Anchoring latent scale values for the EQ-5D-Y at 0 = dead

Koonal K. Shah
Juan Manuel Ramos-Goñi
Simone Kreimeier
Nancy J. Devlin

RESEARCH | REPORT
JANUARY 2020

JANUARY 2020

Anchoring latent scale values for the EQ-5D-Y at 0 = dead

Koonal K. Shah
PHMR, London, UK
School of Health and Related Research, University of Sheffield, UK
Office of Health Economics, London, UK

Juan Manuel Ramos-Goñi
Axentiva Solutions, Tacoronte, Spain
Office of the EuroQol Research Foundation, Rotterdam, Netherlands

Simone Kreimeier
School of Public Health, Bielefeld University

Nancy J. Devlin
School of Population and Global Health, University of Melbourne, Australia
School of Health and Related Research, University of Sheffield, UK
Office of Health Economics, London, UK

OFFICE OF HEALTH ECONOMICS
RESEARCH

A revised version of this paper has been published in the European Journal of Health Economics. Please cite as:
Shah, K.K., Ramos-Goñi, J.M., Kreimeier, S. and Devlin, N.J., 2020. An exploration of methods for obtaining 0 = dead anchors for latent scale EQ 5D Y values. European Journal of Health Economics. https://doi.org/10.1007/s10198-020-01205-9

Corresponding Author:
Koonal K. Shah
koonalshah@phmr.com

For further information please contact:

Dr Koonal Shah
Director, Health Economics
Honorary Research Fellow, University of Sheffield

Tel +44 20 8365 5801
Email koonalshah@phmr.com

About OHE Research Papers

OHE Research Papers are intended to provide information on and encourage discussion about a topic in advance of formal publication. Any views expressed are those of the authors and do not necessarily reflect the views or approval of the OHE's Editorial Panel or Research and Policy Committee, or its sponsors.

Once a version of the Research Paper's content is published in a peer review journal, that supersedes the Research Paper and readers are invited to cite the published version in preference to the original version.

Funding and Acknowledgements

Funding for this research was provided by a EuroQol Research Foundation grant. The views expressed do not necessarily reflect the views of the EuroQol Group.

We thank the interviewers and respondents who participated in the study, and to Alison Lawrence and Su-Hing Lo for leading the fieldwork and cognitive interviews. We are grateful for comments received on an earlier draft by participants at the 2018 EuroQol Scientific Plenary Meeting. We thank John Brazier, Aki Tsuchiya and Oliver Rivero-Arias in particular for their helpful feedback and suggestions.

Table of Contents

Abstract

Objectives: To date there have been no value sets to support the use of the EQ-5D-Y in cost-utility analysis. Discrete choice experiments (DCEs) can be used to obtain values on a latent scale, but these values require anchoring at 0 = dead to meet the conventions of quality-adjusted life year (QALY) estimation. The primary aim of this study is to compare four preference elicitation methods for anchoring EQ-5D-Y values.

Methods: Four methods were tested: visual analogue scale (VAS), DCE (with a duration attribute), lag-time time trade-off (TTO) and the recently developed 'location-of-dead' (LOD) element of the personal utility function approach. In computer-assisted personal interviews, adult members of the UK general public were asked to value both EQ-5D-3L health states from an adult perspective (considering their own health) and EQ-5D-Y health states from a child perspective (considering the health of a 10-year-old child). All respondents completed valuation tasks using all four methods, under both perspectives. For a subset of respondents the instrument was controlled for, i.e. EQ-5D-Y health states were valued under both perspectives.

Results: Three-hundred and forty-nine interviews were conducted. Overall, respondents gave lower values under the adult perspective compared to child perspective, with some variation across methods. The mean TTO value for the worst health state (33333) was about equal to dead in the child perspective and worse than dead in the adult perspective. The mean VAS rescaled value for 33333 was also higher in the child perspective than in the adult perspective. The DCE produced positive child perspective values and negative adult perspective values, though the models were not consistent. The LOD median rescaled value for 33333 was negative under both perspectives, and higher in the child perspective. When asked directly about their prioritisation preferences, 65% of respondents indicated that treating adults and treating children should have same priority.

Discussion: There was broad agreement across all methods. Values for 33333 tended to be negative for the adult perspective and closer to 0 for the child perspective. Potential criteria for selecting a preferred anchoring method are presented. We conclude by discussing the decision-making circumstances under which utilities and QALY estimates for children and adults need to be commensurate in order to achieve allocative efficiency.

1 Introduction

The EQ-5D-Y (Youth; three-level version) has been developed as a measure of health outcomes suitable for children and adolescents (Wille et al., 2010; Ravens-Sieberer, 2010). However, no value sets are available, so EQ-5D-Y data cannot currently be used to estimate quality-adjusted life years (QALYs), as required for cost-utility analysis. The EuroQol Group has recognised the need to establish a protocol for conducting EQ-5D-Y valuation studies.

Two methodological EQ-5D-Y valuation studies – one using visual analogue scale (VAS) (Kind et al., 2015) and the other using composite time trade-off (C-TTO) and a discrete choice experiment (DCE) with death (Kreimeier et al., 2018) – have reported somewhat contradictory results. Both studies reported differences in values elicited under adult health and child health perspectives, but in different directions: Kind et al. reported lower mean VAS ratings for the child perspective compared to the adult perspective, while Kreimeier et al. reported higher mean TTO values for the child perspective. The higher TTO values for the child perspective might have been driven by respondents' aversion or unwillingness to trade off life years for a child (i.e. to choose to effectively shorten a child's life). Both of the valuation techniques used by Kreimeier et al. included direct comparisons of health states with (immediate) death, whereas the VAS approach used by Kind et al. did not include any attempt to compare with or anchor at dead. Evidence from Kreimeier et al. suggests that relative preferences regarding dimensions/levels are different for the EQ-5D-3L elicited under the adult perspective and the EQ-5D-Y elicited under the child perspective. However, the authors did not find statistically significant differences across perspectives in the valuation of health state 33333 (the worst state in both the EQ-5D-3L and EQ-5D-Y descriptive systems). The Kind et al. study did not include health state 33333 in its design.

The 'standard' DCE (as opposed to DCE plus duration/death) seems to be a feasible solution for eliciting preferences under a child perspective as no time is attached to the health states, thus avoiding the issues raised by asking respondents to sacrifice the duration of a child's life. Indeed, such preference data for the EQ-5D-Y have been collected from a sample of the UK general population, and are reported elsewhere (Rivero-Arias et al., 2017; Mott et al., 2019). However, the DCE-estimated utilities based on those relative preferences are on an undefined scale, which cannot be used directly in QALY calculations (Oppe et al., 2014). Latent scale DCE data require an anchor point that must be obtained from an additional task or method.

Based on the evidence described above, a key question remains: if we are to use DCE for valuing EQ-5D-Y health states, what is the appropriate method for anchoring the resulting latent scale values? This study tests and compares four methods:

a) Visual analogue scale (VAS)

b) Lag-time TTO

c) Discrete choice experiment with duration (DCEd; described elsewhere as DCE_{TTO} (Mulhern et al., 2014))

d) Location-of-dead (LOD) method, part of the personal utility function (PUF) approach

The aims of the study are: to compare the use of these four alternative methods for establishing anchors and the resulting values for health state 33333; to compare anchors for the EQ-5D-3L/adult perspective and the EQ-5D-Y/child perspective; and to inform development of a protocol for valuing the EQ-5D-Y.

