but with different scale length. This was expected because both approaches used latent scale DCE model as base analysis but anchored on different health states. The minimum value was smaller when value sets were generated using data from DCE + duration tasks compared to DCE + dead tasks.

Table 4 Utility decrements and value set characteristics from different approaches

	Approach A: DCE + duration as stand-alone	Utility decrements on Latent scale	Anchoring using DCE + duration data		Anchoring using DCE+ dead data		
			Approach B: Linear mapping between latent scale and DCE duration coefficients	Approach C: anchor on 33333 value from DCE + duration data	Approach D1 Anchoring on dead using 32233	Approach D2 Anchoring on dead using 32333	Approach D3 Anchoring on dead using 33333
MO2	-0.114	-0.258	-0.084	-0.055	-0.048	-0.044	-0.041
MO3	-0.187	-0.370	-0.104	-0.078	-0.069	-0.062	-0.059
SC2	-0.085	-0.183	-0.071	-0.039	-0.034	-0.031	-0.029
SC3	-0.110	-0.557	-0.138	-0.118	-0.104	-0.094	-0.088
UA2	-0.047	-0.430	-0.115	-0.091	-0.08	-0.073	-0.068
UA3	-0.132	-0.998	-0.217	-0.211	-0.186	-0.168	-0.158
PD2	-0.175	-0.796	-0.181	-0.168	-0.148	-0.134	-0.126
PD3	-0.463	-2.379	-0.464	-0.503	-0.444	-0.401	-0.378
AD2	-0.156	-0.559	-0.138	-0.118	-0.104	-0.094	-0.089
AD3	-0.439	-1.997	-0.396	-0.422	-0.372	-0.337	-0.317
range	(-0.332, 1)		(-0.319, 1)	(-0.332, 1)	(-0.175, 1)	(-0.063, 1)	(0, 1)
Unique value	206		210	213	204	208	206
No. of HS WTD	28 (11.5%)		25 (10.3%)	26 (10.7%)	12 (4.9%)	2 (0.8%)	0 (0.0%)
ranking	PD	PD	PD	PD	PD	PD	PD
	AD	AD	AD	AD	AD	AD	AD
	MO	UA	UA	UA	UA	UA	UA
	UA	SC	SC	SC	SC	SC	SC
	SC	MO	MO	MO	MO	MO	MO

