



A Value Set for the EQ-5D-Y-3L in the Netherlands

Bram Roudijk¹ · Ayesha Sajjad² · Brigitte Essers³ · Stefan Lipman² · Peep Stalmeier⁴ · Aureliano Paolo Finch¹

Accepted: 5 September 2022 / Published online: 10 October 2022
© The Author(s) 2022

Abstract

Background and Objective There is increasing interest in preference-accompanied measures of health for paediatric populations. The child-friendly EQ-5D version, EQ-5D-Y-3L, is one such instrument, but the lack of a Dutch value set prevents its use in economic evaluations of healthcare interventions in the Netherlands. This study aims at covering this gap by collecting preferences using a standardised protocol for deriving EQ-5D-Y-3L value sets.

Methods Composite time trade-off data were collected using videoconferencing interviews, with each respondent completing ten composite time trade-off tasks. Discrete choice experiment data were collected using an online survey, with respondents each completing 15 paired comparisons. Respondents completed these tasks considering what they prefer for a hypothetical 10-year-old child. Discrete choice experiment data were analysed using a ten-parameter mixed-logit model and anchored to the quality-adjusted life-year scale using the mean observed composite time trade-off values.

Results The study collected preferences for 197 respondents using composite time trade-off and for 959 respondents using a discrete choice experiment. The discrete choice experiment sample was representative for the Dutch population in terms of age and sex. The level 3 weight for pain/discomfort was the largest, followed by feeling worried, sad or unhappy, usual activities, mobility and self-care. Health state values ranged between -0.218 and 1 .

Conclusions This study generated a Dutch value set for the EQ-5D-Y-3L, which can be used for the computation of quality-adjusted life-years for economic evaluations of healthcare interventions in paediatric populations.

1 Introduction

The EQ-5D-Y-3L instrument is a generic health-related quality-of-life measure aimed at paediatric populations, such as children and adolescents between ages 8 and 15 years [1]. It was developed by adapting the adult version of the EQ-5D, to make the labels and descriptors relevant for children and adolescents [2]. Several studies have tested the psychometric properties of the EQ-5D-Y, generally demonstrating the instrument is valid and responsive [3, 4]. Instruments such as the EQ-5D-Y-3L are suitable for both population studies and for economic evaluations. For use in economic evaluations,

a value set, i.e. scoring based on preferences, is commonly employed.

In the Netherlands, the National Health Care Institute, Zorginstituut Nederland, recommends that new technologies and interventions should be assessed using a cost-utility analysis, making quality-adjusted life-years (QALYs) the outcome of reference [5]. The EQ-5D-5L and the accompanying Dutch value set is the preferred instrument for the computation of QALYs [6]. Although value sets for Dutch versions of the EQ-5D family of instruments are available, there is currently no EQ-5D-Y-3L value set for the Netherlands. This is problematic for three reasons. First, preferences for health states might differ for child and adult beneficiaries [7, 8]. Second, the descriptive system of the EQ-5D-Y-3L is different from the adult EQ-5D-3L, resulting de facto in a different classification system. Third, although EQ-5D-Y-3L value sets exist for other countries, it is not recommended to use these in the Netherlands. Earlier work with value sets for adult versions of the EQ-5D instruments has suggested using another countries' value set may misrepresent the preferences of the Dutch population [9]. The same is likely to be true for child versions, thus a Dutch version is needed.

✉ Bram Roudijk
roudijk@euroqol.org

¹ EuroQol Research Foundation, Rotterdam, The Netherlands

² Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

³ Maastricht University Medical Center, Maastricht, The Netherlands

⁴ Radboud University Medical Center, Nijmegen, The Netherlands

Key Points

This study generated a value set for the EQ-5D-Y-3L instrument using a standardised protocol, which enables the use of utility index scores in economic evaluations in paediatric populations in the Netherlands.

Both composite time trade-off data and discrete choice experiment data were collected, and modelled independently. The discrete choice experiment data were used for the final value set, with the composite time trade-off data being used to anchor the discrete choice experiment values onto the quality-adjusted life-year scale.

Pain and feeling worried, sad or unhappy were identified as the most important health dimensions, with looking after yourself receiving the smallest weight.

Preliminary work has been conducted on the valuation of health states for the EQ-5D-Y-3L instrument, leading to the development of a standardised protocol for conducting EQ-5D-Y-3L valuation studies [7, 10]. Four national value sets have already been published using this protocol, in Slovenia, Germany, Spain and Japan [11–14]. In this protocol, valuation data are collected in two samples: (1) a representative sample of respondents each completing paired comparisons in a discrete choice experiment (DCE) and (2) a smaller sample of respondents that each complete composite time trade-off (cTTO) tasks [15]. Discrete choice experiment data are used for the estimation of the value set, using logit models. However, as these models produce values on a latent scale, the cTTO data are used to anchor the DCE values on the full health (1) and dead (0) scale, necessary for QALY computations, using a mapping approach or by rescaling on a single cTTO mean value. In both the cTTO and DCE, respondents are asked to imagine a 10-year-old child, who is experiencing the health states described in the choice tasks.

The aim of this study is to develop a value set for the EQ-5D-Y-3L instrument for the Netherlands. The availability of such a value set would enable a cost-utility analysis for paediatric interventions using values specifically generated for children and adolescents aged between 8 and 15 years.

2 Methods

2.1 General Approach

We largely follow the protocol outlined by Ramos-Goñi et al., that is, we recruited the sample of respondents required to answer the DCE and cTTO questions, as well as using the standard proxy perspective of the protocol [10].

