



# Analytical Considerations When Anchoring Discrete Choice Experiment Values Using Composite Time Trade-Off Data: The Case of EQ-5D-Y-3L

David J. Mott<sup>1</sup> · Nancy J. Devlin<sup>2</sup> · Simone Kreimeier<sup>3</sup> · Richard Norman<sup>4</sup> · Koonal K. Shah<sup>5</sup> · Oliver Rivero-Arias<sup>6</sup>

Accepted: 26 October 2022 / Published online: 18 November 2022  
© The Author(s) 2022

## Abstract

Discrete choice experiments (DCEs) are becoming increasingly used to elicit preferences for children's health states. However, DCE data need to be anchored to produce value sets, and composite time trade-off (cTTO) data are typically used in the context of EQ-5D-Y-3L valuation. The objective of this paper is to compare different anchoring methods, summarise the characteristics of the value sets they produce, and outline key considerations for analysts. Three anchoring methods were compared using data from published studies: (1) rescaling using the mean value for the worst health state; (2) linear mapping; and (3) hybrid modelling. The worst state rescaling value set had the largest range. The worst state rescaling and linear mapping value sets preserved the relative importance of the dimensions from the DCE, whereas the hybrid model value set did not. Overall, the predicted values from the hybrid model value set were more closely aligned with the cTTO values. These findings are relatively generalisable. Deciding upon which anchoring approach to use is challenging, as there are numerous considerations. Where cTTO data are collected for more than one health state, anchoring on the worst health state will arguably be suboptimal. However, the final choice of approach may require value judgements to be made. Researchers should seek input from relevant stakeholders when commencing valuation studies to help guide decisions and should clearly set out their rationale for their preferred anchoring approach in study outputs.

## 1 Introduction

Health state utilities are required to estimate quality-adjusted life years (QALYs) for use in health technology assessment (HTA). These are normally obtained using concise generic measures, such as EQ-5D-5L [1], which

have value sets that provide a utility for any health state described. Value sets are derived using stated preference methods, typically conducted with members of the general population [2].

Traditional methods, such as time trade-off (TTO), provide data on an interval scale anchored at full health (1) and dead (0) as required for estimating QALYs. However, these methods have been criticised due to their difficulty and associated biases [2, 3]. Further issues may arise with these methods when valuing health states for children (aged < 18 years) because in this context adults may be asked to value health in someone else (a child) [4, 5]. These issues include difficulty in making decisions in tasks [6], as well as emerging evidence in some studies that adults are less willing to trade off life years when asked to complete TTO tasks concerning child health [7–10]. The latter has raised broader concerns about the comparability between child QALYs and adult QALYs, which may cause issues in the context of HTA [11].

In recent years, ordinal methods such as discrete choice experiments (DCEs) have had an increasing role in health

✉ David J. Mott  
dmott@ohe.org

<sup>1</sup> Office of Health Economics, London, UK

<sup>2</sup> Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

<sup>3</sup> Department of Health Economics and Health Care Management, School of Public Health, Bielefeld University, Bielefeld, Germany

<sup>4</sup> School of Public Health, Curtin University, Perth, Australia

<sup>5</sup> National Institute of Health and Care Excellence, London, UK

<sup>6</sup> National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

### Key Points

The EQ-5D-Y-3L valuation protocol states that composite time trade-off (cTTO) data should be used to anchor discrete choice experiment (DCE) data onto the full health to dead scale but does not specify which method should be used to do so.

There are several different anchoring approaches that can be used and the final decision on an approach is likely to require some value judgements to be made. This paper sets out the key considerations for analysts when deciding upon an anchoring approach.

Researchers should seek input from relevant stakeholders when commencing valuation studies to help guide decisions and should clearly set out their rationale for their preferred anchoring approach in study outputs.

valuation [12]. In the context of child health valuation, DCEs focused on pairwise choices between health states defined by EQ-5D-Y-3L, without a duration attribute, have been seen as particularly useful as they can provide information about the relative importance of different dimensions of health, without any need for respondents to trade-off years of a child's life. However, unlike DCEs that include a duration attribute, it is necessary for further data to be collected and used to anchor the DCE data onto the required scale. Various methods have been used to generate data for anchoring, including TTO, visual analogue scale (VAS), and the 'location of dead' (LOD) exercise (a component of the 'personal utility function' [PUF] approach) [13–16].

Following an extensive programme of research, a valuation protocol was developed for EQ-5D-Y-3L [17]. The protocol recommends that EQ-5D-Y-3L is valued using a DCE, and that the DCE data are anchored using data from a separate composite TTO (cTTO) task. The aim of this initial protocol was to enable value sets for EQ-5D-Y-3L to be developed, and it has been successful in achieving that aim [18–24]. However, the protocol was only considered to be a starting point, with further updates expected in future due to the various ongoing normative and methodological debates in the context of child health valuation [4, 25]. These debates include, but are not limited to, the source of preference data [26, 27], the perspective taken in the task [6, 7, 9, 28], and the choice of valuation method [29].