Fig 4. Density plots of values generated from different approaches

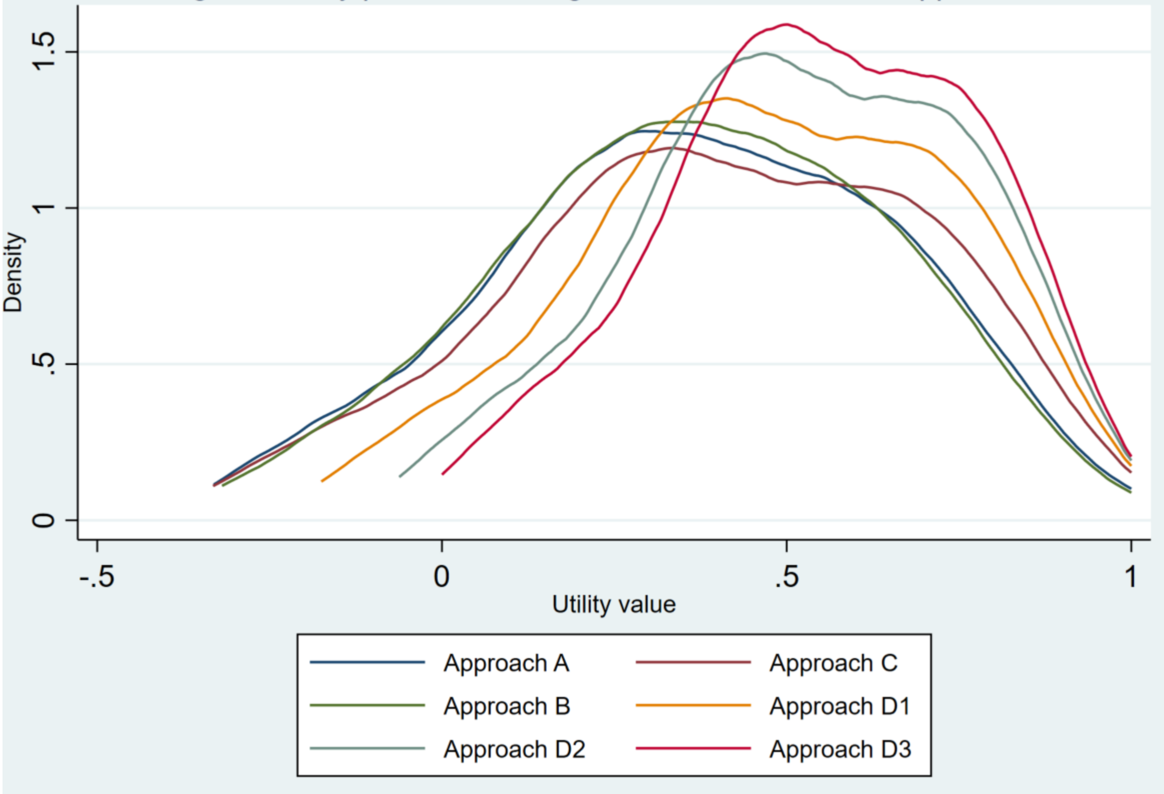
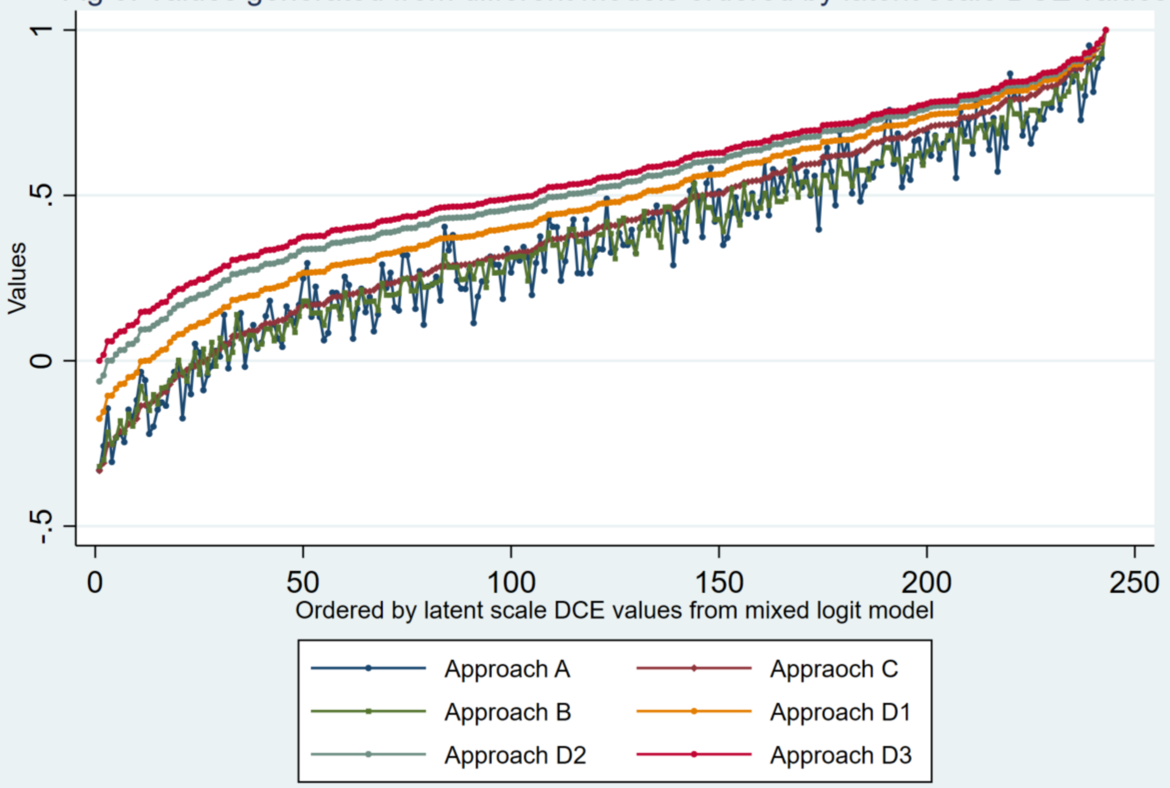