The latter entails asking adults of the general public to imagine a 10-year-old child experiencing health problems. We depart from the standard protocol by including a different and larger set of health states in the cTTO, allowing us to model the cTTO data independently. This enables comparisons between DCE-based and cTTO-based value sets for this instrument. Furthermore, the cTTO data were collected through videoconferencing interviews, rather than the recommended face-to-face interviews, owing to the coronavirus disease 2019 pandemic.

2.2 Instrument

The EQ-5D-Y-3L instrument has five dimensions; mobility (MO), looking after myself (SC), usual activities (UA), pain/discomfort (PD) and feeling worried, sad or unhappy (AD).¹ Mobility is defined as walking about, while looking after myself is defined as washing or dressing. For UA, examples relevant to the paediatric population are provided: for example, going to school, hobbies, sports, playing or doing things with friends.

On each of these five dimensions, EQ-5D-Y-3L differentiates between three levels of severity of problems: (1) no problems; (2) some problems and (3) a lot of problems/very worried, sad or unhappy in the last dimension. With these five dimensions with three levels of severity each, the instrument can distinguish between 243 different unique health states for individuals aged between 8 and 15 years. Five-digit numerical codes are assigned to each health state, with the place of a digit in this code indicating the dimension order as MO, SC, UA, PD and AD. The numbers themselves are then either 1, 2 or 3, referring to no problems, some problems and a lot of problems, respectively. An example of this is state “21323”, which is defined as having some problems with walking about, no problems with washing or dressing, a lot of problems doing usual activities, some pain or discomfort and very worried, sad or unhappy.

2.3 Valuation Methods

Two valuation methods are used in this study: cTTO and DCE. Complete composite time trade-off uses an indifference procedure to elicit the preferences of respondents, whom are asked to choose between two lives for a hypothetical 10-year-old child, A and B, or state that they are indifferent. Life A always equals a number between 0 and 10 of life-years in full health, while life B equals 10 years in some health state X. The number of years in full health in life A is varied until indifference is reached. If a respondent prefers immediate death to living life B, he/she enters the lead-time

¹ Abbreviations were harmonized with those used in the adult instrument.

TTO task, in which more life-years can subsequently be traded. The cTTO task is more elaborately described in Janssen et al. and Stolk et al. [15, 16]. The cTTO produces values on a scale range between 1 and -1 . Examples of the choice tasks are presented in the Electronic Supplementary Material (ESM).

In the DCE, respondents are presented with two EQ-5D-Y-3L health states and are asked to indicate which of the two health states they prefer for the hypothetical 10-year-old child. No specification of the duration of living in such a health state is included. Ordinal information on preferences is collected, and latent values are inferred using a logit model. Examples of the choice tasks are presented in the ESM.

2.4 Study Design, Sampling and Survey Procedure

In the cTTO study, an orthogonal health state design comprising 18 states is used [17, 18]. Five mild health states (21111, 12111, 11211, 11121 and 11112), the worst health state 33333 and four additional intermediate health states were added to the orthogonal array. These were selected to maintain near-orthogonality after adding the mild and worst health states to the design. The 28 states were assigned to three blocks of ten health states each, with all blocks containing health state 33333 and at least one mild health state. The DCE survey uses the efficient design suggested in the protocol, with a level overlap, meaning two identical levels, on two of the five dimensions in each choice situation [10]. The full design comprises 150 pairs of health states, blocked into ten blocks of 15 choice tasks per respondent.

Two samples of data were collected; a small target sample of 200 respondents completing ten cTTO tasks each and a larger target sample of 1000 respondents, each completing the 15 DCE choice situations. Ramos-Goñi et al. provide a detailed justification for using these sample sizes [10]. The respondents for the DCE survey and cTTO interviews were recruited by two different market research companies. For the DCE survey, respondents were recruited from a large online panel. Respondents received an invitation by e-mail to participate in the study with a link to the survey. For the cTTO sample, respondents were recruited from a different smaller panel. Respondents were invited by the panel company to an online-administered interview. Hard quota sampling was applied for age and sex for the DCE study, with soft quotas for education because of the difficulty of recruiting respondents with a lower education. For the cTTO study, soft quotas were set for both age, sex and education. To reach certain demographic groups, recruitment via social media was used as an additional recruitment strategy for the cTTO sample.

The cTTO data were collected by four trained interviewers through videoconferencing interviews. Interviewers

connected with the respondents through videoconferencing software and shared their screens to show the respondents the questions and tasks to be completed. Videoconferencing interviews have been used in several valuation studies and methodological studies, and have so far shown to be feasible [19–21]. The standard EQ-VT 2.1 software was used [16, 22]. First, each respondent completed the EQ-5D-5L questionnaire, followed by a short demographic survey and the EQ-5D-Y-3L, to get acquainted with the instrument. Subsequently, the respondents entered the cTTO task, valuing “being in a wheelchair”, followed by a state either “much better than being in a wheelchair” or “much worse than being in a wheelchair”, depending on their answer to the first question. Next, respondents were presented three practice questions using EQ-5D-Y-3L health states (states 21121, 23332 and 13211), followed by ten more health states from one of the three blocks of the study design, presented in a random order. This was followed by the feedback module, debriefing questions and additional demographic questions. The feedback module is a procedure in which respondents are presented the rank order of their answers in the cTTO, allowing them to flag any mistakes [23].

The DCE data were collected through an online survey. Each respondent completed the self-description of the EQ-5D-5L, followed by several demographic questions and the self-description of the EQ-5D-Y-3L. Subsequently, respondents were presented the DCE survey, followed by additional sociodemographic questions. Each respondent completed 15 DCE paired comparisons; one block of questions from the efficient design, plus three dominant choice pairs used as a quality check. The first paired comparison was always a dominant pair, while the other two dominant pairs were randomly placed between the other 15 choice pairs.

