



# Estimating an EQ-5D-Y-3L Value Set for Indonesia by Mapping the DCE onto TTO Values

Titi Sahidah Fitriana<sup>1,2</sup> · Bram Roudijk<sup>3</sup> · Fredrick Dermawan Purba<sup>4</sup> · Jan J. V. Busschbach<sup>1</sup> · Elly Stolk<sup>3</sup>

Accepted: 11 October 2022 / Published online: 9 November 2022  
© The Author(s) 2022

## Abstract

**Background and Objectives** Methods for estimating health values in adult populations are well developed, but lag behind in children. The EuroQol standard protocol to arrive at value sets for the youth version of the EQ-5D-Y-3L combines discrete choice experiments with ten composite time trade-off values. Whether ten composite time trade-off values are sufficient remains to be seen and this is one of the reasons the protocol allows for experimental expansion. In this study, 23 health states were administered for the composite time trade-off. This methodological research is embedded in a study aimed at generating a representative value set for EQ-5D-Y-3L in Indonesia.

**Methods** A representative sample of 1072 Indonesian adults each completed 15 discrete choice experiment choice pairs via face-to-face interviews. The discrete choice experiment responses were analysed using a mixed-logit model. To anchor the discrete choice experiment values onto the full health-dead quality-adjusted life-year scale, composite time trade-off values were separately obtained from 222 adults living in Java for 23 EQ-5D-Y-3L states. The derived latent discrete choice experiment values were mapped onto the mean observed composite time trade-off values to create a value set for the EQ-5D-Y-3L. Linear and non-linear mapping models were explored to estimate the most efficient and valid model for the value set.

**Results** Coefficients obtained from the choice model were consistent with the monotonic structure of the EQ-5D-Y-3L instrument. The composite time trade-off data showed non-linearity, as the values for the two worst states being evaluated were much lower than predicted by a standard linear model estimated over all composite time trade-off data. Thus, the non-linear mapping strategies with a power term outperformed the linear mapping in terms of mean absolute error. The final model gave a value range from 1.000 for full health (11111) to  $-0.086$  for the worst health state (33333). Values were most affected by pain/discomfort and least by self-care.

**Conclusions** This article presents the first EQ-5D-Y-3L value set for Indonesia based on the stated preferences of adults asked to consider their views about a 10-year-old child. Mapping the mixed-logit discrete choice experiment model with the inclusion of a power term (without a constant) allowed us to generate a consistent value set for Indonesian youth. Our findings support the expansion of the composite time trade-off part of the EQ-5D-Y valuation study design and show that it would be wise to account for possible non-linearities in updates of the design.

✉ Titi Sahidah Fitriana  
t.fitriana@erasmusmc.nl; titi.sahidah@yarsi.ac.id

<sup>1</sup> Section Medical Psychology and Psychotherapy, Department of Psychiatry, Erasmus MC University Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands

<sup>2</sup> Faculty of Psychology, YARSI University, Jakarta, Indonesia

<sup>3</sup> EuroQol Research Foundation, Rotterdam, The Netherlands

<sup>4</sup> Department of Developmental Psychology, Faculty of Psychology, Universitas Padjadjaran, Jatinangor, Indonesia

## Key Points for Decision Makers

An EQ-5D-Y-3L value set for Indonesia is now available and can be used for cost-utility analyses and health-related quality-of-life studies for children in Indonesia.

Values for EQ-5D-Y-3L, provided by adults considering their views about a 10-year-old child, are different from values for the EQ-5D-5L provided by adults from their own perspective. This finding justifies separate value sets for children and adults.

More empirical and theoretical studies are needed to assess the performance of existing youth valuation methods.

## 1 Introduction

Generic measures of health-related quality of life, accompanied by preference weights (values), are increasingly used in healthcare decision making [1]. These values reflect the health-related quality of life of the patient reporting that health state, based on social preferences derived from a separate ‘valuation study’ [2]. These values can be combined with survival data to calculate quality-adjusted life-years (QALYs). The QALY is a preferred outcome in economic evaluations of healthcare interventions. The methodology for estimating health state values in adult populations is well developed, with the inclusion of detailed guidelines by many international agencies, such as the National Institute for Health and Care Excellence [3] and the International Society for Pharmacoeconomics and Outcomes Research [4]. In contrast to adult health valuations, there is considerable debate about basic aspects of the most opportune ways to measure and value child health.

This debate can be illustrated in the study of Ungar [5], and the review by Rowen et al. [6], where these two papers describe the challenges faced in valuing health states for children. These challenges concern who should evaluate the health states (children or adult), which age of children and the method used to elicit preferences. Some researchers use visual analogue scales to arrive at values [7]; others use discrete choice experiments (DCEs) [8] or time trade-off (TTO) and the Standard Gamble [9]. Notably, many valuation studies use a mix of methods, where most health states are valued utilising a method that is potentially more readily understood by respondents, such as visual analogue scales and DCEs [8], and then these values are adjusted with, or ‘mapped’ onto more complex valuation methods, such as TTO and Standard Gamble, to place the values onto the full health (1) to dead (0) scale required for the estimation of QALYs in economic evaluations [10, 11].