Fig 5. Values generated from different models ordered by latent scale DCE values



3.7 Subgroup analysis: Parental Status

In our sample, 63% reported having ever been a parent and 37% as not having been a parent. The utility decrement for parents and non-parents based on Approaches A and D are reported in Table 5. Under Approach A, results differed substantially between the two groups. For parents, the most important dimension was PD, followed by AD, MO, UA, SC. For non-parents, the most important dimension is AD, followed by PD, MO, SC, UA. However, we observed non-monotonicity in SC among the parents' group. We generated two separate values under Approach A for the two groups. The minimum value (i.e. value for 33333) for the parents group was -0.281, which was higher than that for non-parent group (-0.430), suggesting parents were less willing to trade-off time for improved quality of life than non-parents.

Results on latent scale DCE using mixed logit model showed similar pattern in terms of dimension importance rankings between parents and non-parents: PD, AD, UA, SC and MO. For the parent group, 33333 had a predicted probability of 48.6% as being considered WTD; for non-parent group, health state 32333 had a predicted probability of 49.6% which was used to anchor the latent DCE value. Regression results are reported in Tables 3-5 in the Appendix. In general, non-parents were more likely to consider a severe health state as worse than dead.

Table 5 Utility decrements and value set characteristics under Approach A and Approach D, by parental status

	Approach A		Latent scale		Approach D	
	parents	Non-parents	parents	Non-parents	parents	Non-parents
MO2	0.114	0.131	0.209	0.333	0.031	0.063
MO3	0.202	0.171	0.417	0.382	0.062	0.072
SC2	0.083	0.084	0.174	0.208	0.026	0.039
SC3	0.076	0.169	0.557	0.580	0.083	0.110
UA2	0.067	0.013	0.403	0.474	0.060	0.090
UA3	0.156	0.083	1.04	0.92	0.155	0.174
PD2	0.169	0.188	0.912	0.639	0.136	0.121
PD3	0.445	0.489	2.654	1.965	0.396	0.372
AD2	0.143	0.185	0.566	0.535	0.084	0.101
AD3	0.401	0.519	2.037	1.812	0.304	0.343
range	(-0.281, 1)	(-0.430, 1)			(0,1)	(-0.070, 1)
ranking	PD	AD	PD	PD	PD	PD
	AD	PD	AD	AD	AD	AD
	MO	MO	UA	UA	UA	UA
	UA	SC	SC	SC	SC	SC
	SC	UA	MO	MO	MO	MO

4. Discussion

In this study, we explored four DCE-based approaches to valuing EQ-5D-Y. Approach A used DCE + duration data as a stand-alone method. Approaches B, C and D applied different anchoring strategies, by using DCE + duration or DCE + dead to anchor latent scale DCE values onto a QALY scale. Approach C anchors on the 'pits' state whereas approach D anchors on 'dead'. Our results suggested that it is feasible to use DCE-based approach to valuing EQ-5D-Y. However, the different approaches have impacts on relative importance of dimensions and level decrements, and resultant value set characteristics.

The dimension importance ranking is different under Approach A and other approaches. In the DCE + duration tasks, participants gave more importance to PD, followed by AD, MO, UA and SC. The other approaches used latent scale DCE as the base analysis, so these dimension rankings were determined by latent scale DCE data and the chosen model. The main difference between Approach A and others was the position of mobility, which was considered as the least important dimension in the latent scale DCE tasks. It appears that adults' preferences regarding how problematic the five dimensions of HRQoL are for a 10-year-old child changed when a time frame is considered.

The latent scale DCE component in our study follows the international valuation protocol which enables comparison with the four EQ-5D-Y-3L valuation studies published to date. The ordering of the importance of dimensions from the latent scale DCE is the same as that observed in the German EQ-5D-Y-3L valuation study. In Japan and Slovenia, PD, AD and UA have also been considered as the three most important dimensions. These three countries used mixed logit model to model latent DCE data. The pattern whereby heterogeneity is primarily seen in the level 3 coefficients has also been observed in the German study whereas the Slovenian study observed heterogeneity across the levels. In Spanish value sets, PD and AD were the two most important dimensions, followed by MO, UA, SC. Their chosen model to analysis latent DCE data was latent class model with four classes.