2.5 Data Quality

The cTTO interviews were conducted in sets of ten interviews per interviewer, after which the quality of the data was assessed using EuroQol’s quality-control process [21]. This process checked whether a reasonable amount of time was spent on explaining and completing the task. Furthermore, it was assessed whether the lead-time TTO was explained to each respondent and whether there were any inconsistent responses. After these assessments, any necessary feedback was shared with the interviewers. For a more detailed explanation of the quality-control process, please refer to Ramos-Goñi et al. [24].

In the DCE survey, respondents were only included in the final sample if they did not fail any of the three dominance tests and completed the DCE survey in a reasonable amount of time, set at a minimum of 150 seconds. Failing dominance tests may indicate that the respondent is not paying sufficient attention, and 150 seconds was deemed the minimum time

needed to read all the descriptors and make a choice. This cut-off point was also used in the EQ-5D-Y-3L valuation study in Slovenia [11].

2.6 Analyses

The cTTO data and the DCE data are modelled independently in Stata 14.2. The DCE data are used to estimate a value set, while the cTTO data are used to anchor the DCE data on the QALY scale. Modelling the cTTO data separately allows us to compare value sets generated from the cTTO data with value sets generated from the DCE data.

2.7 Modelling Valuation Data

Several ten-parameter models are estimated for the cTTO data, each of the form of Equation 1.

$$U_j = \beta_0 + \beta_1 MO2_j + \beta_2 MO3_j + \beta_3 SC2_j + \beta_4 SC3_j + \beta_5 UA2_j + \beta_6 UA3_j + \beta_7 PD2_j + \beta_8 PD3_j + \beta_9 AD2_j + \beta_{10} AD3_j. \quad (1)$$

Here, the utility assigned to health state j is defined as U_j . β_0 represents the intercept, which can be interpreted as the utility of being without health problems, e.g. in health state 11111. $\beta_1 MO2_j$ to $\beta_{10} AD3_j$ represent the utility decrements assigned to the level-dimension combinations of EQ-5D-Y-3L health states. For example, $\beta_1 MO2_j$ represents the utility decrement (β_1) assigned to level 2 problems in mobility $MO2_j$. As each respondent completes ten cTTO tasks, we assume that responses may be correlated within respondents, for which we account using a random intercept model. Furthermore, models for health state valuations using cTTO often suffer from heteroskedasticity, as the variance in responses for milder health states is much smaller than those for more severe health states. We first test, using a Breusch–Pagan test, whether heteroskedasticity is present, and then account for this by defining an explicit variance function, where the variance σ_j of the error term ϵ_j is dependent on the levels of the dimensions of the EQ-5D-Y-3L of the health state, as in Equation 2:

$$\sigma_j = \exp(\gamma_0 + \gamma_1 MO2_j + \gamma_2 MO3_j + \gamma_3 SC2_j + \gamma_4 SC3_j + \gamma_5 UA2_j + \gamma_6 UA3_j + \gamma_7 PD2_j + \gamma_8 PD3_j + \gamma_9 AD2_j + \gamma_{10} AD3_j). \quad (2)$$

The DCE data are modelled using conditional logit models and mixed-logit models. These take the same form as Equation 1, yet use a logit link function to model the binary data collected in the DCE. We test the model performance of the DCE models by the Akaike Information Criterion (AIC). Furthermore, we perform sensitivity analyses including respondents that were excluded from the final sample because of suspicious response patterns (different cut-offs

for time spent on the DCEs, and allowing for failing one or more dominance tests).

2.8 Anchoring the DCE Data on to the QALY Scale

We explored two methods to anchor the DCE data onto the QALY scale. First, we anchored the DCE data directly on the mean observed value cTTO value of state 33333, by dividing the mean cTTO value for state 33333 by the predicted DCE value for state 33333. Second, we mapped the mean observed cTTO values for the 28 health states included in the design onto the predicted DCE values, using ordinary least-squares regression, with and without specifying an intercept, using a linear equation of the form $U_{TTO} = \alpha * U_{DCE} + \beta$. Here, U_{TTO} is the mean observed cTTO value for a certain health state, U_{DCE} is the predicted DCE value, and α and β (β equals 0 if no intercept is used in the mapping) are the rescaling parameters identified through the mapping. The final rescaling approach was selected based on prediction accuracy criteria such as R-squared, mean absolute error (MAE) and root mean square error (RMSE) and on how well values for the most commonly observed health states in the general population, mild health states, are predicted.

3 Results

The data were collected between early November 2020 and April 2021. One hundred and ninety-seven respondents completed ten cTTO tasks. Fewer than 5% of cTTO interviews were flagged to be of suspicious quality, and all were included in the final sample. Nine hundred and fifty-nine respondents completed the online DCE survey without failing the quality-control criteria (dominance tests or completing the task in less than 150 seconds). An additional 341 respondents either failed at least one dominance task ($n = 72$) or completed the task in less than 150 seconds ($n = 100$) or both ($n = 169$), leading us to exclude these 341 out of 1300 respondents to reach our sample of 959. Table 1 reports on the demographics of both samples and compares them with those of the adult general population [25]. In the cTTO, there was a slight over-sampling of female subjects, as 113 out of 197 respondents were female. Furthermore, although the mean age of the sample (41.9) was similar to that of the Dutch general population (42.3), fewer elderly were sampled. [26] The DCE sample was representative in terms of age and sex, with an over-sampling of the lower-middle and upper-middle educated (9.7 and 12.0 percentage points higher, respectively, compared with the Dutch population). Table 2 reports the responses of both samples to the EQ-5D-5L questionnaire.