This ‘two-step approach’ is said to reduce the complexity of the task in valuing children’s health for the responder. Moreover, the two-step valuation approach reduces the exposure of responders to questions concerning life and death, as is the case with TTO and Standard Gamble. Such reduced exposure to questions concerning life and death is believed to reduce anxiety and increase acceptability during the interview process [8], thus avoiding the death scenario for children. This two-step approach comes with the disadvantage that the model to arrive at the final values for all health states becomes more complex and depends on more assumptions. One of the assumptions is that one can validly scale the DCE scores onto a TTO scale where 0.00 is the value of dead and 1.00 the value of full health, where only a limited number of TTO health states is available. For example, when

the data distribution is linear, then simple anchoring using only the worst health state (33333) is possible. This is a simple solution, but it can also be considered as risky, given that the values for severe health states have the highest variance, leaving a risk of misprediction of the scale due to the high variation in values for state 33333. In light of this risk, the EuroQol group has proposed a standard two-step DCE/TTO valuation protocol for the EQ-5D-Y-3L incorporating the suggestion of ten composite TTO (cTTO) states [12]. Whether ten health states are sufficient remains to be seen, and this is one of the reasons the standard valuation protocol allows for the experimental expansion.

In our investigation, we used five health states from the standard protocol and extended the design by adding 18 cTTO health states to give a total of 23 cTTO states. One might expect when the relationship between DCE and cTTO is simple and linear, the additional cTTO values add little to the validity of the model that generates values for all health states. However, when the relationship is more complex, for example when the relation between cTTO and DCE is concave or convex, the number of health states with cTTO values and the choice of the health states will determine whether a valid model can be constructed. Under these conditions, having more health states means a better estimation of the relation of DCE and cTTO and such observations will also be of value for the future youth valuation study.

Currently, there are several approaches to model values derived from EQ-5D valuation studies, for example modelling the cTTO values in isolation, hybrid modelling [13], where the cTTO and DCE data is modelled jointly, or ‘mapping’, where the DCE data are modelled, and a mathematical relation between the cTTO and DCE data is inferred using some type of regression model. The research question above is embedded in an effort to arrive at the first representative value set for the EQ-5D-Y-3L for Indonesia.

## 2 Methods

### 2.1 Instrument

#### 2.1.1 EQ-5D-Y-3L

EQ-5D-Y-3L is a generic instrument developed by the EuroQol Group to measure the health-related quality of life in children and adolescents. There are five dimensions: mobility (walking about), looking after myself, doing usual activities, having pain or discomfort, and feeling worried, sad or unhappy. In the standard 3L version, the response format has three severity levels: no problems, some problems and many problems [14].

## 2.2 Valuation Tasks

The EQ-5D-Y-3L valuation protocol suggests using DCE as the method to generate the ‘latent’ scale values, that is, the relative distance between points on the scale. A detailed description of that design has been provided by Ramos-Goñi et al.

The DCE design consists of ten blocks of 15 pairs of health states valued in a population not less than 1000 responders [12]. In each block, the health state pair presentation order was randomised, as well as the left/right presentation. Attribute level overlap was imposed on the design where, in each choice task, two attributes were presented at the same level in both alternatives. To help responders identify level severity, color coding was applied on the presented task. The darker the grey color, the more severe the levels are. This method can reduce attribute non-attendance and make the task easier, as suggested by Jonker et al. [15]. The valuation protocol allows DCE data to be collected from other participants who do not complete the cTTO tasks, increasing administrative efficiency. The framing of the DCE task is: “*Considering your views for a 10-year-old child, which health state do you prefer?*”. As DCE produces values on a latent scale, an additional method is needed to arrive at the full-health to dead QALY scale where 1 represents full health, and 0 represents dead. Thus, values below 0 mean health states that are considered worse than dead.

Composite time trade-off combines a conventional TTO to elicit values for states regarded better than dead and a lead-time TTO for states worse than dead [16]. Respondents for cTTO were different than DCE respondents and independently obtained from the study sample. The cTTO values were collected for 23 health states grouped into two blocks in computer-assisted personalised interviews. The preliminary protocol recommended to collect cTTO values for ten states from 200 individuals (200 observations/state). We opted to double the number of states and to reduce the number of observations per state to 100. A previous calculation for the EQ-5D-5L protocol showed that the minimum sample size required was 100 observations per block to reach the desired standard error of 0.01 [13, 17]. While the reduced number still lowers the reliability of each data point, the higher number of health states also reduces the risk of model specification issues because the analyst has more options to verify that the assumption of the specified mapping function is correct.

The 23 health states for the cTTO task were selected based on the following considerations:

- We included 18 health states that represent an orthogonal array. Orthogonal designs have been shown to provide an adequate basis for modelling health state values [18].
- The orthogonal array concentrated on moderate health states; hence we added two health states; 33332 and 32232, to balance the design. These two states have a high level sum score, which is the sum of the levels of the EQ-5D-Y-3L dimensions, which can be considered a crude measure of health state severity. State 33332 has level sum score 14, and state 32232 has level sum score 12. The level sum score ranges from 5 to 15, for which we defined into three groups: mild, moderate and severe. We considered states with level sum scores of 5–7 to be mild, as these states have a maximum of one level 3 on a dimension, or two level 2s. We considered level sum scores of 13–15 to be severe, as this meant the health states would have 3 level 3s, meaning many problems, on at least three dimensions. The remaining states were categorised as moderate. Therefore, both of the added states could be considered relatively severe in terms of a level sum score.
- We added five of ten health states suggested by Ramos-Goñi et al. [12] to provide observations that can be directly compared to data collected in other countries. Two of these health states overlapped with the selected health states, leaving the design with 23 health states as a final design. Details of the health state can be seen in Table 1.

The framing of the cTTO task in the EQ-5D-Y-3L valuation is “*Considering your views for a 10-year-old child. What do you prefer Life A, Life B, or Life A and B are about the same?*”. For health states considered better than dead, life B referred to 10 years of impaired health and life A referred to life in full health, and the length of life was varied over a series of questions until an indifference was reached. If respondents indicated that the health state was worse than dead (below 0), they were offered a slightly different question: the lead-time TTO. Life B was still characterised as 10 years in that impaired state but preceded by 10 years in full health, making the total lifespan 20 years. After valuing all health states, the responses elicited were fed back to respondents to verify their agreement with the severity ordering of the health states that was inferred from their responses.