2 Methods

2.1 Instruments

We used two versions of the EQ-5D instrument: the EQ-5D-3L (van Reenen et al., 2018) to describe adult health states and the EQ-5D-Y (van Reenen et al., 2014) to describe child health states. Both instruments comprise broadly the same five dimensions with three levels of response, usually coded 1, 2 and 3, producing health states that can be summarised using five-digit codes (profiles) – e.g. 11111 represents no problems in any dimension; 33333 represents the worst possible health state in either descriptive system. However, the instruments differ in wording. The EQ-5D-3L uses wording considered appropriate for adults, while the EQ-5D-Y was developed as an adaptation of the EQ-5D-3L for use in child and adolescent populations, with changes made to the labels for various dimension and level descriptions.

2.2 Valuation techniques

There exists a broad range of valuation techniques that produce values on a scale anchored at 0 (dead) and 1 (full health). In this study we focused on the four described below. The first three are widely used by health preference researchers (Szende et al., 2007; Brazier et al., 2017). TTO and DCE are the methods currently favoured for the valuation of the EQ-5D-5L instrument (Oppe et al., 2014), albeit different variants of those methods (composite TTO and DCE without duration, respectively) compared to the variants used in this study. VAS is a relatively simple, non-choice-based method, generally agreed to represent the most feasible of the various valuation techniques (Brazier et al., 2017). The fourth method – LOD – is a novel technique co-developed by two of the authors of this paper (Devlin et al., 2019) and considered promising by the authors for the purpose of establishing the location of dead within a descriptive system.

These methods permit latent scale DCE data to be anchored using the value obtained for health state 33333. Other anchoring methods, such as mapping DCE values onto TTO, and combining DCE and TTO data in a hybrid model, have been examined elsewhere (Rowen et al., 2015).

2.2.1 VAS

The VAS exercise involves rating health states (lasting for 10 years, followed by death) or descriptors on a 0-to-100 scale (ranging from 'The worst health you can imagine' to 'The best health you can imagine'). If ratings for 'Dead' and '11111' are obtained, then the rating for health state h can be rescaled using the formula: $(\text{Rating}_h - \text{Rating}_{dead}) / (\text{Rating}_{11111} - \text{Rating}_{dead})$. The rescaled rating is upper bounded at 1 and anchored at 0 = dead.

2.2.2 TTO

We used the lag-time variant of TTO (Devlin et al., 2013; Augustovski et al., 2013). The lag-time TTO involves, as its starting point, a choice between 20 years in full health followed by death (life A) and 10 years in the EQ-5D health state under evaluation, followed by 10 years in full health (the 'lag-time'), followed by death (option B). Respondents could indicate that they preferred life A, preferred life B, or considered both lives to be 'about the same'. Depending on their choice, the amount of time in full health in life A was varied using the same iterative approach as used in the current EQ-5D-5L valuation protocol (Oppe et al., 2016). The task ended when the respondent indicated that life A and life B are about the same. The value for the health state could be calculated (assuming zero temporal discounting) as follows: U = (t-10)/10, where U is the value (utility) and t is the number of years in full health in life A at the respondent's point of indifference.

Lag-time TTO was used in favour of lead-time TTO (as used by Kreimeier et al. (2018) for the valuation of worse-than-dead health states) because in the former the health state under evaluation occurs at the start of the time frame – i.e. if the scenario were to apply to a 10-year-old child, the health state would be experienced whilst the individual in question is still in childhood. However, in lead-time TTO the health state being evaluated occurs after 10 years of full health – i.e. the health state would not be experienced until adulthood. It is acknowledged that if a 10-year-old child enters a health state which then lasts for 10 years, then part of their time experiencing the health state would be in adulthood (particularly given that the EQ-5D-Y is designed for use in 8-to-15-year olds). However, it was deemed useful to maintain consistency with previous EQ-5D-Y valuation work, which had used standard 10-year timeframes (Kreimeier et al., 2018).

2.2.3 DCEd

The DCEd exercise comprised a series of forced choice paired comparisons. Respondents were asked to choose which they preferred out of two EQ-5D health states, each lasting a specified duration (1, 3, 6 or 10 years), followed by death. No indifference option was available.

2.2.4 LOD

The LOD exercise, developed as part of the PUF approach, seeks to locate each respondent's position of dead within a descriptive system. It is a simplified version of the approach used by Devlin et al. (2019), and comprised two parts. First, a ranking task was presented requiring respondents to rank level 1 descriptors for each of the EQ-5D dimensions (e.g. 'no pain or discomfort') from 'most important' to 'least important', thereby asking respondents to consider on which dimensions it was most important to avoid problems. Ties were not permitted. Second, a series of forced choice paired comparison tasks were presented, each involving a choice between living in a specified EQ-5D health state lasting 10 years (followed by death) and 0 years of life (i.e. immediate death). The information gathered in the ranking task was used to personalise the selection of the health states presented in a series of paired comparison tasks, designed to identify the individual's dividing line between states considered to be better or worse than dead. Hence, the ranking task responses played an indirect role in determining the anchor points using the LOD method.

2.3 Study design

All respondents completed all valuation tasks using two different perspectives. In the adult perspective, they were asked to consider their own health, with the EQ-5D-3L instrument used to describe the health states. In the child perspective, they were asked to consider the health of a 10-year-old child, with the EQ-5D-Y instrument used to describe the health states, following the approach used in previous research (Rivero-Arias et al., 2017; Kreimeier et al., 2018). No specific instruction was provided about the identity of the 10-year-old child. Half of the respondents were randomly allocated to completing the tasks for the adult perspective first; the other half completed the tasks for the child perspective first. At the half-way point, a pop-up message appeared on the screen advising respondents of the change in perspective. Interviewers were also instructed to advise respondents of this change.

The survey design (Figure 1) was developed with the view to minimising respondent burden: given the relatively large number of valuation techniques and perspectives being used, we opted to minimise the numbers of tasks for each valuation technique:

1. Ranking – single task involving ranking of EQ-5D level 1 descriptors (as needed for the LOD technique).

2. VAS – ratings for 33333 and Dead. With these two ratings, and assuming that the rating for 11111 is 100 (assumption not tested as part of this study), we were able to calculate an anchored value for 33333.

3. Lag-time TTO – valuations for 22222 (as a warm-up task) and 33333. Note that the TTO technique produced values on the 0 and 1 anchored scale.

4. DCEd – this technique does not produce values directly. Values were estimated by modelling; therefore, a specific experimental design was needed. We used a six-step approach. First, we prepared the set of all 2,430 possible candidates with overlap in two dimensions, no dominant pairs and no repetitions. Second, we simulated 2,000 designs each including 42 pairs. Using the D-efficiency measure based on a main effects model, we extracted all pairs included in the best 20 designs. Third, based on priors from Rivero-Arias et al. (2017) we estimated the choice probabilities for the pairs from step 2. Fourth, using these estimated probabilities, we divided those pairs into three categories: a) $P<=0.2$; b) $0.2<P<=0.35$; and c) $0.35< P<=0.5$ (same for $P>0.5$ applies to B state). For a) we used high distance between durations of each pair (i.e. 1 year in one state versus 10 years in the other) with the longer duration for the less likely state. For b), we used small distance between durations of each pair and the longer duration is for the less likely state. For c), we used all possible combinations of durations (1, 3, 6, 10 years). Fifth, based on the Bansback et al. (2012) model, where the time was an interaction, we simulated 2,000 designs with all possible pairs and selected the best based on the D-efficiency measure. Finally, we blocked the design into six blocks (thereby allocating seven DCE pairs to be completed by each respondent) by minimising the variance of the level balance between blocks. We used the same design for both perspectives.