Table 1 Sample characteristics of the cTTO and DCE samples

	cTTO (n/%)	DCE (n/%)	Dutch general population (%)
Sex			
Female	113 (57.4)	491 (51.2)	50.7%
Male	84 (42.6)	467 (48.7)	49.4%
Other	0 (0)	1 (0.1)	Not reported
Age, years			
18–24	21 (10.7)	98 (10.2)	10.9%
25–34	52 (26.4)	146 (15.2)	15.9%
35–44	41 (20.8)	135 (14.1)	14.7%
45–54	38 (19.3)	168 (17.5)	17.1%
55–64	29 (14.7)	164 (17.1)	17.0%
65+	16 (8.1)	248 (25.9)	24.4%
Education			
Low	12 (6.1)	75 (7.8)	29.2%
Lower middle	63 (32.0)	360 (37.5)	27.8%
Upper middle	83 (42.1)	409 (42.7)	30.7%
University	39 (19.8)	115 (12.0)	12.3%
Marital status			
Married/registered partners	113 (57.4)	506 (52.8)	45.7%
Single	72 (36.5)	306 (31.9)	39.2%
Divorced	9 (4.6)	82 (8.5)	9.2%
Widowed	3 (1.5)	43 (4.5)	5.8%
Prefer not to say	0 (0)	22 (2.3)	
Having children			
Yes	105 (53.3)	545 (56.8)	N/A
No	92 (46.7)	414 (43.2)	N/A
Income			
<€14.000	17 (8.6)	71 (7.4)	N/A
€14.000–€27.999	22 (11.2)	182 (19.0)	N/A
€28.000–€41.999	34 (17.3)	198 (20.6)	N/A
€42.000–€55.999	30 (15.2)	141 (14.7)	N/A
€56.000–€69.999	23 (11.7)	98 (10.2)	N/A
€70.000–€90.999	16 (8.1)	50 (5.2)	N/A
>€91.000	13 (6.6)	46 (4.8)	N/A
Refuse to answer	42 (21.3)	173 (18.1)	N/A

Frequencies are reported first, followed by percentages of the total sample in brackets. The last column only reports the percentage of the Dutch population [24]

cTTO composite time trade-off, DCE discrete choice experiment, N/A data not available

3.1 cTTO Data

Figure 1 reports the distribution of cTTO responses over the ranges of possible values. 16.3% of responses were negative, indicating worse than dead preferences. The mean value assigned to state 33333 equalled -0.372 . Only for one other health state (13133) a negative mean was reported (-0.038). All other 26 health states reported means between 0.968 (12111) and 0.054 (22333). The means and

Table 2 Responses to the EQ-5D-5L questionnaire for the cTTO and DCE samples

	cTTO (n/%)	DCE (n/%)
Mobility		
No problems	178 (90.4)	718 (74.9)
Slight problems	10 (5.1)	171 (17.8)
Moderate problems	6 (3.0)	40 (4.2)
Severe problems	3 (1.5)	27 (2.8)
Unable to	1 (0.0)	3 (0.3)
Self-care		
No problems	196 (99.5)	907 (94.6)
Slight problems	1 (0.5)	34 (3.5)
Moderate problems	0 (0)	13 (1.4)
Severe problems	0 (0)	5 (0.5)
Unable to	0 (0)	0 (0.0)
Usual activities		
No problems	157 (79.7)	695 (72.5)
Slight problems	28 (14.2)	175 (18.2)
Moderate problems	11 (5.6)	65 (6.8)
Severe problems	1 (0.5)	20 (2.1)
Unable to	0 (0.0)	4 (0.4)
Pain/discomfort		
No problems	122 (61.9)	483 (50.4)
Slight problems	58 (29.4)	337 (35.1)
Moderate problems	15 (7.6)	107 (11.2)
Severe problems	2 (1.0)	31 (3.2)
Extreme problems	0 (0.0)	1 (0.1)
Anxiety/depression		
No problems	155 (78.7)	668 (69.7)
Slight problems	35 (17.8)	199 (20.7)
Moderate problems	7 (3.5)	63 (6.6)
Severe problems	0 (0.0)	25 (2.6)
Extreme problems	0 (0.0)	4 (0.4)

cTTO composite time trade-off, DCE discrete choice experiment

standard deviations of the other health states are reported in the ESM. Table 3 reports the results from the modelling of the cTTO data. The random intercept model (first column) produces logically inconsistent and insignificant parameter estimates for level 2 problems on multiple dimensions. The Breusch–Pagan test showed that heteroskedasticity was present. The other two models, both correcting for heteroskedasticity, performed better and produced logically consistent results, with the constrained intercept model also producing significant parameter estimates. Both models showed a better fit in terms of the Akaike Information Criterion compared with the random intercept model. In all models, AD receives the largest weight, followed by PD. Looking after myself receives the smallest weight.

