



ScienceDirect

Contents lists available at [sciencedirect.com](http://sciencedirect.com)  
Journal homepage: [www.elsevier.com/locate/jval](http://www.elsevier.com/locate/jval)

## Preference-Based Assessments

# 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

Juan M. Ramos-Goñi, PhD, Mark Oppe, PhD, Anabel Estévez-Carrillo, MSc, Oliver Rivero-Arias, DPhil, on behalf of the IMPACT HTA HRQoL Group

## ABSTRACT

**Objectives:** The EuroQol Group published the EQ-5D-Y valuation protocol that recommends 2 valuation techniques to elicit preferences: composite time trade-off (C-TTO) and discrete choice experiments (DCEs). The protocol left the decision of what modeling approach to use open for researchers. Our aims were to explore modeling strategies allowing generation of EQ-5D-Y value sets and to produce an EQ-5D-Y Spanish value set.

**Methods:** We used EQ-5D-Y DCE and C-TTO data collected in Spain following the protocol and adopted a staged approach for our modeling exercise. First, we selected the best performing DCE latent class model and evaluated models from 2 to 10 classes. We selected the preferred model based on best goodness of fit in terms of the Bayesian information criterion. We considered 2 anchoring approaches to estimate utility values: (1) pits state anchoring and (2) hybrid models (using all available C-TTO responses). All analysis were weighted to be representative of the Spanish population.

**Results:** We collected 1005 DCE and 200 C-TTO interviews. We selected a DCE model including 4 classes. Hybrid models using all available C-TTO observations produced a narrower range of values than the pits state anchoring approach.

**Conclusions:** In this article, we have presented an EQ-5D-Y value set that can be used for cost-utility analysis in Spain. The international EQ-5D-Y valuation protocol should be updated to include a different set of health states for the C-TTO experiment if researchers wish to use alternative anchoring approaches to the “pits state.”

**Keywords:** composite time trade-off, discrete choice experiment, EQ-5D-Y, health-related quality of life, value set.

VALUE HEALTH. 2021; ■(■):■-■

## Introduction

An international valuation protocol for the EuroQol youth instrument EQ-5D-Y has recently been published,<sup>1</sup> and research teams have completed or are actively collecting country-specific data using this protocol. Therefore, country-specific value sets for EQ-5D-Y health states for cost-utility analyses in child and adolescent populations are becoming available.<sup>2,3</sup> The EQ-5D-Y valuation protocol was created using a series of international studies in a collaborative effort<sup>4-6</sup> in a similar fashion to the valuation protocol for the 5-level EQ-5D (EQ-5D-5L). It recommends the use of discrete choice experiments (DCEs) and a composite time trade-off (C-TTO) to elicit preferences for EQ-5D-Y health states. Nevertheless, there are significant differences compared with the EQ-5D-5L valuation protocol. For instance, adult participants answer the valuation task under the following perspective “Considering your views about a 10-year-old child, what do you prefer?”. In addition, the sample size for each elicitation task is different (1000 responses for DCE and 200 for C-TTO), and independent samples are used. The idea is that

the DCE informs the relative importance of dimensions and levels and that the only role of the C-TTO is to anchor the latent scale DCE results onto the quality-adjusted life-year (QALY) scale. Nonetheless, the protocol does not provide guidance about how to analyze DCE responses or what anchoring method should be used, leaving such decision open to researchers.

Modeling strategies for the latent scale DCE should focus on the appropriate implementation of a model that represents particular choice behaviors. The conditional multinomial logit is considered the workhorse and is the starting point of many health preference researchers to analyze DCE data.<sup>7</sup> The multinomial logit represents choice behaviors well in many situations, but it relies on the assumptions of independence from irrelevant alternatives and homogeneity of preferences.<sup>8</sup> Alternative discrete choice models can relax these assumptions helping researchers to understand different types of preference heterogeneity, with unobservable preference heterogeneity receiving most of the attention.<sup>9</sup> Unobservable preference heterogeneity across respondents is the variation not explained by observable characteristics of

respondents. Ignoring this type of heterogeneity when it is present biases the coefficients of a discrete choice model.<sup>10</sup> The reality is that most applications to date have found strong evidence that unobservable heterogeneity is always present and that models that account for this perform better than the multinomial logit model.<sup>9</sup> Because of this, it is recommended that health preference researchers use models that account for unobservable preference heterogeneity when producing value sets. Over the past 20 years, discrete choice analysts have developed a battery of models that account for this including the mixed logit<sup>11</sup> and latent class models.<sup>12</sup> Nevertheless, the question of which one to use in practice depends on the specific objectives of the analysis, that is, either researchers use these models because they are interested in data fit (eg, better Bayesian information criterion [BIC] for mixed logit and latent class models) or because they are interested in understanding the impact of unobservable heterogeneity on the estimated values. In this study, we are interested in both understanding the unobservable heterogeneity present in our data and accounting for it when estimating the values; therefore, we report the results of implementing the latent class models to estimate a latent scale value set for DCE data obtained using the international valuation protocol for the EQ-5D-Y using a representative sample of members of the general public in Spain. Nevertheless, for comparison purposes, we also report the results of the logit and mixed logit models.