A further methodological debate relates to the optimal approach for anchoring DCE data using cTTO data. The current valuation protocol does not recommend a specific approach. Furthermore, the protocol is not prescriptive regarding the amount of cTTO data collected, only recommending a minimum of ten health states and 200 observations. Resultingly, in practice, EQ-5D-Y-3L valuation studies have utilised different anchoring approaches, and several have included more than ten health states in the cTTO task.

Approaches which can in principle be taken to anchor DCE data using cTTO data, from rudimentary to more complex, include (1) anchoring on the worst health state; (2) mapping DCE onto mean cTTO values; and (3) hybrid modelling. Whilst these three approaches have been subject to a statistical comparison in the past [30], there has been relatively little discussion around how the methods are employed in practice, and how resulting value sets may differ in their characteristics. As many EQ-5D-Y-3L valuation studies are ongoing [11], research teams will need to consider how to anchor their DCE data. Furthermore, valuation studies for other instruments may face similar issues [31].

This paper aims to address the gap in the literature by explaining the different approaches, illustrating their use using previously collected EQ-5D-Y-3L valuation data [26, 32], and discussing the implications of using each approach for the characteristics of the resulting value sets.

## 2 Methods

### 2.1 Overview of Anchoring Approaches Using cTTO Data

#### 2.1.1 Anchoring on the Worst Health State

This approach requires only the mean value of the worst health state in the descriptive system which, for the EQ-5D-Y-3L instrument, is denoted as 33333, reflecting that each dimension is at the worst level (i.e., level 3). This value is used to create a rescaling parameter,  $\theta$ , using Eq. 1:

$$\theta = \frac{\{U_{DCE}(WS)-1\}}{\{\bar{U}_{TTO}(WS)-1\}} \quad (1)$$

where  $\bar{U}_{TTO}(WS)$  is the mean value of the worst state in the descriptive system from the cTTO task, and  $U_{DCE}(WS)$  is the latent scale value for the worst state in the descriptive system. If regular dummy coding is used, with the best level as the base level for each dimension,  $U_{DCE}(WS)$  is one plus the sum of the coefficients for the worst level on each dimension. To rescale the DCE data, the coefficients must subsequently be divided by  $\theta$ .

### 2.1.2 Mapping DCE onto Mean TTO Values

The simplest mapping approach [30] estimates a rescaling parameter,  $\theta$ , based on the relationship between the observed mean TTO values and the unanchored DCE values using a linear ordinary least squares regression model as in Equation 2:

$$\bar{U}_{TTO}(j) = \theta U_{DCE}(j) + \varepsilon \quad (2)$$

where  $\bar{U}_{TTO}(j)$  is the mean value from the TTO task for all  $j$  health states valued, and  $U_{DCE}(j)$  is the latent scale value for each of the  $j$  health states valued in the cTTO task based on the DCE coefficients. The DCE latent scale coefficients are then multiplied by the rescaling parameter,  $\theta$  (estimated in the mapping model), to produce the value set.

For simplicity, and to enable direct comparisons with the hybrid model, we focus on linear mapping without a constant. However, different specifications can be tested to improve the model fit, such as the inclusion of a constant and/or non-linear terms. Additionally, if sufficient cTTO data are collected such that they can be modelled independently, the predicted cTTO values for every health state in the system (i.e.,  $j = 243$  for EQ-5D-Y-3L) could be included in the model instead.

### 2.1.3 Combining DCE and TTO Data in a Hybrid Model

An alternative anchoring approach that takes into account all of the individual-level DCE and cTTO observations is a hybrid model. This approach has since been employed in many EQ-5D-5L valuation studies and is described in detail elsewhere [30, 33, 34]. Briefly, the hybrid model works under the assumption that the same value function dictates the responses to both DCE and cTTO tasks and therefore the coefficients from DCE and cTTO models should reflect the same relative weights for the dimensions and levels of the instrument being valued. The hybrid model multiplies the likelihood functions obtained through estimation of separate DCE and cTTO models, and uses a rescaling parameter,  $\theta$ , based on the assumption that the DCE model coefficients are proportional to the TTO model coefficients. Thus, the model output is the final anchored value set. User-written commands are available for commonly used software packages that enable analysts to employ this method (*hyreg* in both Stata and R) [35].

## 2.2 Data Sources

To illustrate the different anchoring approaches, and compare the characteristics of the resulting values, previously collected data were used. The DCE data were the adult responses from a UK-based EQ-5D-Y-3L study which compared the preferences of adults and adolescents (aged 11–17 years old) [26]. The cTTO data were from a multinational study which examined the impact of wording and perspective on EQ-5D-Y-3L

valuations [32]. The DCE design is the same as that in the EQ-5D-Y-3L valuation protocol [17]. One-thousand adults completed the DCE survey. The cTTO data were from a composite cTTO task where 17 health states were valued, split into two blocks of nine (both containing 33333). There were four arms in the study, and this analysis uses only the data from the EQ-5D-Y-3L child perspective arm to align with the DCE data. There were 211 respondents in this arm, split almost equally across four countries: England, Germany, Spain, and the Netherlands. Further detail about the data sources can be found in the electronic Supplementary Materials (ESM).