The scale length under each of the four approaches in our studies varies. DCE +duration methods produced the largest range of values in the resultant value set, comparing to using DCE + dead to anchor. This was in line with previous studies that explored different DCE-based approaches to valuing adult health states (Norman et al., 2013; Norman et al., 2016a). When comparing with the Australia EQ-5D-3L and EQ-5D-5L value sets which used a similar approach (i.e. Approach A), it is notable that EQ-5D-Y-3L value sets have a smaller range. For example, the minimal value for EQ-5D-5L is -0.676 and for EQ-5D-3L is -0.516, compared with the minimum value we find of -0.332 for EQ-5D-Y-3L. Although both used DCE + duration as a stand-alone approach, the specific choice tasks, experimental design and modelling methods are different to our study. Particularly, both adult EQ-5D studies used DCE + duration triplets (i.e., the third scenario was immediate death) although this approach usually reduces the scale in comparison to pairs. In an Australian study that explored the same anchoring approaches using dead preference (i.e. our Approach D), the minimum value for EQ-5D-3L is -0.322. Direct comparison of scale length with published EQ-5D-Y-3L data is limited because the different valuation methods and anchoring approaches used, in addition to cultural and context difference. Nevertheless, our finding that value range for EQ-5D-Y-3L is smaller than EQ-5D adult instrument is in line with other valuation studies (Kreimeier et al., 2022; Kreimeier et al., 2018; Shah et al., 2020). It is possible that the underlying factors that lead to higher TTO values for EQ-5D-Y-3L (unwillingness to trade off life years concerning children) also affect responses to DCE +duration concerning child HRQoL.

Among the four approaches, approaches C and D used different anchoring approach. Approach C anchors on the 'pits' state value from DCE +duration, while approaches D anchors on dead. Published EQ-5D-Y-3L studies have used different anchoring methods based on cTTO values, such as anchoring on the mean cTTO value of the worst state (i.e., 33333) or mapping DCE data onto the cTTO data. It is feasible to anchor latent value on the 'pits' state, as shown in Spanish value set (Ramos-Goñi et al., 2021). However, in a recent EQ-5D-Y-3L valuation workshop, there was consensus that it is preferable to avoid using cTTO data only for the purpose of estimating a value for the worst health state, because using all available cTTO data allows for more precise estimates of values for state 33333 and adjacent states. In our approach C, we used a different source of estimates for 33333. Although it is feasible to use DCE + duration to anchor, it raises the question of the purpose of collecting DCE +duration data, and its role in anchoring, or as a standalone approach to valuation. The evidence suggests that it can provide logical, consistent and comparable values as a standalone approach for the EQ-5D-Y-3L (including the perspective specified by the EuroQol valuation protocol). Further work to compare the DCE + duration values and those from the cTTO data that will be collected in Australia in late 2022 will be informative.

Under Approach D, the preferences regarding health states being better or worse than dead were used to anchor. Similar methods have been used to explore EQ-5D-Y-3L and EQ-5D adult instruments in previous methodological studies (Norman et al., 2016a; Roudijk et al., 2018; Shah et al., 2020). Comparing results with these studies, we found that dead was located lower in the descriptive system in the EQ-5D-Y-3L compared to the adult 3L, resulting in narrower range of rescaled values. In our DCE + dead task design, we only included 32 more severe health states, given previous evidence that respondents located dead amongst more severe health states when using child HRQOL instrument (Shah et al., 2020). However, by only including 32 severe health states, we cannot directly rank all the 243 health states and their position relative to dead. This can potentially lead to biased results and have impacts on results of cost–utility studies. A previous study has shown that the position of dead relative to other health states weighs the importance of life-saving and life-improving values, which impact on decision making (Roudijk et al., 2018). As illustrated in Columns D1-D3 in Table 5, the choice of the health state used to anchor dead strongly determines the scale length and the proportion of health states considered worse than dead. Another methodological challenge from this approach is that it lacks a means of identifying the position of dead when respondents locate dead *below* 33333 and therefore beyond the descriptive system. In our subgroup analysis, we found that parents were less likely to consider the worst health state to be worse than dead (e.g. 33333 had a predicted probability of 48.6% as being considered WTD among parents). As we still used 33333 to anchor under approach D, we therefore underestimated its actual value. The rescaled value for 33333 should be positive among parents.