## 2.3 Data Collection

The protocol suggests sample sizes for the DCE and cTTO tasks of  $N = 1000$  and  $N = 200$ , respectively [12]. Discrete choice experiment data were collected using a paper-based survey, face-to-face, to reach people with low educational backgrounds who usually had limited access to the Internet. A 1-day workshop was held to train eight DCE interviewers

**Table 1** Block design for time trade-off

Block 1	Level sum score	Severity	Orthogonal	Valuation protocol	Block 2	Level sum score	Severity	Orthogonal	Valuation protocol
11112	6	Mild	No	Yes	11121	6	Mild	No	Yes
11113	7	Mild	Yes	No	21131	8	Moderate	Yes	No
23111	8	Moderate	Yes	No	12123	9	Moderate	Yes	No
32211	9	Moderate	Yes	No	31221	9	Moderate	Yes	No
13212	9	Moderate	Yes	No	11232	9	Moderate	Yes	No
13321	10	Moderate	Yes	No	12331	10	Moderate	Yes	No
21322	10	Moderate	Yes	No	22312	10	Moderate	Yes	No
22232	11	Moderate	No	Yes	22223	11	Moderate	Yes	Yes
31313	11	Moderate	Yes	No	32132	11	Moderate	Yes	No
33122	11	Moderate	Yes	No	32232	12	Moderate	No	No
33332	14	Severe	No	No	23233	13	Severe	Yes	No
33333	15	Severe	Yes	Yes	33333	15	Severe	Yes	Yes

The choice of the health states was based on the five health states of the protocol, 18 health were chosen on the basis of estimation of an orthogonal design (of which some were already included in the minimal protocol sample), and two severe health states were added, as the ends of the scale tended to be under-represented by orthogonal designs. Given that 33333 is in both blocks of 12, we end up with 23 unique health states

for data collection. After the workshop, interviewers conducted a pilot interview with three respondents with various socio-economic backgrounds. We determined the minimum interview time and daily interview target for interviewers based on the pilot. Data collection was held between 4 August, 2019 and 14 November, 2019. Interviewers were divided into two teams and moved from one city to another. Respondents received Rp 35.000 (equal to 2 euros) for their participation.

Training for the cTTO interviewers consisted of three sections: (1) experiencing the cTTO interview both as interviewer and respondent; (2) theory; and (3) pilot. Three pilot rounds were conducted with five respondents per round per interviewer. Their experiences were discussed, and feedback was given on their performance. After all the interviewers met sufficient performance levels, data collection commenced. Composite time trade-off data collection was conducted between 6 August, 2019 and 2 September, 2019. Local people brought the respondents to the interview site. By conducting the data collection together, interviewers could share their experiences daily and improve their performances. Feedback was given every ten interviews. Respondents received Rp 100.000 (equal to 6 Euros) for their participation. The samples used for the DCE and cTTO were mutually exclusive and collected independently. Ethical approval was obtained from the Health Research Ethics Committee, YARSI University, Indonesia (117/KEP-UY/BIA/VII/2019).

## 2.4 Respondents

Multi-stage stratified quota sampling was used to select a representative sample for the collection of DCE responses.

Quotas were defined by: living area (urban/rural), age (17–30 years/31–50 years/above 50 years), sex (male/female) and level of education (primary/middle/high). The combination of these characteristics resulted in 36 quota groups. Over the complete sample, thus independently of the 36 quotas, we took the proportion of religion (Islam/Christian/Others) and living place (Jawa/Sumatera/Sulawesi/Kalimantan/Other) into account as a second stage. Respondents should have spent at least half of their life on a particular island to represent the local societies. The predefined quotas were based on the Indonesian Bureau of Statistics [19].

The DCE interviews were conducted in 14 cities and their rural surroundings, located on different islands in Indonesia. Interviews were held in public places such as mosques, universities, and markets. For the cTTO, data collection took place on Java, the most populous island in Indonesia, in one city (Jakarta) and one rural area (Bandung district). Respondents were selected based on their living area (urban/rural) and sex (male/female).

## 2.5 Quality Control

The cTTO data were collected using EuroQol Portable Valuation Technology software, obtained from the EuroQol Group, to present the cTTO task and store the data. [12]. The EuroQol Portable Valuation Technology software is similar to the EQ-VT software, but is constructed using Microsoft Powerpoint, and is able to run without being connected to the Internet. Therefore, it is easier to use in areas where Internet connectivity is limited. The data produced were monitored in real-time following the EuroQol quality control procedure, in order to provide interviewers with feedback on their performances and to promote data quality. The quality control

process consisted of minimum quality criteria and cyclical feedback to improve interviewers' skills, as described by Ramos-Goñi et al. [20]. Weekly meetings were organised to discuss the quality control reports with the EQ-VT support team. The minimum quality criteria for each interview were:

1. Time spent in the wheelchair example should be at least 3 min.
2. The interviewer should explain the worse than death element in one of the wheelchair examples.
3. On average, at least 30 s should be spent on each cTTO task.
4. The value for the state '33333' should be the lowest or less than 0.5 higher than the state with the lowest value.

If any of these criteria were not met, the interview was 'flagged' as an indication of poor quality. Feedback was given to improve interviewer performance. The interviewer could terminate the interview or exclude respondents in the following circumstances. (i) The interview was terminated if respondents were unable to understand the cTTO task during the practice session, (ii) data were excluded from the analysis if respondents refused to trade any life-years for children. Although non-trading could reflect true preferences of responders, it could also indicate that responders did not comprehend the task as intended. Because of the ambiguity revealed in the responses, we decided to exclude these from the data. (iii) The interview was terminated if respondents were unable to differentiate between the levels of severity on the dimensions. The quality control report was prepared once per week.