5. LOD – this technique does not produce values directly. Respondents were asked to complete up to five paired comparison tasks, each involving a choice between 10 years in a specified health state followed by death (option A) and 0 years / immediate death (option B). No indifference option was available. The health states presented were selected based on a simple algorithm that used each respondent's responses to the earlier ranking task to generate a personalised ranking of all 243 health states. The algorithm assumed an equal distance (in utility terms) between each dimension rank (i.e. the difference between the first- and second-ranked dimensions was deemed equal to the difference between the second- and third-ranked dimensions), and between levels (i.e. the difference between level 3 and level 2 was deemed equal to the difference between level 2 and level 1). A random number function was used to break ties to generate the ranking. The health state presented in the first task was always 33333 (ranked 243rd for all respondents). Respondents choosing 33333 over immediate death were not given further choice tasks, but were asked if they could think of any health problems that were so bad that they would rather choose immediate death, and if so, to describe those problems using an open-ended text box. Respondents choosing immediate death over 33333 proceeded to a second task in which 33333 was replaced by the health state ranked 122nd (half-way between 1st and 243rd; this health state varied from respondent to respondent). In the subsequent tasks, the health state presented either improved or worsened in ranking/estimated personal utility depending on the respondent's choice in the previous task. An iterative bisection procedure was used for this purpose (Lenert et al., 1998). Following the fifth task, each respondent's location of dead could be estimated to be within a range comprising 15 to 16 health states.

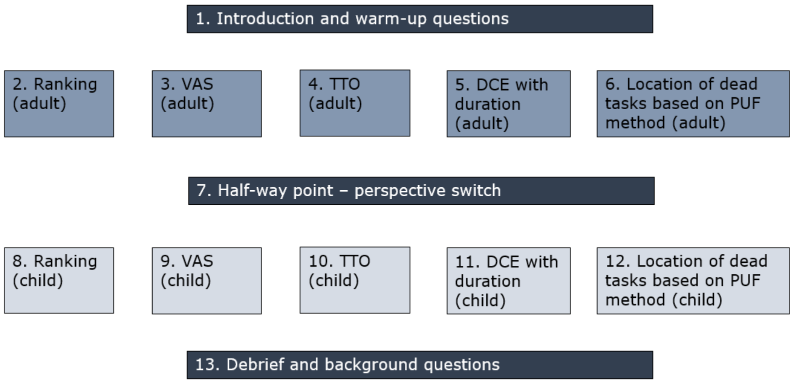

Figure 1. Ordering of the tasks for respondents randomised to the 'adult perspective first' arm

The adult perspective and EQ-5D-3L were used since an aim of the study was to compare anchor points across instruments. However, a small number of additional interviews, using an otherwise identical survey design, were conducted with respondents valuing EQ-5D-Y health states throughout, in both the adult and the child perspectives. This allowed a comparison of the data collected using different perspectives whilst controlling for the descriptive system. Hereafter, the initial interviews are referred to as the 'initial sample' and the additional interviewers are referred to as the 'extended sample'. As with the initial sample, half of the extended sample completed the adult perspective tasks first and half completed the child perspective tasks first.

The valuation tasks were preceded by a small number of warm-up and background questions and followed by debrief and further background questions.

2.4 Data collection

Data were collected from members of the UK general population. The survey was administered via the EuroQol Group Valuation Technology (EQ-VT) platform. The EQ-VT was used as the basis for computer-assisted, one-to-one personal interviews in the homes of respondents, undertaken by a team of five experienced interviewers. The interviewers completed a one-day training session on the methodology and procedures for this study and were asked to follow step-by-step instructions and a script in order to minimise interviewer bias.

The main data collection was preceded by a pilot, which comprised nine cognitive interviews. In addition to completing the valuation tasks using the adapted EQ-VT, pilot respondents were asked probing questions about how they interpreted the tasks, what they found difficult, and how the questionnaire could be improved. All the cognitive interviews were undertaken by two moderators with expertise in qualitative research methods and were carried out in the offices of the moderators' employer. The cognitive interviews were audio recorded and transcribed. Some minor improvements were made to the software (e.g. amendment of on-screen explanatory text) based on the findings of the pilot.

An adapted version of the quality control process developed for EQ-5D-5L valuation studies (Ramos-Goñi et al., 2017) was followed to ensure protocol compliance. Ethics approval for the survey and data collection procedures was granted by the Ethics Committee of the University of Sheffield's School of Health and Related Research (approval reference: 011675).

2.5 Sample

Sample size calculations were based on requirements to estimate DCEd models. We estimated that a minimum of 300 respondents would be needed for the initial sample assuming a requirement of about 50 observations for each of the six blocks of pairs included in the DCEd design. The extended sample comprised a further 50 individuals. The sample comprised adult members of the general population (aged 18 years and older) in two regions in the UK (Midlands and London/Southeast). The sample was recruited using a 'door knock' approach, with interviewers approaching a household member of every third home in a randomly allocated postal area and scheduling interview appointments for those individuals that agreed to participate. A recruitment questionnaire was used to ensure that the sample was broadly representative of the general population in terms of age and gender. Respondents received a shopping voucher worth GBP £10 to thank them for their participation.

The sample for the pilot comprised adult members of the general population in London, recruited using a mixed on-street and 'door knock' approach. Pilot respondents received a shopping voucher worth GBP £40 to thank them for their participation.

2.6 Analysis

Sample background characteristics were described using frequencies and percentages. Box plots were used for describing and comparing lag-time TTO and rescaled VAS values for 33333. TTO values observed at 0 and -1 were not treated as censored. The DCEd data were described using observed choice probabilities for each of the pairs included in the design. DCEd values for 33333 were calculated via different models, including the regular conditional logit model, and conditional logit models assuming non-constant proportionality (Jakubczyk et al., 2018). We estimated models assuming a fixed ½ power and allowing the model to estimate the best-fitted power.

Each respondent's set of choices in the LOD tasks resulted in a range of states within which dead was deduced to be located (for example, for respondents who chose option A in the first task and option B in all subsequent tasks, it was deduced that they located dead between the 228th and the 243rd health states within their own personal ranking). This approach was not possible for respondents who chose option B in the first LOD task, implying that they located dead below 33333 and therefore beyond the descriptive system. For each of the 16 deduced regions, the midpoint rank of the range was calculated and the latent utility corresponding to that midpoint was estimated based on the mixed logit model results from the EQ-5D-Y latent scale DCE study reported by Rivero-Arias et al. (2017). This was done by summing the Rivero-Arias et al. coefficients/disutilities for the relevant dimension-levels for each of the 243 health states. That study produced latent utilities based on the DCE responses of a different sample from the present study (albeit also a representative sample of the UK general public), so combining the data in this way relies on an assumption that respondents in the present study would have responded in the same way as respondents in the Rivero-Arias et al. study had they completed a similar DCE survey. These latent utilities ranged from 0 (corresponding to 11111) to -9.306 (corresponding to 33333; i.e. sum of the five level 3 coefficients/disutilities reported by Rivero-Arias et al.). The value for 33333 was then rescaled onto the 0 (dead) and 1 (full health) scale using the formula: $rescaled_{33333} = (latent_{33333} - latent_{dead}) / (latent_{11111} - latent_{dead})$.

Analyses were undertaken using Microsoft Excel and STATA software.

3 Results

The main interviews were conducted between September and December 2017. The initial sample comprised 299 respondents; a further respondent found the subject matter distressing during the interview and asked to withdraw from the study. No respondents who completed their interview in full were excluded. The mean (median) duration of interview was 40.0 (39.1) minutes for the initial sample; and 40.3 (39.2) minutes for the extended sample. The sample was broadly representative of the general population in terms of age and gender (Office for National Statistics, 2017), though the oldest individuals (aged 70 years and over) are slightly underrepresented (Table 1). The majority of the respondents are parents, though in many cases their children are now adults.