One methodological challenge when anchoring latent scale data is the consistency of preferences across different methods and tasks. This may arguably be more of an issue when using different elicitation methods (e.g. TTO, VAS) other than DCE to anchor DCE latent scale data. Some EQ-5D-Y-3L valuation studies that have completed or underway observed different preferences elicited from DCE and cTTO tasks. However, in our analysis, regression results also showed that the three DCE tasks produce different preference in terms of dimension ranking.

This study also examined whether preference and values differ by parental status, using Approach A and D. Under both approaches, the minimal value was higher among parents compared to non-parents; whereas the relative importance of dimension based on DCE latent scale results were the same among parents and non-parents. Such findings suggest parents were less willing to trade-off time for improved quality of life than non-parents. This is consistent with existing studies. For example, the German EQ-5D-Y-3L study found that the mean cTTO values were greater for parents than for non-parents but the DCE latent scale results did not differ substantially. These suggest that survival and length of life may play a key role in determining preferences for child HRQoL.

A key issue when considering these uses of DCE beyond the role it currently plays in the EQ-5D-Y-3L protocol at present is whether it is acceptable to HTA bodies. The TTO is widely accepted as reasonable and is central to most EQ-5D value sets used in practice. In the development of HTA models, it is important that there is consistency and face validity of utility scores both within models, and between models. Given that most of the existing data centres on TTO-derived weights, this is an ongoing challenge which will require both methodological analysis and outreach from the EuroQol membership if DCE methods are to be accepted more widely in HTA bodies.

There are a number of limitations in this paper. First, online data collection may be limited by data quality issues that cannot easily be controlled by the researcher. However, we did not apply exclusion criteria on data quality. Instead, we conducted several quality checks to explore to the extent that participants provide poorer or random data. In the full sample 32% of respondents reported that they found the tasks difficult or very difficult (Table 6 in Appendix). These participants may contribute to the variance and reduce precision in parameter estimates. We find different preferences for dimensions based on latent scale DCE tasks among the 10% fastest completing respondents and those who were progressively slower, suggesting they may not engage in the tasks (Figures 2 and 3). However, by excluding this group of speeders, the results were qualitatively similar to main analysis (Results available upon request). Coefficients from the three models for the three DCE tasks (i.e., mixed logit model for latent scale DCE, conditional logit model for DCE + duration data and logit model for DCE +dead) were larger compared to the results on full sample. Model performance improved when excluding the fastest speeders. Nevertheless, the domain importance ranking was similar compared to main analysis. The coefficients on standard deviations from mixed logit model on latent scale DCE data were fairly similar to that from the main sample, suggesting that the heterogeneity in preferences among respondents was not purely attributable to speeding behaviour. Another limitation on data quality is that the internal consistency across the three DCE tasks have not been examined. Second, as discussed earlier, we only included 32 severe health states in the DCE + dead tasks; the relative position of all 243 health states to dead therefore cannot be directly estimated. Third, we did not explore further how different modelling methods impact on generated values under the same approach. For example, we chose the mixed logit model to estimate DCE latent scale value, although latent class analysis is also feasible, and both can control unobserved heterogeneity. Under Approach B, we used a simple OLS mapping approach between the 10 coefficients; the model performance was not further examined. Fourth, in the subgroup analysis by parental status, we defined parental status based on the question whether the participant has ever been a parent. Our parent group may include 70-year-old parent of 50 year old as well as people who is currently parent of a child at younger age (e.g. <18 years old). The aspect of parenthood that most affects values may be largely from those who are current parents of young children. Future research may explore further the key drivers of the differences in preference between parents and non-parents and what aspect of parenthood most affects values.

As with all research to explore new methods for valuing HRQoL, we lack a clear basis for determining when one set of values is 'better' than another in the absence of a gold standard. The extended DCE tasks involving duration and dead in the choice sets clearly show these to be feasible to use to create values for the EQ-5D-Y. DCE + duration is a valid method based on our results, and it is also feasible to anchor use DCE + duration.