In this study, the DCE survey was administered via face-to-face interviews. Based on our trial, it was considered implausible that respondents could validly respond to the DCE trade-offs in less than 300 s. Interviews displaying less time than this minimum duration were evaluated to check whether they were administered thoroughly. To make respondents familiar with the DCE task, two DCE practice tasks were provided at the beginning of the questionnaire. The first was a choice pair of common chronic diseases in Indonesian children, namely asthma and malnutrition. We framed the child's health using the EQ-5D-Y-3L problem and dimension levels. The second practice task was the EQ-5D-Y-3L choice pair 21121 vs 11122. If respondents understood the practice task, the interview proceeded to the DCE task. To check for respondents' attentiveness, we placed one dominant pair between DCE tasks. If respondents failed to give a logical answer, the interviewer confirmed their answer and re-explained the instruction. If the illogical response persisted, such responses were excluded from the DCE analysis.

## 2.6 Analysis

The analysis consisted of a two-step approach. First, we modelled the DCE responses to produce values on a latent scale using a mixed-logit model. The mixed-logit model assesses unobserved preference heterogeneity by allowing the model parameters to vary across individuals [21]. Subsequently, we estimated a mapping function to map the DCE derived values onto the observed mean cTTO values.

We explored two methods of anchoring the DCE data onto the cTTO data: (1) linear mapping and (2) non-linear mapping. The mapping employed a regression model to model the relationship between the predicted DCE values and the mean observed cTTO values for each health state included in the cTTO design. Two different mapping models were applied: a linear mapping strategy that took the form of Eq. 1, and a power model taking the form of Eq. 2.

$$V_{\text{cTTO}_k} = \gamma_0 + \gamma_1 V_{\text{DCE}_k} \quad (1)$$

$$V_{\text{cTTO}_k} = \gamma_0 + (\gamma_1 V_{\text{DCE}_k})^\delta \quad (2)$$

Here,  $V_{\text{cTTO}_k}$  is the mean observed value for health state  $k$  in the cTTO and  $V_{\text{DCE}_k}$  is the predicted value in the DCE models for the same health state  $k$ .  $\gamma_0$  is the regression intercept and  $\gamma_1$  is the slope between the cTTO and DCE values.  $\delta$  is a power parameter that may adjust for any possible non-linearity in the relation between the cTTO and DCE data. Each of these models was estimated including and excluding the regression intercept  $\gamma_0$ . To ensure ease of analysis, the cTTO and DCE data were structured as 'disutilities' rather than utilities, which meant that both of these data sets were scaled with 0 being the lowest value, and all other values being positive and indicating greater disutility.

## 2.7 Model Selection

Model selection was evaluated based on the logical consistency and degree of agreement between predicted and observed cTTO values as assessed by the mean absolute error. Logical consistency implies that models should not show any inconsistent patterns in the estimated coefficients, for example, level 3 mobility being assigned less disutility than level 2 mobility. Further, we expected the models to produce an intercept close to the value range of a QALY (0.0 = value of dead to 1.0 = value of full health). For example, if the intercept predicted a value for the best health states and mild health states to be above 1.0, this would be considered a misprediction. A lower mean absolute error would indicate better accuracy of the model.

**Table 2** Characteristics of study respondents compared to the general population

Characteristics	cTTO study sample <i>N</i> = 222 <i>n</i> (%)	DCE study sample <i>N</i> = 1072 <i>n</i> (%)	Indonesian general population, %	Differences between DCE sample and general population, %
<b>Residence</b>				
Urban	112 (50.5)	536 (50)	53.30	+ 3.3
Rural	110 (49.5)	536 (50)	46.70	− 3.3
<b>Sex</b>				
Female	113 (50.9)	532 (49.6)	49.65	+ 0.05
Male	109 (49.1)	540 (50.4)	50.35	− 0.05
<b>Age, years</b>				
17–30	88 (39.6)	423 (39.5)	36.08	+ 3.42
31–50	86 (38.7)	431 (40.2)	40.85	− 0.65
> 50	48 (21.6)	211 (19.7)	23	+ 3.3
Missing	–	7 (0.7%)	–	–
<b>Education</b>				
Low	10 (4.5)	527 (49.2)	50	+0.8
Middle	160 (72.1)	372 (34.7)	35	+0.3
High	52 (23.4)	171 (16)	15	− 1
<b>Living places</b>				
Java	222 (100)	607 (56.6)	56.81	+ 0.21
Sumatera	–	244 (22.8)	21.63	− 1.17
Sulawesi	–	81 (7.6)	7.33	− 0.27
Kalimantan	–	60 (5.6)	6	+ 0.4
Others	–	80 (7.5)	8.2	+ 0.7
<b>Religion</b>				
Islam	216 (97.3)	925 (86.3)	87.18	+ 0.88
Christian	6 (2.7)	103 (9.6)	9.86	+ 0.26
Others	–	44 (4.1)	2.96	− 1.14

cTTO composite time trade-off, DCE discrete choice experiment

### 3 Results

#### 3.1 Respondent Characteristics

There were 1090 respondents who completed the DCE interview, and 228 respondents who completed the cTTO interview. We excluded 18 DCE respondents as they failed the logically dominant pair check. Six cTTO respondents were excluded, two because of non-trading and four as they were unable to understand the task.

Characteristics of the final sample were similar to the Indonesian population in terms of residence, sex, age and education, with an absolute difference of less than 5%. In the final sample, respondents' living places and religion were representative of the general population (see Table 2).