Table 1. Sample background characteristics

	Initial sample		Extended sample		Population
	n	%	n	%	%
Age					
18-29	58	19.4%	11	22.0%	20.0%
30-39	55	18.4%	9	18.0%	16.8%
40-49	44	14.7%	6	12.0%	17.1%
50-59	60	20.1%	10	20.0%	16.7%
60-69	45	15.1%	9	18.0%	13.7%
70+	37	12.4%	5	10.0%	15.8%
Gender					
Female	151	50.5%	26	52.0%	51.1%
Male	148	49.5%	24	48.0%	48.9%
Experience of serious illness					
In self	69	23.1%	4	8.0%	N/A
In family	190	63.5%	20	40.0%	N/A
In caring for others	77	25.8%	4	8.0%	N/A
Self-reported EQ-5D profile					
11111	184	62.5%	37	74.0%	N/A
Any other health state	112	37.5%	13	26.0%	N/A
Children					
No children	66	22.1%	17	34.0%	N/A
Youngest child is <11yrs	84	28.1%	7	14.0%	N/A
Youngest child is 11-18yrs	25	8.4%	3	6.0%	N/A
Youngest child is >18yrs	124	41.5%	23	46.0%	N/A
Experience of working with children					
Yes	60	20.1%	4	8.0%	N/A
No	239	79.9%	46	92.0%	N/A

3.1 Ranking

Amongst respondents in both the initial and extended samples, anxiety/depression was the highest ranked dimension on average (i.e. based on mean rank) in the child perspective but only the third highest ranked in the adult perspective. In the adult perspective, usual activities (initial sample) and pain/discomfort (extended sample) were the highest ranked dimensions; these were the third and second highest ranked in the child perspective, respectively (amongst respondents in both the initial and extended samples). Mobility was found to be the lowest ranked dimension on average amongst both samples and both perspectives.

3.2 VAS

On average, VAS ratings and values (rescaled ratings) given to 33333 were higher in the child perspective than in the adult perspective (Figure 2). A clear majority of respondents considered 33333 to be better than dead when answering from a child perspective; whether 33333 is better than, the same as, or worse than dead under an adult perspective is less clear as the pattern of respondents differed between the initial and extended samples.

3.3 TTO

The average value given to 33333 in the child perspective was close to 0 (or, taking the median, exactly 0), whereas in the adult perspective the average value was clearly negative. The majority of respondents gave a higher value to 33333 in the child perspective than in the adult perspective in both samples (Figure 2). Four of the 349 respondents (1.1%) gave a lower value to 22222 than to 33333.

3.4 DCEd

DCEd model results were in line with VAS and TTO results to the extent that values for 33333 were negative for the adult perspective and positive for the child perspective (this result was consistent across all models). Observed choice probabilities showed a preference for longer life duration in the child perspective (Table 2). This preference for longer duration meant that models were not consistent (i.e. some logically worse health states have higher utilities than logically better, or dominant, health states) in the child perspective. It seems that respondents focused more on the duration of the lives than to the health problems described. The DCEd results indicate that respondents generally avoided shorter life durations and problems with pain/discomfort when considering the health of a 10-year-old child, whereas they focused on problems with mobility and pain/discomfort when considering their own (adult) health. There are considerable differences between the second and fourth columns of Table 2, suggesting low reliability, though it should be noted that the sample size per block and therefore the number of observations for each pair is very small (for the extended sample; which was considerably smaller than the initial sample that had been sized based on DCDd modelling requirements).

3.5 LOD

One respondent (0.3%) chose option B in all of the LOD tasks, implying that all of the health states presented were worse than dead. Conversely, a sizeable minority of respondents chose option A in the first task, implying that 33333 is better than dead. The proportion of respondents making this choice was higher in the child perspective (initial sample: 32.8%; extended sample: 46.0%) than in the adult perspective (initial sample: 23.1%; extended sample: 30.0%). When asked if they could think of any health states that were so bad that they would rather choose immediate death, 57.0% of the respondents in the child perspective and 53.6% of respondents in the adult perspective said that they could. Most of the descriptions of these 'worse than dead' states – in both the child and adult perspectives – focused on being in vegetative states and/or having severe brain damage.

Overall, dead was located lower in the descriptive system in the child perspective than in the adult perspective, resulting in higher rescaled values (Table 3) – in other words, respondents located dead amongst more severe health states in the child perspective. The mean rescaled values shown in

Table 3 underestimate the actual value for 33333, since they do not take into account the fact that for respondents who chose option A in the first task, the rescaled value for 33333 should be positive. Including such positive values would have an upward effect on the mean; it is worth noting that this effect would be stronger in the child perspective since more respondents chose option A in the first task in this version. The median rescaled values are unaffected by this issue since the median respondent chose option B on at least one occasion.

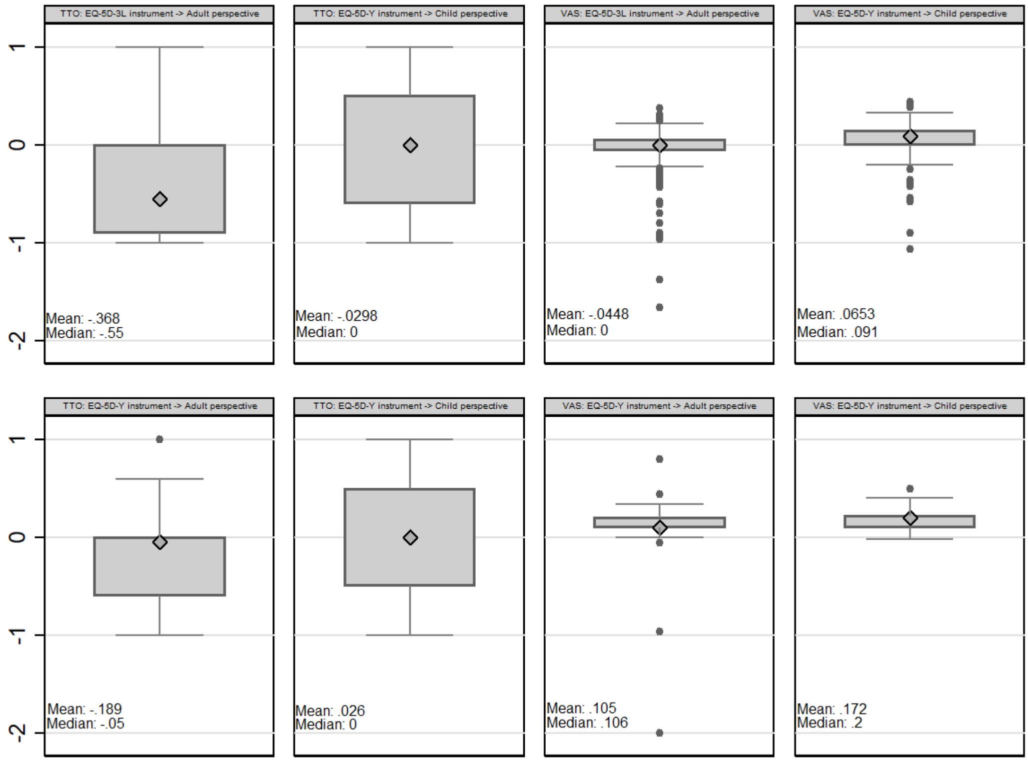

Figure 2. Box-plots of TTO and rescaled VAS values for health state 33333 (upper plots show results for initial sample; lower plots show results for extended sample)