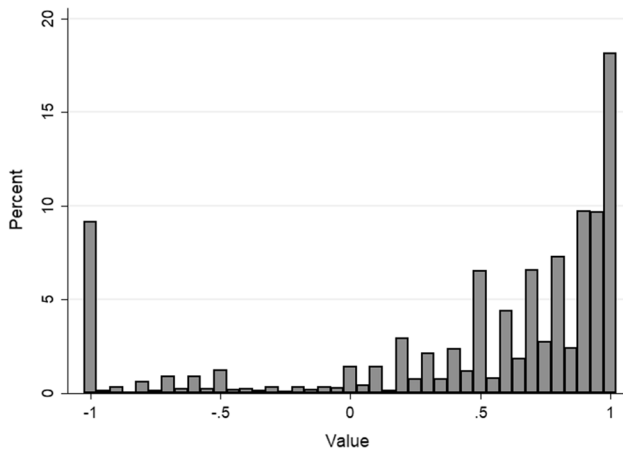


Fig. 1 Distribution of cTTO responses

Table 3 Modelling results of cTTO data

	Random intercept regression	Heteroskedastic model	Heteroskedastic model, intercept constrained at 1
mo2	0.003	0.028*	-0.036**
mo3	-0.102**	-0.086**	-0.090**
sc2	-0.022	-0.020	-0.025*
sc3	-0.105**	-0.065**	-0.069**
ua2	-0.023	-0.034**	-0.041**
ua3	-0.130**	-0.103**	-0.108**
pd2	-0.037	-0.047**	-0.052**
pd3	-0.450**	-0.450**	-0.453**
ad2	-0.070**	-0.062**	-0.066**
ad3	-0.503**	-0.493**	-0.496**
Constant	0.989 [Ⓢ]	0.984 [Ⓢ]	1.000
Observations	1970	1970	1970
Respondents	197	197	197
AIC	2160	1545	1545

The rows represent the coefficient level dimensions (e.g. mo2 represents mobility level 2)

AIC Akaike Information Criterion, CTTO composite time trade-off, * indicates significance at the 5% level, ** indicates significance at the 1% level, [Ⓢ] indicates constants in these models did not significantly differ from 1

3.2 DCE Data

Table 4 reports the results from the modelling of the DCE data, using conditional logit (column 1) and mixed-logit (column 2) models. Columns 3 and 4 report rescaled coefficients, after anchoring these on the QALY scale (see next section). In each of these models, PD receives the largest weight, followed by AD. Looking after myself is assigned the smallest weight. The mixed-logit model outperformed the conditional logit model in terms of model fit, as assessed

by the AIC and was therefore selected as the model to be used for the final value set.

We assessed the effect of including the respondents that failed the DCE response quality criteria along with the final sample in both a conditional logit model and a mixed-logit model. For both models, the relative change in AIC was much larger than the relative change in sample size (see ESM for more details). This large marginal decrease in AIC shows that the responses from the excluded respondents are much noisier than the responses included in the final sample. Using less stringent quality criteria (e.g. respondents can fail one or two dominant tasks) produced similar results, with the criteria used in the current sampling producing the best AIC to sample size ratio.

3.3 Anchoring and Final Value Set

Table 5 reports the results of the anchoring of mean observed cTTO values onto the predicted DCE values. Three anchoring strategies were explored: linear mapping (column 2), linear mapping while constraining the constant at 1 (column 3) and directly rescaling on the mean observed value for state 33333 (column 4). All anchorings were done on the mixed logit model. Rescaling on the mean observed value for state 33333 resulted in MAEs and RMSEs that are consistently larger than those of the other two strategies. This is visualised in Fig. 2. The linear mapping without a constant produced the highest R-squared and showed smaller MAEs and RMSEs for the empirically more prevalent milder health states, while reporting only slightly higher MAEs and RMSEs for the mean values of the observed 28 health states. Therefore, the mapping without a constant was selected as the final model, to be used for the value set. All mixed-logit coefficients were therefore multiplied by 0.0606, the rescaling factor identified by the mapping without a constant. The rescaled DCE models are reported in the last two columns of Table 4. The final model used for the value set, the rescaled mixed-logit model 2 (Table 4, column 4), is represented mathematically as:

$$\begin{aligned}
 U(HS) = & 1 - 0.036 * MO2 - 0.191 * MO3 - 0.028 * SC2 \\
 & - 0.139 * SC3 - 0.058 * UA2 - 0.211 * UA3 - 0.111 \\
 & * PD2 - 0.363 * PD3 - 0.096 * AD2 - 0.314 * AD3.
 \end{aligned}$$

Using this value set, health state 13213 would for example be assigned the value $U = 1 - 0.139 - 0.058 - 0.314 = 0.489$.

Out of all predicted values, 8 of the 243 (3.3%) health states were negative. Figure 3 plots the values generated by the best fitting cTTO model (Table 3, column 3) and the rescaled mixed-logit model (Table 4, column 4) against each other, to compare cTTO-only and DCE-rescaled value sets, to see whether the two methods produce the same values.

Table 4 Modelling results of DCE data

	Conditional logit	Mixed logit	Conditional logit rescaled	Mixed logit rescaled
mo2	- 0.386**	- 0.587**	- 0.048**	- 0.036**
mo3	- 1.890**	- 3.146**	- 0.235**	- 0.191**
sc2	- 0.319**	- 0.458**	- 0.040**	- 0.028**
sc3	- 1.315**	- 2.293**	- 0.163**	- 0.139**
ua2	- 0.683**	- 0.949**	- 0.085**	- 0.058**
ua3	- 2.117**	- 3.474**	- 0.263**	- 0.211**
pd2	- 1.125**	- 1.826**	- 0.140**	- 0.111**
pd3	- 3.336**	- 5.995**	- 0.415**	- 0.363**
ad2	- 0.849**	- 1.581**	- 0.106**	- 0.096**
ad3	- 2.707**	- 5.175**	- 0.336**	- 0.314**
Pseudo R^2	52.16%	-	52.16%	-
Observations	14,385	14,385	14,385	14,385
Respondents	959	959	959	959
AIC	9579	8202	9579	8202

The rows represent the coefficient level dimensions (e.g. mo2 represents mobility level 2)
AIC Akaike Information Criterion, * indicates significance at the 5% level, ** indicates significance at the 1% level
 Rescaling factor mixed logit: 0.0606. Rescaling factor conditional logit: 0.1243, both based on linear mappings without constant