Another aspect that received little attention in the protocol is how to anchor latent scale DCE values onto the QALY scale using the gathered C-TTO information.<sup>13–15</sup> We aim to further clarify this issue and explore different approaches to rescale DCE values onto the QALY scale for the estimation of a value set using this new protocol.

The aim of this study is to produce an EQ-5D-Y value set to be used in the Spanish setting for the conduct of cost-utility analysis in pediatric and young populations.

## Methods

### Protocol

The EQ-5D-Y valuation protocol was used in this study and the reader is referred to the main publication for full details.<sup>1</sup> Briefly, an online DCE and a face-to-face C-TTO were the elicitation techniques used to obtain preferences for EQ-5D-Y states. The DCE design included 150 pairs divided into 10 blocks (15 pairs per block). This design was estimated in Stata using a D-efficiency maximization procedure without specifying any previous information as it aimed to create a valid design for any country.<sup>1</sup> This DCE design used a 2-dimension overlap for all pairs, where the 2 health states descriptions included in each pair have different levels in 3 dimensions whereas the other 2 remain at the same level. For quality control purpose, we included 3 fixed dominant pairs (same for all participants) in each block yielding a total of 18 pairs per respondents.

The C-TTO experimental design was simpler and consisted of 1 single block of 10 health states, 3 mild, 2 moderate, and 5 severe health states (mild, 11112, 11121, 21111; moderate, 22223, 22232; and severe, 31133, 32223, 33233, 33323, 33333).

### DCE Interview

The DCE was completed online and included the following elements:

1. Information sheet with the aims of the study and requesting participant's consent. If a respondent did not provide consent,

the survey was terminated and only showed a message thanking the respondent for considering participation.

2. Demographic questions including geographical region, age, and gender used to delimit quotas and ensure sample representativeness
3. Self-reported health with EQ-5D-Y instrument as a warm-up task
4. The 3 questions related to experience with illness
5. The 15 + 3 DCE tasks. Each participant completed first a fixed dominant pair. Next the block of 15 + 1 pairs including 1 fixed dominant pair, all presented in a random order. Lastly the third fixed dominant pair. The left-right positions of the health states within a pair were also randomized. To mitigate the occurrence of nonattendance, level differences between health states were emphasized using bold fonts.
6. Self-reported health with EQ-5D-Y-5L instrument<sup>16</sup>
7. Additional background questions, including whether the responses would be different if the health states were experienced by the respondent themselves instead of by a 10 years old child, whether a child should have priority over an adult in a limited health resources context, education status, employment status, whether the respondent has worked with children or whether the respondent has children, whether the respondent knows a child that has experienced a serious illness, and whether the respondent is suffering from any disease.

### Time Trade-Off Interview

Participants of the C-TTO experiment completed a face-to-face interview with same structure as the DCE interview except for element 5, which was replaced with the following:

1. Two wheelchair examples (to allow interviewer to explain the C-TTO task)
2. Three practice states (to allow participants to practice alone before the real tasks)
3. The main 10 C-TTO states (same for all participants) presented in random order
4. The standard feedback module as used in EQ-5D-5L valuation studies<sup>17</sup>

### Sampling, Data Collection, and Quality Control

#### DCE

A target sample size of 1000 participants as suggested by the protocol was followed in this study for the DCE component. Spanish sample representativeness was defined in terms of Spanish regions, age, and gender obtained from the National Institute of Statistics.<sup>18</sup> A panel company was employed to recruit participants to complete the survey. The company distributed individual links to participants who accessed the online DCE. As quality control for data collection, we established 2 rules for determining whether a participant was sufficiently engaged while completing the survey. First, we excluded speeders by excluding participants who did not invest at least 2 minutes and 30 seconds in completing the DCE part of the survey. Second, we excluded participants who incorrectly answered at least 2 of the 3 dominant pairs. Participants included in the final sample received points from the panel company as a reward for their participation.