Note: Two outlier VAS values lower than -3 were removed from the graphs for scaling purposes – one was rating EQ-5D-3L under an adult perspective; the other was rating EQ-5D-Y under a child perspective

Table 2. Discrete choice experiment with duration observed choice probabilities

Pair definition	EQ-5D-3L -> Adult perspective vs. EQ-5D-Y -> Child perspective			EQ-5D-Y -> Adult perspective vs. EQ-5D-Y -> Child perspective		
	Adult perspective	Child perspective	Diff Adult-child	Adult Perspective	Child perspective	Diff Adult-child
10 years in 11321 vs 1 year in 31211	0.633	0.653	-0.020	0.250	0.250	0.000
3 years in 11321 vs 6 years in 31211	0.479	0.313	0.167	0.333	0.333	0.000
6 years in 11322 vs 1 year in 12221	0.540	0.540	0.000	0.500	0.500	0.000
3 years in 11323 vs 1 year in 31222	0.563	0.604	-0.042	0.500	0.667	0.167
1 year in 12112 vs 10 years in 11213	0.438	0.333	0.104	0.500	0.417	-0.083
10 years in 12122 vs 1 year in 31112	0.569	0.549	0.020	0.500	0.500	0.000
3 years in 12211 vs 6 years in 11222	0.404	0.447	-0.043	0.625	0.625	0.000
10 years in 12313 vs 1 year in 13111	0.447	0.553	-0.106	0.375	0.250	-0.125
6 years in 12322 vs 3 years in 32221	0.596	0.617	-0.021	0.500	0.250	-0.250
10 years in 13113 vs 1 year in 22112	0.633	0.653	-0.020	0.250	0.625	0.375
10 years in 13233 vs 3 years in 33113	0.588	0.510	0.078	0.667	0.333	-0.333
10 years in 13331 vs 3 years in 23211	0.451	0.510	-0.059	0.333	0.167	-0.167
10 years in 13332 vs 3 years in 22322	0.426	0.574	-0.149	0.500	0.250	-0.250
6 years in 13332 vs 1 year in 32312	0.537	0.519	0.019	0.625	0.625	0.000
10 years in 21133 vs 1 year in 22122	0.500	0.521	-0.021	0.500	0.500	0.000
6 years in 21223 vs 3 years in 31211	0.537	0.537	0.000	0.625	0.250	-0.375
10 years in 21233 vs 1 year in 21322	0.556	0.481	0.074	0.625	0.500	-0.125
6 years in 21322 vs 10 years in 31311	0.480	0.46	0.020	0.500	0.375	-0.125
6 years in 22233 vs 10 years in 31133	0.429	0.388	0.041	0.500	0.500	0.000
10 years in 22323 vs 6 years in 31321	0.520	0.500	0.020	0.375	0.500	0.125
10 years in 22332 vs 3 years in 23311	0.438	0.396	0.042	0.250	0.500	0.250
10 years in 22333 vs 3 years in 23132	0.519	0.519	0.000	0.500	0.500	0.000
1 year in 23111 vs 10 years in 13331	0.490	0.408	0.082	0.500	0.250	-0.250
10 years in 23213 vs 6 years in 31211	0.551	0.633	-0.082	0.375	0.500	0.125
6 years in 23223 vs 10 years in 32123	0.611	0.407	0.204	0.375	0.500	0.125
1 year in 23312 vs 6 years in 31311	0.468	0.404	0.064	0.750	0.500	-0.250
1 year in 23321 vs 6 years in 22333	0.553	0.426	0.128	0.500	0.625	0.125
10 years in 31111 vs 3 years in 21212	0.520	0.500	0.020	0.500	0.500	0.000
1 year in 31111 vs 6 years in 21123	0.375	0.354	0.021	0.583	0.333	-0.250
3 years in 31111 vs 10 years in 12112	0.333	0.313	0.021	0.333	0.250	-0.083
6 years in 31111 vs 10 years in 11312	0.388	0.327	0.061	0.500	0.625	0.125
10 years in 31231 vs 3 years in 33111	0.556	0.481	0.074	0.625	0.625	0.000
10 years in 31233 vs 1 year in 32221	0.400	0.540	-0.140	0.625	0.500	-0.125
10 years in 31323 vs 3 years in 32122	0.420	0.480	-0.060	0.375	0.375	0.000
6 years in 32111 vs 10 years in 23311	0.389	0.370	0.019	0.625	0.500	-0.125
1 year in 32133 vs 10 years in 13233	0.431	0.471	-0.039	0.167	0.500	0.333
3 years in 32211 vs 10 years in 13212	0.383	0.404	-0.021	0.500	0.625	0.125
1 year in 33122 vs 10 years in 23332	0.431	0.412	0.020	0.500	0.667	0.167
3 years in 33211 vs 10 years in 33132	0.520	0.480	0.040	0.375	0.375	0.000
1 year in 33212 vs 10 years in 23233	0.490	0.408	0.082	0.500	0.250	-0.250
3 years in 33212 vs 6 years in 13223	0.392	0.412	-0.020	0.500	0.500	0.000
6 years in 33212 vs 10 years in 23223	0.451	0.412	0.039	0.500	0.333	-0.167
Predicted values for 33333						
Logit model	-0.796	0.059				
Power model (Power =1/2)	-0.468	0.280				
Power model (Power = 0.296)	-0.227	0.188				

*Models coefficients are reported in the Appendix (Table A1)

Table 3. Summary of LOD results

Set of choices	Deduced range in which dead is located	Midpoint of deduced range (rank)	Latent utility of midpoint	Rescaled utility for 33333	EQ-5D-3L -> Adult perspective vs. EQ-5D-Y -> Child perspective				EQ-5D-Y -> Adult perspective vs. EQ-5D-Y -> Child perspective			
					Adult perspective		Child perspective		Adult perspective		Child perspective	
					N	%	n	%	n	%	n	%
BBBBB	Between 1st and 17th ranked states	9	-1.015	-8.170	1	0.3%	0	0.0%	0	0.0%	1	2.0%
BBBBA	Between 17th and 32nd ranked states	24.5	-1.826	-4.098	6	2.0%	0	0.0%	1	2.0%	1	2.0%
BBBAB	Between 32nd and 47th ranked states	39.5	-2.290	-3.064	4	1.3%	1	0.3%	0	0.0%	0	0.0%
BBBAA	Between 47th and 62nd ranked states	54.5	-2.690	-2.459	16	5.4%	9	3.0%	3	6.0%	2	4.0%
BBABB	Between 62nd and 77th ranked states	69.5	-3.048	-2.053	6	2.0%	1	0.3%	1	2.0%	1	2.0%
BBABA	Between 77th and 92nd ranked states	84.5	-3.415	-1.725	15	5.0%	13	4.3%	2	4.0%	2	4.0%
BBAAB	Between 92nd and 107th ranked states	99.5	-3.728	-1.496	9	3.0%	5	1.7%	1	2.0%	1	2.0%
BBAAA	Between 107th and 122nd ranked states	114.5	-4.033	-1.307	25	8.4%	19	6.4%	2	4.0%	1	2.0%
BABBB	Between 122nd and 138th ranked states	130	-4.399	-1.116	11	3.7%	4	1.3%	5	10.0%	2	4.0%
BABBA	Between 138th and 153rd ranked states	145.5	-4.717	-0.973	9	3.0%	11	3.7%	4	8.0%	2	4.0%
BABAB	Between 153rd and 168th ranked states	160.5	-5.005	-0.859	11	3.7%	14	4.7%	2	4.0%	1	2.0%
BABAA	Between 168th and 183rd ranked states	175.5	-5.383	-0.729	18	6.0%	18	6.0%	4	8.0%	1	2.0%
BAABB	Between 183rd and 198th ranked states	190.5	-5.776	-0.611	17	5.7%	14	4.7%	1	2.0%	3	6.0%
BAABA	Between 198th and 213th ranked states	205.5	-6.218	-0.497	18	6.0%	21	7.0%	4	8.0%	4	8.0%
BAAAB	Between 213th and 228th ranked states	220.5	-6.822	-0.364	21	7.0%	20	6.7%	3	6.0%	1	2.0%
BAAAA	Between 228th and 243rd ranked states	235.5	-7.825	-0.189	43	14.4%	51	17.1%	2	4.0%	4	8.0%
A	Dead cannot be located using LOD tasks	N/A	N/A	N/A	69	23.1%	98	32.8%	15	30.0%	23	46.0%
Mean rescaled utility for 33333 (excluding respondents who considered 33333 to be better than dead)					-1.076		-0.787		-1.122		-1.341	
Mean rescaled utility for 33333 (assuming a rescaled utility of 0 for respondents who considered 33333 to be better than dead)					-0.828		-0.529		-0.785		-0.724	
Median rescaled utility for 33333					-0.497		-0.364		-0.729		-0.189	