Table 5 Results anchoring analyses

	Linear mapping	Linear mapping (no constant)	Rescaling on mean value state 33333
Coefficient	0.0669	0.0606	0.0683
Intercept	1.0700	-	-
R-squared	89.22%	95.38%	-
MAE (means 28 states)	0.0897	0.0929	0.1108
MAE (mild states)	0.0506	0.0228	0.0279
RMSE (means 28 states)	0.1183	0.1244	0.1436
RMSE (mild states)	0.0573	0.0309	0.0383

cTTO composite time trade-off, *MAE* mean absolute error, *RMSE* root mean square error
 MAE and RMSE of three different anchoring techniques; linear mapping, linear mapping without a constant and rescaling directly on the mean observed value for state 33333 in the *cTTO* data. MAEs/RMSEs are calculated on the mean *cTTO* values for the 28 health states in the health state design (fourth/sixth row) and the means for the mild health states, which have at most one level 3 or two level 2 problems (fifth/seventh row)

Figure 3 identifies differences in the values assigned to health states, which may differ between the *cTTO* and DCE methods. In the extreme case, this difference is more than 0.3.

4 Discussion

4.1 Main Findings

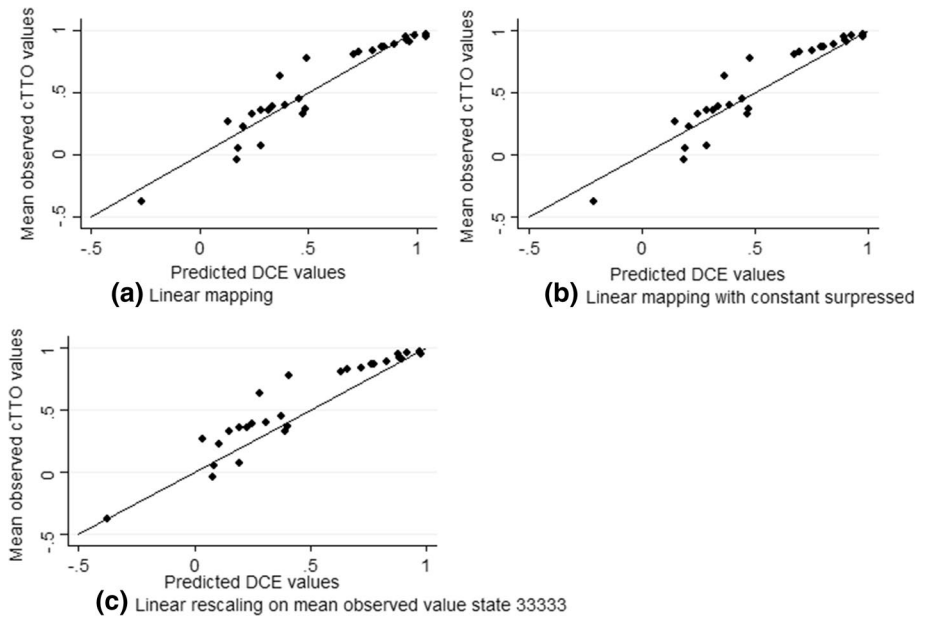
This study aimed to estimate a value set for the EQ-5D-Y-3L instrument for the Netherlands, and by doing so investigated the preferences of the Dutch population for health states for

children. Respondents assigned the highest value to PD, followed by AD, UA, MO and SC. Although the ordering of the dimensions was roughly similar between the *cTTO* and DCE models, there were substantial differences between the health state values generated by the two methods, as is illustrated in Fig. 3.

4.2 Interpretation and Comparison with Existing Value Sets

The ordering of the weights assigned to the levels of the five dimensions was similar in the Netherlands as compared to Spain, Germany, Japan and Slovenia, suggesting

Fig. 2 Scatter plots for the predicted, rescaled DCE values and the mean observed cTTO values for the 28 health states included in the health state design, for all three anchoring strategies.



that the same health dimensions are relatively important in all five countries [11–14]. The Dutch weights assigned to health problems were much larger than those found in Japan, smaller than those found in Slovenia and Spain, and relatively similar to those found in Germany. In all countries, PD was the most important dimension, and SC the least important, except for Germany, where MO was the least important. However, the relative ratio of the largest level 3 weight compared to the smallest level 3 weight was 2.62 in the Netherlands, 2 in Slovenia, 3.86 in Japan, 2.37 in Spain and 3.56 in Germany. This shows that Dutch respondents placed relatively more emphasis on a single health dimension for children as compared with Slovenian respondents, but relatively less than Japanese and German respondents. Generally, the Dutch value set was most similar to the German value set, with roughly the same scale length, but slightly higher values for milder health states and lower values for more severe states.

Negative values accounted for 3.3% of health states in the Dutch value set, indicating that they are considered worse than dead. The lowest score predicted in the current study was -0.218 , for state 33333. In Germany, Spain, Japan and Slovenia, these were -0.283 , -0.539 , 0.288 and -0.691 , respectively. In Japan, no negative values were predicted in the final model, compared to 6.6% in Germany, 16.1% in Spain and 20.6% in Slovenia. This shows that Dutch respondents found only a small set of health states to be worse than dead for children. These findings are similar to findings for the EQ-5D-5L instrument, with Japan showing higher values and Germany lower values [12, 13]. An exception is Spain, showing similar values for the worst health state in the EQ-5D-5L instrument, but lower values for the