## 2.3 Data Analysis

Two of the anchoring approaches require a latent scale choice model to be estimated. A random utility framework was employed, and a conditional logit model was estimated with a linear, additive utility function, as in Eq. 3:

$$\begin{aligned} V_j = & \beta_1 MO2 + \beta_2 MO3 + \beta_3 LAM2 + \beta_4 LAM3 \\ & + \beta_5 UA2 + \beta_6 UA3 + \beta_7 PD2 \\ & + \beta_8 PD3 + \beta_9 WSU2 + \beta_{10} WSU3 \end{aligned} \quad (3)$$

where each independent variable is a dummy variable indicating whether that level was observed. Although alternative models have been shown to provide a better fit for these data elsewhere [26] as might be expected, the conditional logit model was preferred for this analysis as it is consistent with the hybrid model approach. The coefficients from the DCE model (as well as the value set coefficients) were transformed to relative attribute importance (RAI) scores by dividing the level range for each dimension by the sum of all level ranges.

Due to the censoring in the cTTO task (i.e., by design, the cTTO task used in the EQ-5D-Y-3L protocol yields a minimum value of  $-1$ ), the mean cTTO values arguably should be adjusted accordingly before they are used in the worst health state anchoring and mapping approaches. This was done by estimating Tobit models for each health state with left censoring at  $-1$ . A censored hybrid model was estimated using the *hyreg* user-written command in Stata [35]. A linear, additive function as in Eq. 3 was also used. The performance of the value sets was explored by examining the alignment between the predicted values and the Tobit-adjusted cTTO values, using mean absolute error (MAE) and root mean squared error (RMSE).

## 3 Results

Table 1 provides the value sets produced using the three anchoring methods. The underlying data and model output can be found in the ESM. The value sets based on worst

state rescaling and linear mapping approaches perfectly reflect the RAI scores from the DCE (Table S1), whereas the hybrid model RAI scores differ slightly. The most notable difference is that, in the hybrid model value set, mobility is the least important dimension rather than ‘looking after myself’. This appears to be driven by a direct switch in the level rankings between mobility level 3 and pain/discomfort level 2.

The predicted values for the selected health states vary between the three value sets. As expected, the worst state rescaling value set has the lowest value for 33333 (−0.271), which is identical to the observed Tobit-adjusted mean from the cTTO task (Table S2). The mapping value set has the highest value for 33333 (−0.052) and therefore has the smallest overall range and the smallest proportion of health states that are worse than dead ( $n = 1$ ; 0.4%). However, of all three value sets, the hybrid model value set provides the highest values for all but one of the mild health states, illustrating that the differences between these value sets is not limited to their overall range.

Figure 1 compares the values of the 243 health states between the three value sets by rank order. The reference set of rankings is based on the worst state rescaling value set (identical to the DCE rankings). The rank ordering is identical with the linear mapping value set (indicated

by the smooth line), but the range is smaller. In contrast, the rank ordering differs from the hybrid model value set (indicated by the oscillation), and the range is similar to that of the linear mapping value set.

Figure 2 presents the comparison between the predicted values from the three value sets and the Tobit-adjusted means from the cTTO task. All three value sets underpredict the majority of the mild health states. The worst state rescaling value set performs worst overall (higher MAE and RMSE), and the hybrid model value set performs best overall.