#### 3.2 DCE Model

Table 3 shows the result from the DCE mixed-logit model. The coefficients were logically ordered, with all coefficients

statistically significant at the 0.01 level. Pain/discomfort received the largest weight, followed by mobility, usual activities, worried/sad/unhappy and self-care. The overall fit as assessed by Akaike Information Criteria and Bayesian Information Criteria shows that all dimensions (levels) indicated a wide variation in preferences among respondents.

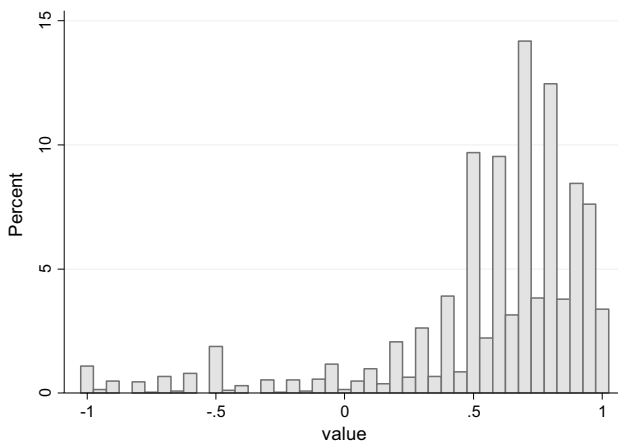
#### 3.3 Composite TTO Data Characteristics

The 222 respondents provided 2664 cTTO observations. In the feedback module, the respondents pointed out 222 (8.3%) responses that they felt in hindsight did not validly represent their preferences. There were four responses (0.1%) with value 0.0, and 238 (8.9%) negative values (Fig. 1). Seventy-six percent of these negative values were concentrated in health states 33332 and 33333. The observed means of the 23 cTTO values ranged from − 0.185 for state 33333 to 0.939 for state 11121. Figure 2 shows the mean observed value for the 23 health states included in the cTTO task.

**Table 3** Mixed-logit discrete choice experiment model

Level	Coefficients mixed logit	SE	SD	p-value
Mobility 2	- 1.257	0.072	0.850	0.000
Mobility 3	- 2.161	0.103	0.943	0.000
Self-care 2	- 0.970	0.066	0.969	0.000
Self-care 3	- 1.711	0.083	1.014	0.000
Usual activities 2	- 1.375	0.060	1.048	0.000
Usual activities 3	- 1.997	0.080	1.265	0.000
Pain/discomfort 2	- 1.198	0.055	0.978	0.000
Pain/discomfort 3	- 2.173	0.083	1.411	0.000
Worried/sad/unhappy 2	- 1.224	0.059	1.069	0.000
Worried/sad/unhappy 3	- 1.924	0.078	1.541	0.000
AIC	17257.43			
BIC	17802.03			

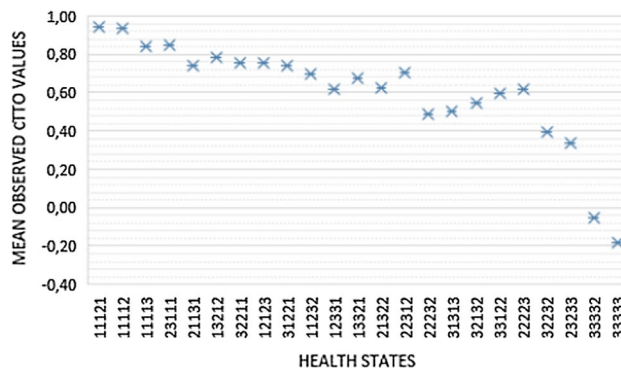
AIC Akaike Information Criteria, BIC Bayesian Information Criteria, SD standard deviation, SE standard error



**Fig. 1** Distribution of composite time trade-off values

**Anchoring Results**

Table 4 presents the results of four different mapping functions used to define the relationship between DCE-derived values and cTTO-derived values. Figure 3 shows the correspondence between the mapped DCE-derived values and the observed cTTO values for all tested mapping functions. The linear models provided reasonable results for most health states, but correspondence at the extremes of the scale was a concern (i.e. for mild and poor health states). The non-linear models provided a markedly better overall fit, as indicated by higher R-squares and a lower mean absolute error in Table 3 and as shown in Fig. 3. The power model with a constant slightly outperformed the power model without a constant in terms of the fit statistics. Nevertheless, we preferred the latter because the power model with a constant showed a stronger misprediction for the milder health states



**Fig. 2** Mean observed composite time trade-off (cTTO) values over the health states

(see Fig. 3c, d). Milder health states are commonly observed in the general population and patient population [22–24], hence precise estimation for the milder health states was prioritised over precise estimates for more uncommon severe states.

We thus recommend mapping the mixed-logit model to the cTTO data using non-linear mapping (power without constant) as the preferred model for the value set for EQ-5D-Y-3L in Indonesia. This meant that the coefficients of the latent class model were multiplied by the rescaling parameter of the power model without a constant, 0.1048, and the power of 1.9013 was taken of the sum of these rescaled coefficients, and subsequently subtracted from 1, the value for full health. This led to the following algorithm for the value set:

$$U = 1 - (0.1317 MO2 + 0.2265 MO3 + 0.1017 SC2 + 0.1793 SC3 + 0.1441 UA2 + 0.2093 UA3 + 0.1256 PD2 + 0.2277 PD3 + 0.1283 AD2 + 0.2016 AD3)^{1.9013}$$

For health state 12312, this meant that the value would be:  $U = 1 - (0.1017 + 0.2093 + 0.1283)^{1.9013}$ , which equals 0.791.

The final model ranged from 1.000 for full health (11111) to - 0.086 for the worst health state (33333). The model shows that the most important dimension was pain/discomfort, and the least important was self-care.