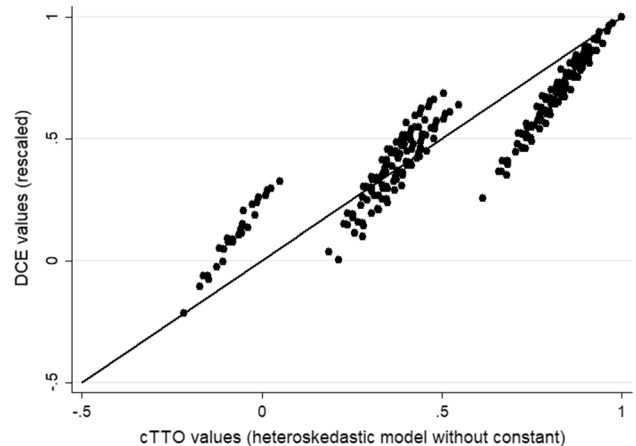


Fig. 3 Values for all 243 health states, generated through either DCE (mixed logit model rescaled on the mean cTTO value for state 33333) or cTTO responses (heteroskedastic model without constant)

EQ-5D-Y-3L instrument [14]. For Slovenia, no EQ-5D-5L value set is currently available [11]

The relative importance of the health dimensions is roughly similar between EQ-5D-Y-3L and EQ-5D-5L in the Netherlands [6]. For EQ-5D-5L, PD and AD receive the largest weight, followed by MO, UA and SC, while for EQ-5D-Y-3L, PD receives the largest weight, followed by AD, UA, MO and SC. The values assigned to the worst health states (-0.446 for 55555 in the EQ-5D-5L, and -0.218 for 33333 in the EQ-5D-Y-3L) differ substantially, with the percentage of negative values being higher in the EQ-5D-5L value set at 15.5%, compared with 3.3% for the EQ-5D-Y-3L value set. This may mean that Dutch respondents consider the same health dimensions important for both adults and

children, but differ in the degree in which they are willing to give up life-years to avoid these health problems, which may reflect a different willingness to trade between adults and children. These differences may be partially caused by a different level structure of the instruments, different descriptors (e.g. anxiety/depression in the EQ-5D-5L vs worried, sad or unhappy in the EQ-5D-Y-3L) or different measurement strategies (EQ-5D-5L value set is cTTO based, while the EQ-5D-Y-3L value set is DCE based). Figure 3 shows a substantial disagreement between cTTO-based and DCE-based value sets for the EQ-5D-Y-3L, which may also reflect the differences in valuation protocols between the EQ-5D-Y-3L and EQ-5D-5L value sets. There is an ongoing discussion on how to interpret differences between youth and adult value sets. One study provides qualitative evidence that respondents differ in the rate at which they are willing to trade life-years to relieve health problems for adults or children [27]. Furthermore, there appears to be more variance in the values assigned to health states for children [8]. One reason for this could be that respondents assign values to health states for other, hypothetical persons in the valuation of health states for children. Incorporating adolescent preferences into the valuations may overcome this issue, and has been shown to be feasible in the valuation of other instruments [28]. A review by Rowen et al. identifies future directions for research that may help us better understand the difference between youth and adult health state values [29].

4.3 Limitations and Strengths

Composite time trade-off data were collected through videoconferencing interviews rather than the recommended face-to-face interviews. Collecting cTTO data via face-to-face interviews was not feasible, owing to the coronavirus disease 2019 pandemic, with the Netherlands experiencing lockdown measures between October 2020 until June 2021. Therefore, it was decided to interview respondents using videoconferencing software. Preliminary studies have shown the feasibility of this mode of administration [19, 20]. A more recent study shows that the mode of administration is unlikely to lead to different values when employing videoconferencing interviews. However, sample selection, i.e. differences in responders due to the recruitment strategy, may affect the collected values [21]. Another limitation of this study is that for the cTTO data, soft quota sampling was used, leading to an over-sampling of the highly educated, while obtaining a reasonable balance of sex and age in the cTTO sample. Furthermore, in the DCE survey, there was an under-sampling of those with the lowest attained education, compensated by those with lower-level and upper-middle-level education. Over-sampling highly educated respondents is a problem observed more frequently in valuation studies for various instruments, regardless of the mode of

administration [30–34]. The sampling issue may be partially related to the choice of recruitment strategy, by employing panel companies to recruit respondents for the cTTO and DCE samples. Last, a limitation of the current study is that the performance of the health state design used for the cTTO cannot be assessed directly with the one recommended by Ramos-Goñi et al., as the Ramos-Goñi et al. design is not a subset of the currently used design [10]. Although anchoring the DCE values onto the cTTO data seems to work well with a larger set of health states, we make no claim of superiority, as no direct comparison can be made.

A strength of this study is regardless of restrictions related to the coronavirus disease 2019 pandemic, it was still possible to collect cTTO data using computer-assisted personal interviews. Furthermore, the data were collected following a strict protocol as outlined by Ramos-Goñi et al., with quality-control procedures being implemented in both the cTTO and DCE data collection [10]. No issues with data quality were identified during the collection of the cTTO data, as defined by EuroQol's quality-control process [24].

5 Conclusions

This study yields a value set for the EQ-5D-Y-3L instrument, enabling a cost utility analysis for paediatric interventions in the Netherlands. Pain/discomfort was identified as the most important dimension, followed by feeling worried, sad or unhappy, usual activities, mobility and looking after myself. Of all health states, 3.3% had a negative value, and the lowest value assigned to any health state was -0.218 . The relative importance of the health dimensions seems to be the same between the adult EQ-5D-5L instrument and the paediatric EQ-5D-Y-3L instrument. However, other properties, such as the scale length and the percentage of negative values, seem to differ between the value sets.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40273-022-01192-0>.

Declarations

Funding This study was funded by the EuroQol Research Foundation (EQ project 100-2020RA).