3.6 Comparison across methods

It is possible to report whether each individual respondent valued 33333 as better than dead via the TTO, VAS and LOD tasks (Table 4). Respondents were more likely to value 33333 as better than dead in the child perspective than in the adult perspective. This finding was consistent across all three methods. Respondents valued 33333 as better than dead via VAS more frequently than via the other two methods. The majority of respondents did not provide internally consistent valuations, in that they valued 33333 as better than dead via one of the methods but as worse than or equal to dead via another of the methods.

Table 4. Comparison across methods: valuation of 33333 in relation to 0 = dead

	Child perspective		Adult perspective		Both perspectives	
	n	%	N	%	n	%
TTO – respondents valuing 33333 as better than dead	146	41.8%	79	22.6%	65	18.6%
VAS – respondents valuing 33333 as better than dead	252	72.2%	154	44.1%	135	38.7%
LOD – respondents valuing 33333 as better than dead	121	34.7%	84	24.1%	77	22.1%
All three methods – respondents providing internally consistent valuations (i.e. 33333 valued as better than dead using all three methods *OR* 33333 valued as worse than or equal to dead across all three methods)	126	36.1%	169	48.4%	76	21.8%

3.7 Debrief questions

The majority of respondents (77.7%) found the child perspective questions more difficult, with a slight majority (51.9%) claiming that they found it somewhat or very difficult to imagine the health of a 10-year-old child (Table 5). Respondents were varied in terms of what sort of child they were thinking of; the most common approach was to think of 'no particular child'. The vast majority of respondents (82.8%) claimed that their responses might have been different if they had been asked to consider a child of a different age, though no information is available about how their responses would have differed. The majority of respondents (65.6%) indicated that the health system should give equal priority to the treatment of adults and children.

Table 5. Responses to debrief questions

Question/response options	n	%
Which questions did you find more difficult – the questions about your own health or the questions about the health of a 10-year-old child?		
The questions about my own health were more difficult	10	2.9%
The questions about the health of a 10-year-old child were more difficult	271	77.7%
Both types of questions were equally difficult	68	19.5%
None of the above / don't know	0	0.0%
How easy or difficult did you find it to imagine the health of a 10-year-old child?		
Very easy	19	5.4%
Somewhat easy	72	20.6%
Neither easy nor difficult	77	22.1%
Somewhat difficult	103	29.5%
Very difficult	78	22.3%
What sort of child were you thinking of when responding to the questions?		
My own child	111	31.8%
A child that I know (but not my own child)	54	15.5%
No particular child	151	43.3%
Myself as a child	15	4.3%
None of the above / don't know	18	5.2%
Would your responses to the questions have been different if you had been asked to imagine a child of a different age – for example, a 5 year old child?		
Yes	289	82.8%
No	60	17.2%
How do you think a health care system with a limited budget should prioritise resources?		
The health system should prioritise the treatment of adults	0	0.0%
The health system should prioritise the treatment of children	119	34.1%
The health system should give equal priority to the treatment of adults and children	229	65.6%
Don't know	1	0.3%

4 Discussion

Our findings in this study were that three of the methods we tested are feasible to use to obtain stated preference-based anchors for a potential EQ-5D-Y value set (LOD's failure to handle cases where 33333 is considered better then dead arguably makes it the least feasible). This opens the possibility that the relative importance of dimensions could be rapidly and inexpensively obtained for EQ-5D-Y via DCE, then subsequently anchored at dead = 0 via a smaller-scale (but more resource-intensive) study applying one of the methods reported here. Indeed, while our study was focused on valuation of the EQ-5D-Y instrument, it is worth noting that in principle this approach could also be followed for the valuation of adult health states using other instruments.

A strong finding from this study was the broad level of agreement across the four very different methods used to locate the relative position of dead = 0 for adult versus child perspectives. Previous studies of EQ-5D-Y valuation, as noted in the introduction, had found contradictory results for TTO and VAS tasks, with values for child health states being higher or lower than corresponding adult health states depending on the method used. However, it is worth noting that the VAS study reported by Kind et al. (2015) did not include 33333 or the rating of dead. Our results are in line with those reported by Kreimeier et al. (2018) to the extent that values for 33333 were higher in the child perspective. However, our study found this pattern more clearly in all methods employed.

There are many improvements and alterations that could be made to the specific approaches used to implement all four methods. Notwithstanding that, the evidence from this study suggests that none of the four can be immediately ruled out as being infeasible or not working (though the way in which the LOD data were combined with data from a separate study may be problematic as it requires a high level of agreement between the preferences of the two study samples in order to be valid). This in turn suggests either that multiple methods could continue to be used in future studies (with conclusions somehow triangulated across methods) or that a choice between them must be made. We have considered the criteria that might be used to guide this choice – our thinking about this is provided here for discussion.

Criteria for choosing between anchoring methods could arguably include:

- Feasibility. We consider multiple methods to be feasible, so in this case feasibility does not identify a single preferred option out of the candidate methods. It should be noted that one respondent in the initial sample and one respondent in the pilot found the subject matter distressing and their interviews were terminated. This issue does not appear to be linked to any particular valuation technique but rather to the general task of considering the severe ill health and death of children (necessary for all of the candidate methods). Hence, it is worth acknowledging that these kinds of studies are not easy to undertake and can pose considerable emotional burden on respondents.

- Acceptability to decision makers. This includes any prior beliefs decision makers may have about desirable theoretical properties of methods. For example, NICE (2013) requires utilities to be based on 'choice-based methods'. TTO and DCEd are generally accepted as being choice-based; the LOD approach is also based on choice-based tasks, though the novelty and relative lack of research using the technique is likely to make it less attractive to decision makers. VAS has tended to be rejected by health economists (with rare exceptions e.g. Parkin and Devlin, 2006) on the grounds that it is not choice-based,

- Potential for administration on-line. While the current study was undertaken using face-to-face interviews, it may be desirable for future studies to be capable of being completed online. This would probably preclude the lag-time TTO or other TTO variants, because of the complexity of the tasks, but would favour VAS, DCEd and potentially the LOD approach (e.g. as implemented by Sullivan et al., 2018).