#### C-TTO

A target sample size of 200 participants as suggested by the protocol was followed in this study for the C-TTO component. Spanish

**Table 1.** Sample description and comparison with Spanish population.

| Variables                      | DCE                |                       | TTO                  | Spanish general pop. (%) <sup>*</sup> |
|--------------------------------|--------------------|-----------------------|----------------------|---------------------------------------|
|                                | Excluded (n = 434) | Estimation (n = 1005) | Estimation (n = 200) |                                       |
| Age, mean (SD)                 | 41.0 (12.1)        | 46.2 (13.5)           | 43.4 (15.1)          | 43.3 (NA)                             |
| Age groups                     |                    |                       |                      |                                       |
| 18-24                          | 36.0 (8.3)         | 58.0 (5.8)            | 25.0 (12.5)          | 8.4                                   |
| 25-29                          | 46.0 (10.6)        | 62.0 (6.2)            | 19.0 (9.5)           | 6.6                                   |
| 30-39                          | 115.0 (26.5)       | 184.0 (18.3)          | 36.0 (18.0)          | 15.9                                  |
| 40-49                          | 140.0 (32.3)       | 287.0 (28.6)          | 54.0 (27.0)          | 20.1                                  |
| 50-59                          | 68.0 (15.7)        | 240.0 (23.9)          | 34.0 (17.0)          | 18.0                                  |
| 60-69                          | 23.0 (5.3)         | 134.0 (13.3)          | 19.0 (9.5)           | 13.6                                  |
| 70+                            | 6.0 (1.4)          | 40.0 (4.0)            | 13.0 (6.5)           | 17.3                                  |
| Gender                         |                    |                       |                      |                                       |
| Male                           | 264.0 (60.8)       | 496.0 (49.4)          | 88.0 (44.0)          | 49.0                                  |
| Female                         | 170.0 (39.2)       | 509.0 (50.6)          | 112.0 (56.0)         | 51.0                                  |
| Employment status              |                    |                       |                      |                                       |
| Employed or freelance          | 333.0 (76.7)       | 644.0 (64.1)          | 103.0 (51.5)         | 85.9                                  |
| Retired                        | 20.0 (4.6)         | 103.0 (10.2)          | 17.0 (8.5)           | 4.5                                   |
| Student                        | 19.0 (4.4)         | 49.0 (4.9)            | 30.0 (15)            | 3.5                                   |
| Housewife/house husband        | 25.0 (5.8)         | 87.0 (8.7)            | 18.0 (9)             | 1.9                                   |
| Disabled                       | 6.0 (1.4)          | 28.0 (2.8)            | 5.0 (2.5)            | 2.1                                   |
| None                           | 21.0 (4.8)         | 71.0 (7.1)            | 27.0 (13.5)          | 1.9                                   |
| Missing                        | 10 (2.3)           | 23 (2.2)              | —                    | —                                     |
| Education                      |                    |                       |                      |                                       |
| Study after minimum age        | 372.0 (85.7)       | 889.0 (88.5)          | 182.0 (91.0)         | 59.2                                  |
| University degree              | 295.0 (68.0)       | 643.0 (64.0)          | 119.0 (59.5)         | 34.0                                  |
| Experience with illness        |                    |                       |                      |                                       |
| Personal (% yes)               | 135.0 (31.1)       | 324.0 (32.2)          | 67.0 (33.5)          | NA                                    |
| Relatives (% yes)              | 246.0 (56.7)       | 640.0 (63.7)          | 163.0 (81.5)         | NA                                    |
| Others (% yes)                 | 108.0 (24.9)       | 261.0 (26.0)          | 64.0 (32.0)          | NA                                    |
| Self-reported EQ-5D-Y          |                    |                       |                      |                                       |
| Mobility                       |                    |                       |                      |                                       |
| No problems                    | 358.0 (82.5)       | 872.0 (86.8)          | 175.0 (87.5)         | 86.1                                  |
| Some problems                  | 70.0 (16.1)        | 124.0 (12.3)          | 23.0 (11.5)          |                                       |
| A lot of problems              | 6.0 (1.4)          | 9.0 (0.9)             | 2.0 (1.0)            |                                       |
| Self-care                      |                    |                       |                      |                                       |
| No problems                    | 396.0 (91.2)       | 958.0 (95.3)          | 193.0 (96.5)         | 93.9                                  |
| Some problems                  | 33.0 (7.6)         | 46.0 (4.6)            | 6.0 (3.0)            |                                       |
| A lot of problems              | 5.0 (1.2)          | 1.0 (0.1)             | 1.0 (0.5)            |                                       |
| Usual activities               |                    |                       |                      |                                       |
| No problems                    | 361.0 (83.2)       | 859.0 (85.5)          | 179.0 (89.5)         | 92.2                                  |
| Some problems                  | 67.0 (15.4)        | 126.0 (12.5)          | 20.0 (10.0)          |                                       |
| A lot of problems              | 6.0 (1.4)          | 20.0 (2.0)            | 1.0 (0.5)            |                                       |
| Pain/discomfort                |                    |                       |                      |                                       |
| No pain or discomfort          | 301.0 (69.4)       | 695.0 (69.2)          | 150.0 (75.0)         | 75.2                                  |
| Some pain or discomfort        | 125.0 (28.8)       | 293.0 (29.2)          | 48.0 (24.0)          |                                       |
| A lot of pain or discomfort    | 8.0 (1.8)          | 17.0 (1.7)            | 2.0 (1.0)            |                                       |
| Anxiety/depression             |                    |                       |                      |                                       |
| Not worried, sad, or unhappy   | 218.0 (50.2)       | 429.0 (42.7)          | 124.0 (62.0)         | 85.4                                  |
| A bit worried, sad, or unhappy | 186.0 (42.9)       | 507.0 (50.4)          | 72.0 (36.0)          |                                       |
| Very worried, sad, or unhappy  | 30.0 (6.9)         | 69.0 (6.9)            | 4.0 (2.0)            |                                       |