## 4 Discussion

### 4.1 Comparison of the Characteristics of the Three Value Sets

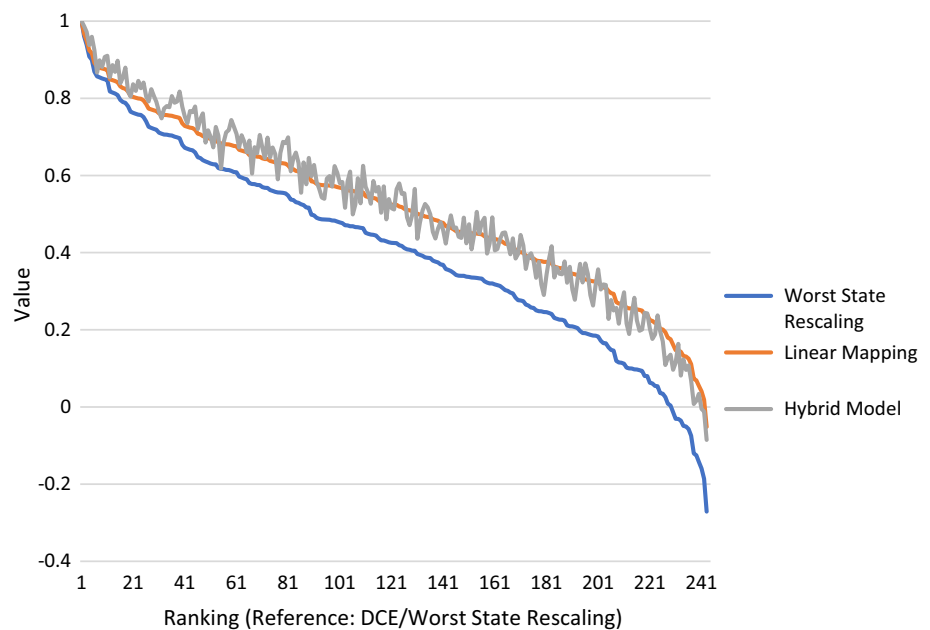
Whilst the value sets produced by each anchoring method are similar, there are some notable differences. The values for 33333 are significantly higher when using the mapping and hybrid model approaches, relative to the worst state anchoring approach, which, by design, fixes the value set at the observed mean cTTO value for the worst health state.

**Table 1** Value sets generated using the three different methods, and selected summary statistics

	Worst state rescaling				Linear mapping				Hybrid model			
	Coef.	Level rank	RAI	Dim. rank	Coef.	Level rank	RAI	Dim. rank	Coef.	Level rank	RAI	Dim. rank
MO2	−0.039	10	11.9%	4	−0.032	10	11.9%	4	−0.013	10	8.6%	5
MO3	−0.151	4			−0.125	4			−0.093	7		
LAM2	−0.061	9	11.5%	5	−0.050	9	11.5%	5	−0.028	9	9.3%	4
LAM3	−0.146	6			−0.121	6			−0.101	6		
UA2	−0.092	8	17.3%	3	−0.076	8	17.3%	3	−0.062	8	16.7%	3
UA3	−0.220	3			−0.182	3			−0.181	3		
PD2	−0.143	7	30.1%	1	−0.118	7	30.1%	1	−0.133	4	35.2%	1
PD3	−0.383	1			−0.317	1			−0.382	1		
WSU2	−0.149	5	29.2%	2	−0.123	5	29.2%	2	−0.119	5	30.1%	2
WSU3	−0.371	2			−0.307	2			−0.327	2		
Predicted values for selected health states (difference from observed Tobit-adjusted mean TTO value in parentheses)												
21111	0.961 (0.006)				0.968 (0.013)				0.987 (0.033)			
12111	0.939 (−0.005)				0.950 (0.006)				0.972 (0.028)			
11211	0.908 (−0.039)				0.924 (−0.023)				0.938 (−0.009)			
11121	0.857 (−0.067)				0.882 (−0.042)				0.867 (−0.057)			
11112	0.851 (−0.109)				0.877 (−0.084)				0.881 (−0.080)			
22222	0.517 (−0.313)				0.600 (−0.230)				0.645 (−0.185)			
33333	−0.271				−0.052 (0.219)				−0.085 (0.186)			
<i>n</i> (%) WTD	14 (6.0%)				1 (0.4%)				3 (1.2%)			

*Coef.* coefficient, *Dim.* dimension, *LAM* looking after myself, *MO* mobility, *PD* pain/discomfort, *RAI* relative attribute importance score, *UA* usual activities, *WSU* feeling worried, sad or unhappy, *WTD* states worse than dead (i.e., a value of below zero. A ranking of 1 denotes the most important level or dimension)

**Fig. 1** Values of EQ-5D-Y-3L health states with the three different value sets. *DCE* discrete choice experiment



Therefore, in comparison, the value sets derived from mapping and the hybrid model have a smaller range.

In terms of relative importance, the worst health state anchoring and mapping approaches perfectly maintain the RAI scores from the DCE, as they are simply linear transformations of the DCE coefficients. In contrast, the hybrid model combines all the individual-level DCE and cTTO data, resulting in different relative importance scores.

In terms of alignment between the predicted values and the observed mean cTTO values (Fig. 2), differences in alignment were observed between all three value sets. This suggests that preferences differ in the DCE and cTTO datasets. Overall, the value set from the hybrid model has the best alignment.

#### 4.2 How Generalisable Are These Results to Other EQ-5D-Y-3L Valuation Studies?

To understand the generalisability of these results, it is first worth noting how these data might differ from other EQ-5D-Y-3L valuation studies. The DCE data used here is based on the same experimental design as the EQ-5D-Y-3L valuation protocol [17, 26]. However, the selection of health states used in the cTTO differs from other EQ-5D-Y-3L valuation studies [11]. Furthermore, our cTTO data are from respondents from four countries [32], whereas the DCE data are from one country. Thus, it may be that the alignment between the cTTO and DCE data in our case study is worse than might be expected when both datasets are made up of respondents from the same country.