The DCE + dead task is more experimental but does seem to have promise. The DCE data were obtained as part of the ongoing EQ-5D-Y-3L valuation study in Australia. The cTTO data are yet to be collected. Therefore, at this point we are unable to directly compare the characteristics of the alternative DCE values reported here with the corresponding TTO data from the Australian general public. Nevertheless, the characteristics of the values are plausible, when judged against other EQ-5D-Y-3L value sets. When compared against adult 5L valuations, generated using the same method from the Australian general public, it appears the value ranges are smaller. This suggests that any method for anchoring EQ-5D-Y-3L at dead = 0 might encounter the same issue as TTO with respect to the importance placed on length of life in children – at least to some degree. We presented different DCE based approaches; however, the criteria for choosing between DCE-based methods, and more broadly, among different elicitation methods and anchoring methods, have not been discussed. The evidence from this study suggests that the four approaches are feasible, but many improvements and alterations could be made. In follow up work, we will assess DCE + duration and DCE + dead relative to the cTTO data soon to be collected. Further research is also needed to identify a preferred method, or criteria to identify a preferred method.

Reference

Bailey C, Howell M, Raghunandan R, Salisbury A, Chen G, Coast J, Craig JC, Devlin NJ, Huynh E, Lancsar E, Mulhern BJ, Norman R, Petrou S, Ratcliffe J, Street DJ, Howard K, Viney R, Dalziel K, Hiscock H, Hayes A, Wong G, Donaldson C, Carter S, the Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia TfORtMVCHPT. Preference Elicitation Techniques Used in Valuing Children's Health-Related Quality-of-Life: A Systematic Review. PharmacoEconomics 2022.

Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ 2012;31; 306-318.

Craig BM, Greiner W, Brown DS, Reeve BB. Valuation of Child Health-Related Quality of Life in the United States. Health Economics 2016;25; 768-777.

Devlin N, Pan T, Kreimeier S, Verstraete J, Stolk E, Rand K, Herdman M. Valuing EQ-5D-Y: the current state of play. Health and Quality of Life Outcomes 2022 *(forthcoming)*.

Hole AR. Fitting Mixed Logit Models by Using Maximum Simulated Likelihood. The Stata Journal 2007;7; 388-401.

Jonker MF, Roudijk B, Maas M. The Sensitivity and Specificity of Repeated and Dominant Choice Tasks in Discrete Choice Experiments. Value in Health 2022.

Kreimeier S, Mott D, Ludwig K, Greiner W, Prevolnik Rupel V, Ramos-Goñi JM, Group IHH. EQ-5D-Y Value Set for Germany. PharmacoEconomics 2022.

Kreimeier S, Oppe M, Ramos-Goñi JM, Cole A, Devlin N, Herdman M, Mulhern B, Shah KK, Stolk E, Rivero-Arias O, Greiner W. Valuation of EuroQol Five-Dimensional Questionnaire, Youth Version (EQ-5D-Y) and EuroQol Five-Dimensional Questionnaire, Three-Level Version (EQ-5D-3L) Health States: The Impact of Wording and Perspective. Value in Health 2018;21; 1291-1298.

Matza LS, Boye KS, Feeny DH, Johnston JA, Bowman L, Jordan JB. Impact of caregiver and parenting status on time trade-off and standard gamble utility scores for health state descriptions. Health and Quality of Life Outcomes 2014;12; 48.

Mulhern B, Norman R, Street DJ, Viney R. One Method, Many Methodological Choices: A Structured Review of Discrete-Choice Experiments for Health State Valuation. Pharmacoeconomics 2019;37; 29-43.

Norman R, Cronin P, Viney R. A pilot discrete choice experiment to explore preferences for EQ-5D-5L health states. Applied health economics and health policy 2013;11; 287-298.

Norman R, Mulhern B, Viney R. The Impact of Different DCE-Based Approaches When Anchoring Utility Scores. Pharmacoeconomics 2016a;34; 805-814.

Norman R, Robinson S, Dickinson H, Williams I, Meshcheriakova E, Manipis K, Anstey M. Public Preferences for Allocating Ventilators in an Intensive Care Unit: A Discrete Choice Experiment. The Patient - Patient-Centered Outcomes Research 2021;14; 319-330.

Norman R, Viney R, Aaronson NK, Brazier JE, Cella D, Costa DS, Fayers PM, Kemmler G, Peacock S, Pickard AS, Rowen D, Street DJ, Velikova G, Young TA, King MT. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation 2016b;25; 637-649.

Prevolnik Rupel V, Ogorevc M. EQ-5D-Y Value Set for Slovenia. Pharmacoeconomics 2021;39; 463-471.