**4 Discussion**

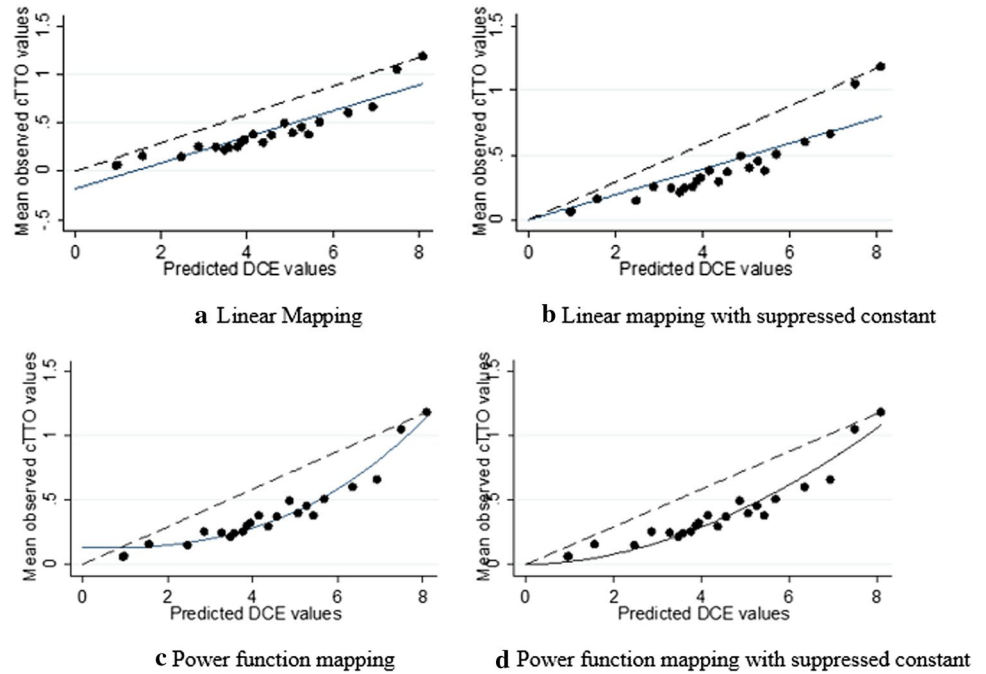
Our study aimed to generate an EQ-5D-Y-3L value set for Indonesia using the standard EuroQol valuation protocol for EQ-5D-Y-3L, with an extension to the number of cTTO health states. We collected DCE and cTTO responses, using the DCE responses as a basis for developing the value set and the cTTO responses for anchoring. Because the DCE

**Table 4** Estimation results for linear and non-linear mapping

	Linear mapping	Linear mapping (no constant)	Power mapping	Power mapping (no constant)
Intercept	- 0.1804	-	0.1302	-
Coefficient	0.1079	0.0794	0.1012	0.1048
Power	-	-	2.7363	1.9013
R-squared	0.827	0.926	0.9856	0.9773
MAE	0.0853	0.0916	0.0458	0.0598

MAE mean absolute error

**Fig. 3** Scatterplots for mapping strategies. Dashed lines spring from the origin to the worst health state. Solid lines represent the mapping strategies. *cTTO* composite time trade-off, *DCE* discrete choice experiment



responses were obtained on a latent scale, they were mapped onto the *cTTO* responses. This paper reports upon two types of mapping functions tested to transform *DCE* responses: linear mapping and non-linear mapping (power model). As we had more than twice the number of *cTTO* health states than the standard protocol, it was clearly visible that a non-linear mapping function without a constant was the preferable model, based on a good model fit and adequate performance on all parts of the severity scale.

One of the big challenges that we faced for the joint analysis of *DCE* and *cTTO* was the non-linearity observed in our *cTTO* data. In the published EQ-5D-Y-3L valuation studies, linearity was observed in the distribution of mean *cTTO* values [25-28]. Thus, joint analysis can be undertaken in a variety of ways including mapping, hybrid modelling or anchoring on 33333. In our study, we had fewer options. The limited number of *cTTO* data points available, particularly for severe states where the distribution bends the most, made it hard to identify what model specification, in this case of

the mapping function, was appropriate. The amount of *cTTO* data was too limited to serve as a basis for a value set, both because the sample was not representative and because the selection of *cTTO* states was tailored to estimation of the rejected main-effect additive linear model. Hybrid modeling was not an option because of the imbalance in the number of *DCE* ( $1072 \times 15 = 16080$ ) and *cTTO* ( $222 \times 10 = 2220$ ) observations. Thus, the hybrid model yielded an estimation that almost totally relies on the *DCE* model (see Appendix 1), which did not reflect the scale use of the respondents in the *cTTO* task. Mapping strategies, linear and non-linear, appeared to be viable anchoring strategies instead, and the non-linear mapping appeared to predict the relation between the *cTTO* and *DCE* data with the smallest error. Choosing non-linear mapping is not without drawbacks. Under the assumption that *DCE* and *cTTO* produce interval scales, only linear transformations are allowed to maintain the interval scale properties for the combined *cTTO/DCE* scale. However, the linear mapping approach either produced some



values above 1 for mild health states or mispredicted the severe states by a difference of 0.4, which can be considered as too large. Thus, although our final model violated the interval scale assumption, we considered this as a preferable solution to increase the fit between the DCE and cTTO data. The ordering of the estimated health states on the utility scale was preserved under the transformation chosen to anchor the health states on 0.0 as the value of dead and 1.0 as the value of full health. It was helpful that we utilised 23 instead of ten cTTO health states, as this better marked the differences between the linear and non-linear mapping. Even with 23 health states, critics may argue that two outlier observations supply a narrow basis for our decision.