Three published EQ-5D-Y-3L valuation studies have made comparisons using different anchoring approaches,

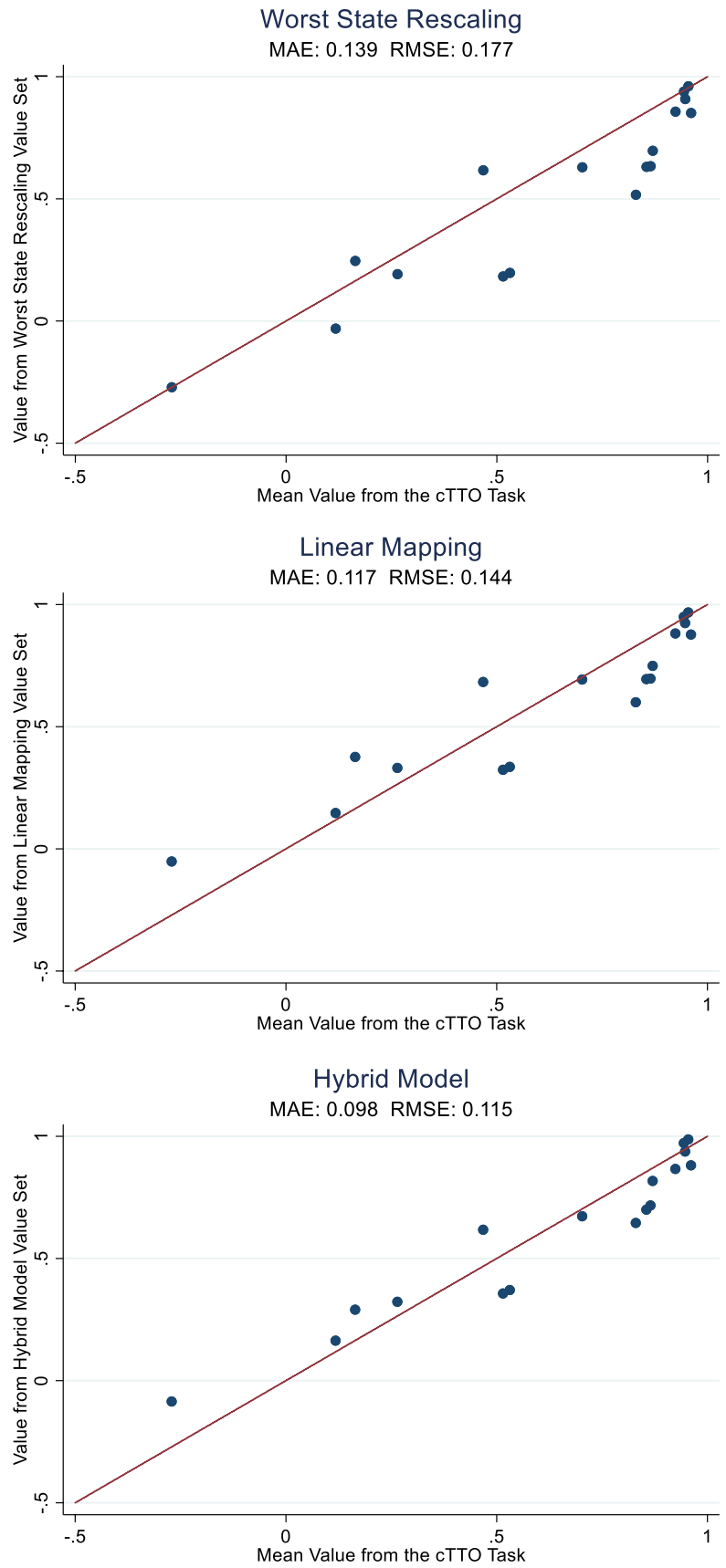
finding that (in line with this case study) alignment is worse with the worst state rescaling approach compared with other approaches [20, 22, 23]. Furthermore, the worst state rescaling approach consistently produced a lower predicted value for the worst state relative to other approaches (resulting in greater value set ranges) [20, 22, 23]. Finally, differences were also observed in the importance of dimensions between the hybrid model and DCE results [20].

#### 4.3 How Should an Anchoring Approach Be Selected?

The extent of the potential differences in value set characteristics based on the anchoring approach chosen highlights how important this choice can be. Analysts must weigh up multiple factors when making their decision. Table 2 summarises some key analytical considerations.