Conflicts of Interest Bram Roudijk, Stefan Lipman, Peep Stalmeier and Aureliano Paolo Finch reported receiving grants from the EuroQol Research Foundation during the conduct of the study. Bram Roudijk, Peep Stalmeier and Aureliano Paolo Finch are members of the EuroQol Group. Ayesha Sajjad and Brigitte Essers report no conflict of interest.

Ethics Approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Erasmus School of Health Policy and Management (07-01-2020/Lipman 20-25).

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent for Publication Not applicable.

Availability of Data and Material The anonymised datasets generated during the current study are available from the corresponding author on reasonable request.

Code Availability Not applicable.

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

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

1. EuroQol. EQ-5D-Y user guide: basic information on how to use the EQ-5D-Y instrument. Version 2.0. Rotterdam: EuroQol Research Foundation; 2020. Available from: <https://euroqol.org/publications/user-guides/>. [Accessed Sep 2021].
2. Wille N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res.* 2010;19(6):875–86.
3. Ravens-Sieberer U, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res.* 2010;19(6):887–97.
4. 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(1):31–7.
5. Nederland Z. Guideline for economic evaluations in healthcare. Diemen: Zorginstituut Nederland; 2016.
6. Versteegh MM, et al. Dutch tariff for the five-level version of EQ-5D. *Value Health.* 2016;19(4):343–52.
7. Kreimeier S, et al. 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(11):1291–8.
8. 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(9):1507–18.
9. Roudijk B, Donders ART, Stalmeier PF. Cultural values: can they explain differences in health utilities between countries? *Med Decis Making.* 2019;39(5):605–16.
10. Ramos-Goñi JM, Oppe M, Stolk E, Shah K, Kreimeier S, Rivero-Arias O, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics.* 2020;38(7):653–63.
11. Rupel VP, Ogorevc M. EQ-5D-Y value set for Slovenia. *Pharmacoeconomics.* 2021;39(4):463–71.
12. Shirowa 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(5):597–606.
13. Kreimeier S, Mott D, Ludwig K, Greiner W. EQ-5D-Y value set for Germany. *Pharmacoeconomics.* 2022;2:1–13.
14. Ramos-Goñi JM, Oppe M, Estévez-Carrillo A, Rivero-Arias O, Wolfgang G, Simone K, et al. 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.* 2022;25(5):835–43.
15. Janssen BM, Oppe M, Versteegh MM, Stolk EA. Introducing the composite time trade-off: a test of feasibility and face validity. *Eur J Health Econ.* 2013;14(1):5–13.
16. Stolk E, et al. Overview, update, and lessons learned from the International EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. *Value Health.* 2019;22(1):23–30.
17. 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(4):456–61.
18. Yang Z, Luo N, van Busschbach J, Stolk E. Using orthogonal design in selecting health states for the construction of EQ-5D-3L value set. *Value Health.* 2016;19:A386.
19. Finch AP, Mereaglia M, Ciani O, Roudijk B, Jommi C. An EQ-5D-5L value set for Italy using videoconferencing interviews and feasibility of a new mode of administration. *Soc Sci Med.* 2022;292: 114519.
20. Lipman SA. Time for tele-TTO? Lessons learned from digital interviewer-assisted time trade-off data collection. *Patient.* 2021;14(5):459–69.
21. Rowen D, Mukuria C, Bray N, Carlton J, Longworth L, Meads D, et al. Assessing the comparative feasibility, acceptability and equivalence of videoconference interviews and face-to-face interviews using the time trade-off technique. *Soc Sci Med.* 2022;309: 115227.
22. Oppe M, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health.* 2014;17(4):445–53.
23. Wong EL, Ramos-Goni JM, Cheung AW, Wong AY, Rivero-Arias O. Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. *Patient.* 2018;11(2):235–47.
24. Ramos-Goñi JM, et al. Quality control process for EQ-5D-5L valuation studies. *Value Health.* 2017;20(3):466–73.
25. Central Bureau of Statistics. Netherlands, population characteristics. Available from: <https://opendata.cbs.nl/statline/#/CBS/nl/>. [Accessed 29 Apr 2021].
26. Central Bureau of Statistics. Netherlands, age characteristics. Available from: <https://www.cbs.nl/nl-nl/visualisaties/dashboard-bevolking/leeftijd/bevolking>. [Accessed 25 Nov 2021].
27. 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(1):1–14.
28. Ratcliffe J, Huynh E, Chen G, Stevens K, Swait J, Brazier J, et al. Valuing the child health utility 9D: using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. *Soc Sci Med.* 2016;157:48–59.
29. 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(4):325–40.
30. Gamper EM, King MT, Norman R, Efficace F, Cottone F, Holzner B, et al. EORTC QLU-C10D value sets for Austria, Italy, and Poland. *Qual Life Res.* 2020;29(9):2485–95.

31. Finch AP, Gamper E, Norman R, Viney R, Holzner B, King M, et al. Estimation of an EORTC QLU-C10 value set for Spain using a discrete choice experiment. *Pharmacoeconomics*. 2021;39(9):1085–98.
32. Purba FD, Hunfeld JA, Iskandarsyah A, Fitriana TS, Sadarjoen SS, Ramos-Goñi JM, et al. The Indonesian EQ-5D-5L value set. *Pharmacoeconomics*. 2017;35(11):1153–65.
33. Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care*. 2016;54(1):98.
34. Pickard AS, Law EH, Jiang R, Pullenayegum E, Shaw JW, Xie F, et al. United States valuation of EQ-5D-5L health states using an international protocol. *Value Health*. 2019;22(8):931–41.