A broader issue raised by our results is a concern about the commensurability of the cTTO-derived and DCE-derived values, as these methods produced different results that forced a choice between the two methods. The decision to select the mapping function that best approximated all observed cTTO values was a matter of judgement. One reason for making cTTO values the target is that cTTO values reflect a direct trade-off between time and quality of life, similar to the QALY. With DCE, we must assume that the 'values' derived from DCE relate to time in the same way, while to date, evidence is lacking that this has indeed been the case. More reflection is needed on reasons why the DCE and cTTO results did diverge, and on possible implications for a protocol requiring input from both methods.

This study had several limitations. First, although we adopted a larger cTTO design than suggested [14], we would have benefited from a larger design. When the study was planned, it was expected that our cTTO design would enable us to produce a cTTO value set without using the DCE data [29], and we could then avoid the discussion above. However, this assumed that a main effects model would be adequate; the cTTO design was too small to explore other model specifications. The distribution of cTTO responses showed a remarkable pattern where the value for the two worst states was much lower than the other observed values (see Fig. 2). Unfortunately, where the curves bent the most, few cTTO values were gathered. Second, participants who completed the DCE and cTTO tasks were sampled in a different way. Age, sex, residence, education level, religion and geographical location were considered in the DCE sampling approach, whereas all cTTO data were sampled in one location (Java) using quotas based on age and sex. In this respect, the DCE values were more representative of the Indonesian adult population.

Comparing our findings with the recent adult EQ-5D-5L valuation in Indonesia, the negative values on the cTTO observed scores of the EQ-5D-Y-3L valuation were much smaller (2.5% compared with 35.4%) [30]. The dimension importance was also different: the most important dimension for children was pain/discomfort, while for adults, pain/discomfort was the least important. Because there has been

a limited number of youth valuation studies published to date, we do not yet know whether this result was specific for the Indonesian population. In any case, these differences between adult and youth values justify separate value sets.

## 5 Conclusions

This article presents the first EQ-5D-Y-3L value set for Indonesia that employs an adult perspective. Mapping the mixed-logit DCE model including a power term (without a constant) allowed us to generate a consistent Indonesian youth value set. Our findings support the expansion of the cTTO section of the EQ-5D-Y valuation study design and show that it would be wise to account for possible non-linearities in updates of the design.

## Appendix 1

Level	Coefficients	SE	<i>p</i> value
Hybrid modelling			
Mobility 2	0.085	0.004	0.000
Mobility 3	0.164	0.005	0.000
Self-care 2	0.064	0.004	0.000
Self-care 3	0.125	0.004	0.000
Usual activities 2	0.096	0.004	0.000
Usual activities 3	0.147	0.004	0.000
Pain/Discomfort 2	0.083	0.003	0.000
Pain/Discomfort 3	0.163	0.005	0.000
Worried/sad/unhappy 2	0.085	0.003	0.000
Worried/sad/unhappy 3	0.137	0.004	0.000
DCE model			
Mobility 2	0.085	0.004	0.000
Mobility 3	0.164	0.005	0.000
Self-care 2	0.064	0.004	0.000
Self-care 3	0.125	0.004	0.000
Usual activities 2	0.096	0.004	0.000
Usual activities 3	0.147	0.004	0.000
Pain/discomfort 2	0.083	0.003	0.000
Pain/discomfort 3	0.163	0.005	0.000
Worried/sad/unhappy 2	0.085	0.003	0.000
Worried/sad/unhappy 3	0.137	0.004	0.000
TTO model			
Mobility 2	0.066	0.012	0.000
Mobility 3	0.232	0.012	0.000
Self-care 2	0.020	0.012	0.000
Self-care 3	0.172	0.012	0.000
Usual activities 2	0.052	0.012	0.000
Usual activities 3	0.227	0.012	0.000
Pain/discomfort 2	0.088	0.013	0.000

Level	Coefficients	SE	p value
Pain/discomfort 3	0.334	0.012	0.000
Worried/sad/unhappy 2	0.088	0.012	0.000
Worried/sad/unhappy 3	0.252	0.016	0.000

*DCE* discrete choice experiment, *SE* standard error, *TTO* time trade-off

**Acknowledgements** The authors thank the EuroQol Group (EQ Project 20180490) and Indonesian Endowment Fund for Education (LPDP) for providing funding for this study. The first author thanks Rina Rahmatika, Cindi Anggraini and all the hardworking interviewers for their endless energy during data collection. Last but not least, we thank all the members of EuroQol Younger Population Working Group for every fruitful discussion. Opinions expressed in this paper are those of the authors and are not necessarily those of the EuroQol Research Foundation.

## Declarations

**Funding** Funding for this study was provided by the EuroQol Group grant EQ Project 20180490 and Indonesian Endowment Fund for Education (LPDP). Both grants were unrestricted.

**Conflicts of Interest/Competing Interests** Titi Fitriana reports no conflicts of interest. Fredrick Purba, Bram Roudijk, Elly Stolk and Jan Busschbach are either members of the EuroQol Group or are directly employed by the EuroQol Research Foundation.

**Ethics Approval** All procedures performed in this study were in accordance with the ethical standards of the Health Research Ethics Committee of YARSI University, Jakarta, Indonesia (ethics approval number: 117/KEP-UY/BIA/VII/2019).

**Consent to Participate** All participants enrolled in this study gave their consent to participate in this study.

**Consent for Publication** All participants understand the data obtained from the study will be used only in educational publications and no personal identification will be attached to the publication. A copy of the consent form is available for review by the Editor of this journal.

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

**Code Availability** Not applicable.

**Authors' Contributions** TF, ES, FP and JB designed the study, obtained funding and were involved in the data acquisition. Quality control and analysis were performed by BR, TF and FP. TF and BR drafted the initial version of the manuscript, with all authors providing substantial feedback.