In most cases, worst state rescaling could be viewed as suboptimal. It is unlikely that a cTTO exercise would be conducted with only the worst health state being valued—nor would it be advisable, given that this state tends to be seen as ‘worse than dead’ by many respondents. Using all the cTTO data that has been collected in a linear mapping would be advantageous, as it would improve the alignment between the predicted and observed cTTO values, without compromising the alignment with the DCE results. If we therefore regard worst state rescaling as suboptimal, and both linear mapping and hybrid modelling are feasible, the choice between them could be based solely on alignment with the cTTO data, which will typically be greater with hybrid models. However, it is worth noting that the proportion of

**Fig. 2** Comparison of predicted values and observed Tobit-adjusted mean cTTO values. *cTTO* composite time trade-off, *MAE* mean absolute error, *RMSE* root mean squared error



**Table 2** Key analytical considerations when deciding upon an anchoring approach

Consideration	Worst state rescaling	Linear mapping	Hybrid model
1. Amount of cTTO data required	Very little—mean value of one health state	Relatively little—mean value of at least two health states	Substantial—multiple health states with many observations
2. Range of resulting value set	Maximised—fixed at observed mean value of worst state	Constrained due to prediction error	Constrained due to prediction error
3. Alignment of resulting value set with DCE data	Perfect	Perfect	Variable—depends on similarity between cTTO and DCE results
4. Alignment of resulting value set with cTTO data	Typically worse than other anchoring approaches	Variable—may outperform hybrid model in some circumstances*	Variable—but with adequate data it is likely to outperform other approaches
5. Extent to which results may be impacted by the amount of cTTO observations	Minimal—only impacted to the extent that the mean value of the worst state changes	Minimal—only impacted to the extent that the mean values change	The relative proportions of DCE and cTTO data may impact results

cTTO composite time trade-off, DCE discrete choice experiment

\*Such as when the number of observations for the hybrid model are suboptimal, and the DCE and cTTO data are well aligned

cTTO and DCE data may have an impact on the results of hybrid models. Whilst further research is needed to better understand the extent of the impact on resulting value sets, ultimately this means that the hybrid model approach is less predictable than linear mapping.

However, for normative reasons, some relevant stakeholders (i.e., local end-users) may have priorities that extend beyond the alignment between the resulting value set and the observed cTTO values. There may be cases where alignment with the DCE data is prioritised, such as when mixed samples are sought [27]. For example, if adolescents complete the DCE and adults complete the cTTO, the use of linear mapping will ensure that adolescents' preferences (in relation to the relative importance of dimensions) are fully represented in the value set. Alternatively, some stakeholders may be particularly concerned about the range of the value set due to concerns about comparability between child and adult QALYs and therefore seek to maximise it (possibly using cTTO data from the valuation of an adult instrument instead), which could be achieved relatively simply with worst state rescaling.

These normative considerations, coupled with the lack of a 'gold standard' anchoring approach, present challenges for analysts and we therefore make two recommendations. Firstly, research teams should engage with relevant stakeholders before commencing valuation studies to better gauge priorities relating to the normative issues, which can later inform analytical approaches. Secondly, analysts should ensure that they test the sensitivity of their results using different anchoring approaches, and clearly report the rationale for their preferred approach in study outputs (with results reported based on different anchoring approaches, where relevant and feasible).

#### 4.4 What Future Research is Needed?

There are some areas where further methodological research would be advantageous. Firstly, a better understanding of the hybrid model's experimental design and sample size requirements, and the impact of DCE-cTTO data proportions, would be useful. Future updates to the EQ-5D-Y-3L valuation protocol are likely to include a cTTO design that is optimised for the estimation of hybrid models, which based on past empirical work is likely to involve a greater number of health states [36, 37].

However, an arguably bigger issue relates to the source of data for anchoring. As recent work has shown [6–10], cTTO data in this context may be biased due to participants being reluctant to trade years of life of a child. It may therefore be advantageous to consider alternative methods of obtaining stated preference data for anchoring that avoid this trade-off or present it differently. Prior studies have considered the use of data from a VAS and more novel tasks such as the LOD approach [15, 16]. Another alternative may be to use methods that can generate value sets without additional data collection, such as DCE with duration [38, 39], or the (online) PUF approach [13, 14]. Less is known about these methods and the similarities of the data that they produce compared with cTTO data in this context.

## 5 Conclusion

There are numerous analytical considerations when anchoring DCE data using TTO data and, in the context of EQ-5D-Y-3L valuation (or other child/adolescent-specific instruments), many of these involve value judgements.

There is no single criterion on which to judge the performance of a value set, and therefore no anchoring approach is guaranteed to outperform the others. However, where cTTO data are collected for more than one health state, anchoring on the worst health state is arguably suboptimal. Researchers should seek input from relevant stakeholders when commencing valuation studies to help guide decisions and should clearly set out their rationale for their preferred anchoring approach in study outputs.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40273-022-01214-x>.

**Acknowledgements** The authors are grateful to Dr Juan Manuel Ramos-Goñi, Dr Bram Roudijk, and Dr Elly Stolk for their guidance in this and related projects. The authors are also grateful to Dr Bernarda Zamora for her analytical assistance.

## Declarations

**Disclosure** This article is published in a special edition journal supplement wholly funded by the EuroQol Research Foundation.