- Theoretical and empirical coherence with the preference data to be anchored. If unanchored preference data are to be collected via DCE and a second task used for anchoring, it may be considered desirable that there be some degree of consistency or coherence between these two sets of preference data. Our study has proceeded on the basis that this is legitimate, and serves to compare different methods for anchoring the data. VAS valuation may present issues in anchoring latent scale DCE data because the preferences are elicited using completely different sorts of tasks with different biases affecting each. This might favour the use of DCEd – although this raises the question of why DCEd would not then be favoured as the sole approach to eliciting preferences (likewise, if TTO emerges as the preferred anchoring method, this raises the question of why TTO would not be used as the sole valuation method rather than obtaining latent scale DCE data that need to be anchored using a second method. Our response to this is that all child health valuation techniques involving duration pose issues, so it is preferable to focus the majority of resources on a non-duration-based approach – i.e.. DCE – in order to obtain as accurate as possible an estimation of the relative importance of different dimensions and levels). In addition, the current state of the art in DCEd, particularly in terms of design and modelling, has yet to achieve a final solution, meaning that further research is needed to understand the dependency of certain kinds of designs on modelling results as we have found in this study.

- Theoretical and empirical consistency with adult valuations in use in HTA. This raises a fundamental consideration: should the values for the EQ-5D-Y, and QALYs estimated from them, be commensurate with those for adult EQ-5D instruments? That is, should a QALY estimated for a child be equal to a QALY estimated for an adult? Where resource allocation decisions are made from a single health care budget, the achievement of allocative efficiency would rely on being able to consider QALYs gained and foregone across both adult and child interventions. Alternatively, if budgets for health care for children are ring-fenced, then the only decisions for which EQ-5D-Y values would be used are to assess the incremental QALY gains and cost-effectiveness of alternative ways of treating children. In the latter case, commensurability with adult values would not be a requirement. So, for example, and given results reported in this paper, the value set for the EQ-5D-Y might contain no states worse than dead. The extent to which budgets, and therefore cost-effectiveness thresholds, might be characterised as being distinct between adults and children, depends on the nature of the health care system. These normative issues would appropriately be informed by discussions with those responsible for HTA, rather than resting on our judgements as researchers. However, even where the child health care budget is ring-fenced, it is important to note that interventions that avoid the premature death of children involve QALY gains both in childhood and in adulthood, so in practice the complete separation of utilities and QALY estimates is difficult if not impossible.

All four methods used in this paper have their own limitations. The lag-time TTO results relate to a child aged 10 years experiencing health states for 10 years, which takes them to adulthood at 20 years of age, and then experiencing a lag-time period of full health. The time being traded off is therefore partly years in young adulthood and (for negative values) partly years in childhood. In addition, a feature of both lead and lag-time TTO is that the minimum value is determined by the ratio of duration in health to lead/lag time (in the current study, -1) (Devlin et al., 2013). Further, the amount of lead or lag-time available to trade will affect the distribution of values for severe health states (the more time available, the more time is traded).

Similarly, the LOD approach to locating the position of dead within the descriptive system was, in this study, based on quite limited information about the nature of respondents' utility functions. Further, there lacks an agreed means of identifying the position of dead when respondents consider it to be worse than 33333 and therefore to lie outside the EQ-5D descriptive system. More sophisticated approaches to this task are possible and can be rendered suitable for use online (e.g. see Sullivan et al., 2019, where a similar approach was embedded within an online adaptive DCE to create an EQ-5D-5L value set for New Zealand).

A further limitation of this study is that anchors for the EQ-5D-Y were obtained by eliciting stated preferences regarding health states pertaining to a child aged 10 years. We judged that specifying the age for the child to be considered in these tasks was important, or else respondents would have introduced their own, varying and unobserved, assumptions about that. Our choice of 10 years of age in this study was influenced by this being the age also used in the UK latent scale DCE study of EQ-5D-Y values (Rivero-Arias et al., 2017), which produced the data that we wished to re-scale using the anchors derived in the current study. It is also consistent with previous research by Kind et al. (2015) and Kreimeier et al. (2018). It is also the mid-point between the ages of 8 and 12 years where use of the EQ-5D-Y is recommended (ages 12-15 being regarded as an area of overlap where EQ-5D-Y is recommended but the adult EQ-5D can also be used) (van Reenen et al., 2014). Nevertheless, the specification of age means that the anchoring results reported here may be specific to that age and might be different for younger or older children. There is some suggestion from our respondents that this is the case, with 83% saying their responses to the tasks might have been different for children of different ages. This is an issue which does not arise in the valuation of adult health states, where respondents are asked to consider health states as if experienced by themselves, at their current age is. However, in both adult and child valuation tasks, there is no guarantee that the preferences obtained and the age of the person imagined to be experiencing the state match the age of the patients reporting EQ-5D-Y data to which those utilities are then applied.

A related limitation is that under the adult perspective, respondents were asked to consider their own health, whereas under the child perspective they were asked to consider the health of another individual. Hence, some of the differences may be due to respondents' preferences about other individuals rather than about children *per se*. The importance of differences in perspective when eliciting preferences in health has been examined by Dolan et al. (2003), Tsuchiya and Watson (2017) and Cubi-Molla et al. (2018).

The fact that the majority of respondents did not provide internally consistent valuations across the VAS, TTO and LOD methods is potentially concerning. Further research should focus on the reasons why respondents respond differently to different valuation techniques. Approaches that encourage respondents to 'think aloud' and/or to reflect and deliberate on their choices would likely be useful for this kind of research (Devlin et al., 2019; Karimi et al., 2019).

The decision to include four valuation methods and two perspectives in the study resulted in a rather complex study design (Figure 1). In order to minimise respondent burden, the number of tasks included for each method was restricted. This meant that the average interview duration for this study was similar to that for typical EQ-5D-5L valuation studies (e.g. Ludwig et al., 2018). However, it may have been beneficial to have included more VAS and TTO health states in order to assess whether the response patterns observed for 33333 were consistent over the full range.

In conclusion, this study has shown that multiple options exist for providing post-hoc anchors for latent scale DCE preferences. The stated preference methods tested were mostly feasible to use and produced plausible anchors. There was broad agreement between the methods in terms of the placement of the anchor for dead for children versus adults, with the value for 33333 being higher (and more likely to be positive) for children than for adults. The choice between methods, and on what basis that choice should be made, requires further consideration. The choice of anchors raises wider questions about the extent to which the use of values in cost-effectiveness analysis imposes a requirement of commensurability between adult and child health state values.

References

Augustovski, F., Rey-Ares, L., Irazola, V., Oppe, M. and Devlin, N.J., 2013. Lead versus lag-time trade-off variants: does it make any difference? *European Journal of Health Economics*, 14(1), pp.25-31.

Bansback, N., Brazier, J., Tsuchiya, A. and Anis, A., 2012. Using a discrete choice experiment to estimate societal health state utility values. *Journal of Health Economics*, 31, pp.306-218.

Brazier, J., Ratcliffe, J., Salomon, J.A. and Tsuchiya, A., 2017. *Measuring and valuing health benefits for economic evaluation*. Oxford, UK: Oxford University Press.

Cubi-Molla, P., Shah, K. and Burström, K., 2018. Experience-Based Values: A Framework for Classifying Different Types of Experience in Health Valuation Research. *The Patient*, 11(3), pp.253–270.

Devlin, N., Buckingham, K., Shah, K., Tsuchiya, A., Tilling, C., Wilkinson, G. and Hout, B., 2013. A comparison of alternative variants of the lead and lag time TTO. *Health Economics*, 22(5), pp.517-532.

Devlin, N.J., Shah, K.K., Mulhern, B.J., Pantiri, K. and van Hout, B., 2019. A new method for valuing health: directly eliciting personal utility functions. *European Journal of Health Economics*, 20(2), pp.257-270.

Dolan, P., Olsen, J.A., Menzel, P. and Richardson, J., 2003. An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Economics*, 12(7), pp.545–551.