**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. Wang A, Rand K, Yang Z, et al. *The remarkably frequent use of EQ-5D in non-economic research*. Eur J Health Econ. 2021; p. 1–8.
2. Brazier J, Ara R, Rowen D, et al. *A review of generic preference-based measures for use in cost-effectiveness models*. Pharmacoeconomics. 2017;35(1):21–31.
3. National Institute for Health and Care Excellence (NICE). Centre for Health Technology Evaluation (CHTE) methods review—task and finish group specifications. 2019 [cited 22 Augustus 2022]. <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/chte-methods-consultation>.
4. Wolowacz SE, Briggs A, Belozeroff V, et al. *Estimating health-state utility for economic models in clinical studies: an ISPOR good research practices task force report*. Value Health. 2016;19(6):704–19.
5. Ungar WJ. *Challenges in health state valuation in paediatric economic evaluation*. Pharmacoeconomics. 2011;29(8):641–52.
6. Rowen D, Rivero-Arias O, Devlin N, et al. *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; p. 1–16.
7. Wu XY, Ohinmaa A, Johnson JA, et al. *Assessment of children's own health status using visual analogue scale and descriptive system of the EQ-5D-Y: linkage between two systems*. Qual Life Res. 2014;23(2):393–402.
8. Ratcliffe J, Couzner L, Flynn T, et al. *Valuing child health utility 9D health states with a young adolescent sample*. Appl Health Econ Health Policy. 2011;9(1):15–27.
9. Michael SY, Britto MT, Wilmott RW, et al. *Health values of adolescents with cystic fibrosis*. J Pediatr. 2003;142(2):133–40.
10. Stolk EA, Oppe M, Scalone L, et al. *Discrete choice modeling for the quantification of health states: the case of the EQ-5D*. Value Health. 2010;13(8):1005–13.
11. Rowen D, Brazier J, Van Hout B. *A comparison of methods for converting DCE values onto the full health-dead QALY scale*. Med Decis Making. 2015;35(3):328–40.
12. Ramos-Goñi JM, Oppe M, Stolk E, et al. *International Valuation Protocol for the EQ-5D-Y-3L*. Pharmacoeconomics. 2020; p. 1–11.
13. Ramos-Goñi JM, Pinto-Prades JL, Oppe M, et al. *Valuation and modeling of EQ-5D-5L health states using a hybrid approach*. Med Care. 2017;55(7):e51–8.
14. Wille N, Badia X, Bonsel G, et al. *Development of the EQ-5D-Y: a child-friendly version of the EQ-5D*. Qual Life Res. 2010;19(6):875–86.
15. Jonker MF, Donkers B, de Bekker-Grob EW, et al. *Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments*. Value Health. 2018;21(7):767–71.
16. Janssen BM, Oppe M, Versteegh MM, et al. *Introducing the composite time trade-off: a test of feasibility and face validity*. Eur J Health Econ. 2013;14(1):5–13.
17. Oppe M, van Hout B. The “power” of eliciting EQ-5D-5L values: the experimental design of the EQ-VT. . 2017 [cited 01 September

- 2022; <https://euroqol.org/wp-content/uploads/2016/10/EuroQol-Working-Paper-Series-Manuscript-17003-Mark-Oppe.pdf>.
18. Yang Z, Luo N, Bonsel G, et al. *Selecting health states for EQ-5D-3L valuation studies: statistical considerations matter*. Value Health. 2018;21(4):456–61.
  19. Statistik BP. *Result of Population Census 2010*. 2012. BPS-Statistics Indonesia.
  20. Ramos-Goñi JM, Oppe M, Slaap B, et al. *Quality control process for EQ-5D-5L valuation studies*. Value Health. 2017;20(3):466–73.
  21. Eberth B, Watson V, Ryan M, et al. *Does one size fit all? Investigating heterogeneity in men's preferences for benign prostatic hyperplasia treatment using mixed logit analysis*. Med Decis Making. 2009;29(6):707–15.
  22. Scott D, Ferguson GD, Jelsma J. *The use of the EQ-5D-Y health related quality of life outcome measure in children in the Western Cape, South Africa: psychometric properties, feasibility and usefulness—a longitudinal, analytical study*. Health Qual Life Outcomes. 2017;15(1):12.
  23. Wong CKH, Cheung PWH, Luo N, et al. *A head-to-head comparison of five-level (EQ-5D-5L-Y) and three-level EQ-5D-Y questionnaires in paediatric patients*. Eur J Health Econ. 2019; p. 1–10.
  24. Szende A, Janssen B, Cabases J. *Self-reported population health: an international perspective based on EQ-5D*. London: Springer Open; 2014.
  25. Rupel VP, Ogorevc M. *EQ-5D-Y Value Set for Slovenia*. Pharmacoeconomics. 2021;39(4):463–71.
  26. Shiroiwa T, Ikeda S, Noto S, et al. *Valuation survey of EQ-5D-Y based on the international common protocol: development of a value set in Japan*. Med Decis Making. 2021;41(5):597–606.
  27. Ramos-Goñi JM, Oppe M, Estévez-Carrillo A, 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.
  28. Kreimeier S, Mott D, Ludwig K, et al., *EQ-5D-Y value set for Germany*. Pharmacoeconomics. 2022; p. 1–13.
  29. Lamers LM, McDonnell J, Stalmeier PFM, et al. *The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies*. Health Econ. 2006;15(10):1121–32.
  30. Purba FD, Hunfeld JAM, Iskandarsyah A, et al. *The Indonesian EQ-5D-5L value set*. Pharmacoeconomics. 2017;35(11):1153–65.