**Funding** This manuscript was supported by a grant (104-2020RA) from the EuroQol Research Foundation.

**Conflict of Interest** All authors are members of the EuroQol Group Association. However, the views expressed by the authors do not necessarily reflect the views of the EuroQol Group, nor do they necessarily reflect the views of the authors' institutions. David Mott is an employee of the Office of Health Economics, a registered charity, and Independent Research Organisation, which receives funding from a variety of private and public sector sources. There are no other conflicts of interest.

**Data Availability** The study materials are available from the corresponding author upon reasonable request.

**Authorship statement** All authors contributed study conception and design. The analysis was performed by DM with guidance from ORA. The first draft of the manuscript was written by DM, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

- Herdman M, Gudex C, Lloyd A, Mf J, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011;20:1727–36. <https://doi.org/10.1007/s11136-011-9903-x>.
- Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. 2nd ed. Oxford: Oxford University Press; 2016.
- Lipman SA, Brouwer WBF, Attema AE. What is it going to be, TTO or SG? A direct test of the validity of health state valuation. *Health Econ.* 2020. <https://doi.org/10.1002/hec.4131>.
- Rowen D, Rivero-Arias O, Devlin N, Ratcliffe J. Review of valuation methods of preference-based measures of health for economic evaluation in child and adolescent populations: where are we now and where are we going? *Pharmacoeconomics.* 2020;38:325–40. <https://doi.org/10.1007/s40273-019-00873-7>.
- Lipman SA, Reckers-Droog VT, Kreimeier S. Think of the children: a discussion of the rationale for and implications of the perspective used for EQ-5D-Y health state valuation. *Value Health.* 2021;24:976–82. <https://doi.org/10.1016/j.jval.2021.01.011>.
- Powell PA, Rowen D, Rivero-Arias O, Tsuchiya A, Brazier JE. Valuing child and adolescent health: a qualitative study on different perspectives and priorities taken by the adult general public. *Health Qual Life Outcomes.* 2021;19:222. <https://doi.org/10.1186/s12955-021-01858-x>.
- Lipman SA, Reckers-Droog VT, Karimi M, Jakubczyk M, Attema AE. Self vs. other, child vs. adult. An experimental comparison of valuation perspectives for valuation of EQ-5D-Y-3L health states. *Eur J Health Econ.* 2021;22:1507–18. <https://doi.org/10.1007/s10198-021-01377-y>.
- Åström M, Conte H, Berg J, Burström K. 'Like holding the axe on who should live or not': adolescents' and adults' perceptions of valuing children's health states using a standardised valuation protocol for the EQ-5D-Y-3L. *Qual Life Res.* 2022. <https://doi.org/10.1007/s11136-022-03107-0>.
- Dewilde S, Janssen MF, Lloyd AJ, Shah K. Exploration of the reasons why health state valuation differs for children compared with adults: a mixed methods approach. *Value Health.* 2022. <https://doi.org/10.1016/j.jval.2021.11.1377>.
- 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 Health.* 2022. <https://doi.org/10.1016/j.jval.2021.12.014>.
- Devlin N, Pan T, Stolk E, Verstraete J, Kreimeier S, Rand K, Herdman M. Valuing EQ-5D-Y: the current state of play. *Health Qual Life Outcomes.* 2022;2:2.
- 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. Quality of Life in Kids: Key Evidence to Strengthen Decisions in Australia (QUOKKA), Tools for Outcomes Research to Measure, Value Child Health (TORCH) Project Teams. Preference elicitation techniques used in valuing children's health-related quality-of-life: a systematic review. *Pharmacoeconomics.* 2022;40:663–98. <https://doi.org/10.1007/s40273-022-01149-3>.
- Devlin NJ, Shah KK, Mulhern BJ, Pantiri K, van Hout B. A new method for valuing health: directly eliciting personal utility functions. *Eur J Health Econ.* 2019;20:257–70. <https://doi.org/10.1007/s10198-018-0993-z>.
- Schneider P, van Hout B, Heisen M, Brazier J, Devlin N. The online elicitation of personal utility functions (OPUF) tool: a new method for valuing health states. *Wellcome Open Res.* 2022;7:14. <https://doi.org/10.12688/wellcomeopenres.17518.1>.