Pullenayegum E, Stolk E, Jakubczyck M, Roudijk B, Viney R, Gu NY, Yang Y, Norman R. 2020. Using Discrete Choice Experiments as a stand-alone approach to valuation: a draft protocol. (Ed)^(Eds), 37th EuroQol Plenary Meeting. 2020.

Ramos-Goñi JM, Oppe M, Estévez-Carrillo A, Rivero-Arias O, Wolfgang G, Simone K, Kristina L, Valentina R. Accounting for Unobservable Preference Heterogeneity and Evaluating Alternative Anchoring Approaches to Estimate Country-Specific EQ-5D-Y Value Sets: A Case Study Using Spanish Preference Data. Value in Health 2021.

Ramos-Goñi JM, Oppe M, Stolk E, Shah K, Kreimeier S, Rivero-Arias O, Devlin N. International Valuation Protocol for the EQ-5D-Y-3L. PharmacoEconomics 2020;38; 653-663.

Reckers-Droog V, Karimi M, Lipman S, Verstraete J. Why Do Adults Value EQ-5D-Y-3L Health States Differently for Themselves Than for Children and Adolescents: A Think-Aloud Study. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2022.

Roudijk B, Donders ART, Stalmeier PFM. Setting Dead at Zero: Applying Scale Properties to the QALY Model. Medical decision making : an international journal of the Society for Medical Decision Making 2018;38; 627-634.

Roudijk B, Donders ART, Stalmeier PFM. A Head-On Ordinal Comparison of the Composite Time Trade-Off and the Better-Than-Dead Method. Value in Health 2020;23; 236-241.

Rowen D, Brazier J, Van Hout B. A Comparison of Methods for Converting DCE Values onto the Full Health-Dead QALY Scale. Medical Decision Making 2014;35; 328-340.

Rowen D, Mulhern B, Stevens K, Vermaire JH. Estimating a Dutch Value Set for the Pediatric Preference-Based CHU9D Using a Discrete Choice Experiment with Duration. Value in Health 2018;21; 1234-1242.

Shah KK, Ramos-Goñi JM, Kreimeier S, Devlin NJ. An exploration of methods for obtaining 0 = dead anchors for latent scale EQ-5D-Y values. The European Journal of Health Economics 2020;21; 1091-1103.

Shiroiwa T, Ikeda S, Noto S, Fukuda T, Stolk E. Valuation Survey of EQ-5D-Y Based on the International Common Protocol: Development of a Value Set in Japan. Medical decision making : an international journal of the Society for Medical Decision Making 2021;41; 597-606.

Stolk EA, Oppe M, Scalone L, Krabbe PFM. Discrete Choice Modeling for the Quantification of Health States: The Case of the EQ-5D. Value in Health 2010;13; 1005-1013.

Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, Street D. An Australian discrete choice experiment to value eq-5d health states. Health Econ 2014;23; 729-742.

-

Appendix: Results from analysis DCE + dead data

Table S 1 Modelling results of DCE + dead data using logit model

VARIABLES	logit
3.MO	0.449***
	0.082
3.SC	0.208**
	0.082
3.UA	0.178**
	0.084
3.PD	0.987***
	0.095
3.AD	0.643***
	0.087
Constant	-2.197***
	0.131
Observations	3,006
Log pseudolikelihood	-1700
Wald Chi-square	176.7
Prob > chi2	0
Pseudo R2	0.0644
Degree of freedom	6
AIC	3412.702
BIC	3448.753

Table S2 Predicted probability of a health state being considered worse than dead (WTD) for the 32 severe health states