Jakubczyk, M., Craig, B.M., Barra, M., Groothuis-Oudshoorn, C.G.M., Hartman, J.D., Huynh, E., Ramos-Goñi, J.M., Stolk, E.A. and Rand, K., 2018. Choice Defines value: a predictive modeling competition in health preference research. *Value in Health*, 21(2), pp.229-238.

Karimi, M., Brazier, J. and Paisley, S., 2019. Effect of Reflection and Deliberation on Health State Values: A Mixed-Methods Study. *Value in Health*, 22(11), pp.1311-1317.

Kind, P., Klose, K., Gusi, N., Olivares, P.R. and Greiner, W., 2015. Can adult weights be used to value child health states? Testing the influence of perspective in valuing EQ-5D-Y. *Quality of Life Research*, 24(10), pp.2519-2539.

Kreimeier, S., Cole, A., Devlin, N.J., Herdman, M., Mulhern, B., Oppe, M., Shah, K.K., Stolk, E., Ramos-Goñi, J.M., Rivero-Arias, O. and Greiner, W., 2018. Valuation of EQ-5D-Y and EQ-5D-3L health states – the impact of wording and perspective. *Value in Health*, 21(11), pp.1291-1298.

Lenert, L.A., Cher, D.J., Goldstein, M.K., Bergen, M.R. and Garber, A., 1998. The effect of search procedures on utility elicitations. *Medical Decision Making*, 18(1), pp.76-83.

Ludwig, K., von der Schulenburg, J.M.G. and Greiner, W., 2018. German value set for the EQ-5D-5L. *Pharmacoeconomics*, 36(6), pp.663-674.

Mott, D.J., Shah, K.K., Ramos-Goñi, J.M., Devlin, N.J., Rivero-Arias, O., 2019. *Valuing EQ-5D-Y health states using a discrete choice experiment: do adult and adolescent preferences differ?* OHE Research Paper. London: Office of Health Economics.

Mulhern, B., Bansback, N., Brazier, J., Buckingham, K., Cairns, J., Devlin, N., Dolan, P., Hole, A.R., Kavetsos, G., Longworth, L. and Rowen, D., 2014. The feasibility of the DCE_{TTO} for deriving health-state values for EQ-5D-5L. In: Preparatory study for the revaluation of the EQ-5D tariff: methodology report. *Health Technology Assessment*, 18(12).

NICE [National Institute for Health and Care Excellence], 2013. *Guide to the methods of technology appraisal 2013*. London, UK: National Institute of Health and Care Excellence.

Office for National Statistics, 2017. *Population estimates for UK, England and Wales, Scotland and Northern Ireland*. [dataset] Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland [accessed 4 Oct 2017]

Oppe, M., Devlin, N.J., van Hout, B. and Krabbe, P.F.M. and de Charro, F., 2014. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health*, 17(4), pp.445-453.

Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goñi, J.M. and Luo, N., 2016. EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics*, 34(10), pp.993-1004.

Parkin, D. and Devlin, N., 2006. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics*, 15(7), pp.653-664.

Ramos-Goñi, J.M., Oppe, M., Slaap, B., Busschbach, J.J. and Stolk, E., 2017. Quality control process for EQ-5D-5L valuation studies. *Value in Health*, 20(3), pp.466-473.

Ratcliffe, J., Flynn, T., Terlich, F., et al., 2012. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. *Pharmacoeconomics*, 30(8), pp.713-727.

Ratcliffe, J., Huynh, E., Stevens, K., et al., 2016. Nothing about us without us? A comparison of adolescent and adult health-state values for the Child Health Utility-9D using profile case best-worst scaling. *Health Economics,* 25(4), pp.486-496.

Ravens-Sieberer, U., Wille, N., Badia, X., et al., 2010. Feasibility, reliability and validity of the EQ-5D-Y: results from a multinational study. *Quality of Life Research*, 19(6), pp.887-897.

Rivero-Arias, O., Shah, K., Ramos-Goñi, J.M., Mott, D. and Devlin, N., 2017. *Estimating latent scale discrete choice utilities to develop an EQ-5D-Y value set in the UK*. Paper presented at the 34th EuroQol Plenary Meeting. Barcelona. 21-22 September.

Rowen, D., Brazier, J. and Van Hout, B., 2015. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Medical Decision Making*, 35(3), pp.328-340.

Sullivan, T., Ward, J., Hansen, P., Devlin, N., Ombler, F. and Derrett, S., 2019. *A new tool for creating personal and social EQ-5D-5L value sets, including valuing 'dead'*. Economics Discussion Paper 1903, Economics Department, University of Otago. Available at: https://www.otago.ac.nz/economics/otago705521.pdf [accessed 11 Nov 2019]

Szende, A., Oppe, M. and Devlin, N., 2007. *EQ-5D value sets: Inventory, comparative review and user guide*. Dordrecht, Netherlands: Springer.

Tsuchiya, A. and Watson, V., 2017. Re-Thinking 'The Different Perspectives That can be Used When Eliciting Preferences in Health'. *Health Economics*, 26(12), e103–e107.

van Reenen, M., Janssen, B., Oppe, M., Kreimeier, S. and Greiner, W., 2014. *EQ-5D-Y User Guide. Basic information on how to use the EQ-5D-Y instrument*. Available at: https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-Y_User_Guide_v1.0_2014.pdf [accessed 24 May 2018]

van Reenen, M., Oppe, M., Boye, K.S., Herdman, M., Kennedy-Martin, M., Kennedy-Martin, T. and Slaap, B., 2018. EQ-5D-3L User Guide. Basic information on how to use the EQ-5D-3L instrument. Available at: https://euroqol.org/wp-content/uploads/2019/10/EQ-5D-3L-User-Guide_version-6.0.pdf [accessed 13 Nov 2019]

Wille, N., Badia, X., Bonsel, G., et al., 2010. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Quality of Life Research*, 19(6), pp.875-886.

About us

Founded in 1962 by the Association of the British Pharmaceutical Society, the Office of Health Economics (OHE) is not only the world's oldest health economics research group, but also one of the most prestigious and influential.

OHE provides market-leading insights and in-depth analyses into health economics & health policy. Our pioneering work informs health care and pharmaceutical decision-making across the globe, enabling clients to think differently and to find alternative solutions to the industry's most complex problems.

Our mission is to guide and inform the healthcare industry through today's era of unprecedented change and evolution. We are dedicated to helping policy makers and the pharmaceutical industry make better decisions that ultimately benefit patients, the industry and society as a whole.

OHE. For better healthcare decisions.

Areas of expertise

- Evaluation of health care policy
- The economics of health care systems
- Health technology assessment (HTA) methodology and approaches
- HTA's impact on decision making, health care spending and the delivery of care
- Pricing and reimbursement for biologics and pharmaceuticals, including value-based pricing, risk sharing and biosimilars market competition
- The costs of treating, or failing to treat, specific diseases and conditions
- Drivers of, and incentives for, the uptake of pharmaceuticals and prescription medicines
- Competition and incentives for improving the quality and efficiency of health care
- Incentives, disincentives, regulation and the costs of R&D for pharmaceuticals and innovation in medicine
- Capturing preferences using patient-reported outcomes measures (PROMs) and time trade-off (TTO) methodology
- Roles of the private and charity sectors in health care and research
- Health and health care statistics

RESEARCH | REPORT
JANUARY 2020

ohe.org

The Office of Health Economics A Company Limited by Guarantee of Registered No. 09848965
OHE Consulting Ltd Registered Company No. 09853113
OHE is a Charity Registration No. 1170829
Registration Office Southside, 105 Victoria Street, London SW1E 6QT