15. Webb EJD, O'Dwyer J, Meads D, Kind P, Wright P. Transforming discrete choice experiment latent scale values for EQ-5D-3L using the visual analogue scale. *Eur J Health Econ.* 2020. <https://doi.org/10.1007/s10198-020-01173-0>.
16. 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. *Eur J Health Econ.* 2020;21:1091–103. <https://doi.org/10.1007/s10198-020-01205-9>.
17. 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;38:653–63. <https://doi.org/10.1007/s40273-020-00909-3>.
18. Kreimeier S, Mott D, Ludwig K, Greiner W, IMPACT HTA HRQoL Group. EQ-5D-Y value set for Germany. *Pharmacoeconomics.* 2022. <https://doi.org/10.1007/s40273-022-01143-9>.
19. Prevolnik Rupel V, Ogorevc M, IMPACT HTA HRQoL Group. EQ-5D-Y value set for slovenia. *Pharmacoeconomics.* 2021;39:463–71. <https://doi.org/10.1007/s40273-020-00994-4>.
20. 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 Health.* 2021. <https://doi.org/10.1016/j.jval.2021.10.013>.
21. Shirowai 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. *Med Decis Making.* 2021;41:597–606. <https://doi.org/10.1177/0272989X211001859>.
22. Rencz F, Ruzsa G, Bató A, Yang Z, Finch AP, Brodzsky V. Value set for the EQ-5D-Y-3L in Hungary. *Pharmacoeconomics.* 2022. <https://doi.org/10.1007/s40273-022-01190-2>.
23. Roudijk B, Sajjad A, Essers B, Lipman S, Stalmeier P, Finch AP. A value set for the EQ-5D-Y-3L in the Netherlands. *Pharmacoeconomics.* 2022. <https://doi.org/10.1007/s40273-022-01192-0>.
24. Fitriana TS, Roudijk B, Purba FD, Busschbach JJV, Stolk E. Estimating an EQ-5D-Y-3L value set for Indonesia and testing the assumptions of a two-step DCE/TTO approach. *Pharmacoecon Rev.* 2022;2:2.
25. Kreimeier S, Greiner W. EQ-5D-Y as a health-related quality of life instrument for children and adolescents: the instrument's characteristics, development, current use, and challenges of developing its value set. *Value Health.* 2019;22:31–7. <https://doi.org/10.1016/j.jval.2018.11.001>.
26. Mott DJ, Shah KK, Ramos-Goñi JM, Devlin NJ, Rivero-Arias O. Valuing EQ-5D-Y-3L health states using a discrete choice experiment: do adult and adolescent preferences differ? *Med Decis Making.* 2021;41:584–96. <https://doi.org/10.1177/0272989X21999607>.
27. Rowen D, Mukuria C, Powell PA, Wailoo A. Exploring the issues of valuing child and adolescent health states using a mixed sample of adolescents and adults. *Pharmacoeconomics.* 2022;40:479–88. <https://doi.org/10.1007/s40273-022-01133-x>.
28. Ramos-Goñi JM, Estévez-Carrillo A, Rivero-Arias O, Rowen D, Mott D, Shah K, Oppe M. Does changing the age of a child to be considered in 3-level version of EQ-5D-Y discrete choice experiment-based valuation studies affect health preferences? *Value Health.* 2022. <https://doi.org/10.1016/j.jval.2022.03.001>.
29. Dalziel K, Catchpool M, García-Lorenzo B, Gorostiza I, Norman R, Rivero-Arias O. Feasibility, validity and differences in adolescent and adult EQ-5D-Y health state valuation in Australia and Spain: an application of best-worst scaling. *Pharmacoeconomics.* 2020;38:499–513. <https://doi.org/10.1007/s40273-020-00884-9>.
30. Rowen D, Brazier J, van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Med Decis Making.* 2015;35:328–40. <https://doi.org/10.1177/0272989X14559542>.
31. Robinson T, Hill S, Oluboyede Y. Developing a preference-based measure for weight-specific health-related quality of life in adolescence: the WAITe UK valuation study protocol. *BMJ Open.* 2021;11: e054203. <https://doi.org/10.1136/bmjopen-2021-054203>.
32. 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 Health.* 2018;21:1291–8. <https://doi.org/10.1016/j.jval.2018.05.002>.
33. Feng Y, Devlin NJ, Shah KK, Mulhern B, van Hout B. New methods for modelling EQ-5D-5L value sets: an application to English data. *Health Econ.* 2018;27:23–38. <https://doi.org/10.1002/hec.3560>.
34. Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Med Care.* 2017;55:e51–8. <https://doi.org/10.1097/MLR.0000000000000283>.
35. Ramos-Goñi JM, Craig BM, Oppe M, Van Hout B. Combining continuous and dichotomous responses in a hybrid model. Rotterdam: EuroQol Research Foundation; 2016.
36. Yang Z, Luo N, Bonsel G, Busschbach J, Stolk E. Selecting health states for EQ-5D-3L valuation studies: statistical considerations matter. *Value Health.* 2018;21:456–61. <https://doi.org/10.1016/j.jval.2017.09.001>.
37. Yang Z, Luo N, Bonsel G, Busschbach J, Stolk E. Effect of health state sampling methods on model predictions of EQ-5D-5L values: small designs can suffice. *Value Health.* 2019;22:38–44. <https://doi.org/10.1016/j.jval.2018.06.015>.
38. Norman R, Mulhern B, Viney R. The impact of different DCE-based approaches when anchoring utility scores. *Pharmacoeconomics.* 2016;34:805–14. <https://doi.org/10.1007/s40273-016-0399-7>.
39. 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. <https://doi.org/10.1007/s40273-018-0714-6>.