MO	SC	UA	PD	AD	DCE latent scale value	Predicted probability of WTD from logit model for the 32 health states	Predicted probability of WTD, when regressing latent scale DCE values for each state on the p(WTD) in the DCE dead task, and predicting p(WTD) for all 243 states
3	3	3	3	3	-6.301	0.567	0.505
3	3	2	3	3	-5.733	0.523	0.446
3	2	3	3	3	-5.927	0.515	0.466
3	2	2	3	3	-5.359	0.471	0.407
2	3	3	3	3	-6.189	0.455	0.493
2	3	2	3	3	-5.621	0.411	0.434
3	3	3	3	2	-4.863	0.407	0.355
2	2	3	3	3	-5.815	0.404	0.454
3	3	2	3	2	-4.295	0.365	0.296
2	2	2	3	3	-5.247	0.362	0.395
3	2	3	3	2	-4.489	0.358	0.316
3	3	3	2	3	-4.718	0.328	0.340
3	2	2	3	2	-3.921	0.318	0.257
2	3	3	3	2	-4.751	0.305	0.344
3	3	2	2	3	-4.150	0.290	0.281
3	2	3	2	3	-4.344	0.284	0.301
2	3	2	3	2	-4.183	0.269	0.284
2	2	3	3	2	-4.377	0.263	0.305
3	2	2	2	3	-3.776	0.249	0.242
2	3	3	2	3	-4.606	0.237	0.329
2	2	2	3	2	-3.809	0.230	0.245
2	3	2	2	3	-4.038	0.207	0.269
3	3	3	2	2	-3.28	0.204	0.190
2	2	3	2	3	-4.232	0.202	0.290
3	3	2	2	2	-2.712	0.177	0.131
2	2	2	2	3	-3.664	0.175	0.230
3	2	3	2	2	-2.906	0.172	0.151
3	2	2	2	2	-2.338	0.148	0.092
2	3	3	2	2	-3.168	0.141	0.179
2	3	2	2	2	-2.600	0.120	0.120
2	2	3	2	2	-2.794	0.117	0.140
2	2	2	2	2	-2.226	0.100	0.81

Appendix: Subgroup analysis results: by Parental status

Table S3 regression results of DCE + duration data, by parental status

VARIABLES	Parents	Non-parents
Duration	1.101***	0.686***
SE	0.092	0.091
MO2 * DUR	-0.125***	-0.090**
SE	0.034	0.039
MO3 * DUR	-0.222***	-0.117***
SE	0.036	0.041
SC2 * DUR	-0.092***	-0.058
SE	0.035	0.038
SC3 * DUR	-0.084**	-0.116***
SE	0.040	0.042
UA2 * DUR	-0.074**	-0.009
SE	0.036	0.039
UA3* DUR	-0.172***	-0.057
SE	0.034	0.040
PD2 * DUR	-0.187***	-0.129***
SE	0.039	0.044
PD2 * DUR	-0.490***	-0.335***
SE	0.041	0.044
AD2 * DUR	-0.158***	-0.127***
SE	0.034	0.037
AD3 * DUR	-0.442***	-0.356***
SE	0.046	0.048

Table S4. regression results of latent scale DCE data using mixed logit model (using 200 Halton draws), by parental status

	Parents		Non-parents	
MO2	-0.209**	-0.050	-0.333***	-0.161
	0.091	0.101	0.114	0.130
MO3	-0.417***	1.144***	-0.382**	1.306***
	0.138	0.122	0.180	0.159
SC2	-0.174**	-0.018	-0.208**	-0.057
	0.068	0.071	0.087	0.058
SC3	-0.557***	0.368**	-0.580***	0.315**
	0.099	0.149	0.120	0.139
UA2	-0.403***	-0.061	-0.474***	-0.068
	0.061	0.064	0.076	0.106
UA3	-1.040***	-0.641***	-0.920***	0.573***
	0.093	0.110	0.117	0.123
PD2	-0.912***	-0.227	-0.639***	0.122
	0.074	0.191	0.086	0.112
PD3	-2.654***	1.923***	-1.965***	1.603***
	0.171	0.154	0.178	0.125
AD2	-0.566***	-0.009	-0.535***	-0.127
	0.066	0.109	0.077	0.115
AD3	-2.037***	1.621***	-1.812***	1.687***
	0.125	0.114	0.157	0.123

Table S5 Regression results of DCE data data using logit model, by parental status

	Parents	Non-parents
3.MO	0.453***	0.429***
	0.105	0.134
3.SC	0.192*	0.230*
	0.103	0.138
3.UA	0.188*	0.164
	0.106	0.139
3.PD	0.991***	0.991***
	0.124	0.150
3.AD	0.583***	0.742***
	0.114	0.136
Constant	-2.240***	-2.127***
	0.166	0.215

Appendix: Quality assessment

Table S6 Respondents' feedback

clearness		
very unclear	39	3.9%
unclear	35	3.5%
neither	123	12.3%
clear	418	41.7%
very clear	387	38.6%
difficulty		
very difficult	91	9.1%
difficult	230	23.0%
neither	244	24.4%
easy	269	26.8%
very easy	168	16.8%