DCE indicates discrete choice experiment; EQ-5D-5L, 5-level EQ-5D; NA, not available; pop., population; TTO, time trade-off.

<sup>\*</sup>Data extracted from the National Institute of Statistics. Percentage of no problems using self-reported EQ-5D-5L, collected in the 2012 to 2013 National Health Survey.

sample representativeness was defined in terms of age and gender using quotas. A different market research company was hired to conduct the face-to-face interviews; nevertheless, because of the coronavirus disease 2019 (COVID-19) pandemic, 77 interviews were conducted remotely using video conference facilities. Our interviewers had previous experience conducting C-TTO interviews,<sup>4,19</sup> but they also received a 1-day training course for this specific study. Initial participants were invited to a central location for being interviewed in a face-to-face setting. The participants who were interviewed by video conferencing were safely at home while the

interviewers were at home as well. We followed the standard quality control procedure developed by the EuroQol for C-TTO studies, which mainly focus on interviewer's behavior and between interviewer variability.<sup>20</sup> Participants received a €10 voucher as a reward for their participation.

### Statistical Analysis

Sample characteristics were described using proportions for each categorical variable. Differences between our sample and the

**Table 2.** C-TTO descriptive results.

| Profile | Observed |            | Censoring-adjusted |            |
|---------|----------|------------|--------------------|------------|
|         | Mean     | Std. error | Mean               | Std. error |
| 11112   | 0.989    | 0.003      | 0.989              | 0.002      |
| 11121   | 0.969    | 0.004      | 0.971              | 0.003      |
| 21111   | 0.986    | 0.003      | 0.987              | 0.002      |
| 22223   | 0.604    | 0.028      | 0.603              | 0.021      |
| 22232   | 0.221    | 0.041      | 0.183              | 0.038      |
| 31133   | −0.039   | 0.044      | −0.086             | 0.045      |
| 32223   | 0.445    | 0.035      | 0.401              | 0.030      |
| 33233   | −0.217   | 0.042      | −0.318             | 0.045      |
| 33323   | 0.012    | 0.042      | −0.063             | 0.042      |
| 33333   | −0.389   | 0.039      | −0.539             | 0.050      |

C-TTO indicates composite time trade-off; Std., standard.

Spanish general population structure were handled using age and gender weighted analysis.

Observed C-TTO values are reported as the population weighted means and standard errors of the observed C-TTO values for each of the 10 health states included in the design. In addition, given that C-TTO values are by construction censored at  $-1$ , we also estimated the population weighted censoring-adjusted mean value for each of the 10 health states using regression methods.<sup>21</sup> A total of 10 separate weighted Tobit models censored at  $-1$  (ie, separate regressions for each health state) with observed C-TTO values as the dependent variable and a constant as independent variable were estimated using the calculated population weights.

DCE responses were analyzed using latent class models (lclogit2 Stata command) that account for unobservable heterogeneity of preferences. A main effects specification using incremental dummies, where estimated coefficients represented movement between consecutive levels within dimensions, was estimated in each model. For each model, we estimated the accuracy in predicting the observed choice probabilities and calculated the mean square error (MSE) and mean absolute error (MAE) of the observed choice probabilities against the model predicted probabilities. These were then plotted using scatter plots of observed versus predicted probabilities. Goodness of fit of each model specification was assessed according to the BIC. The appropriate number of classes in the latent class model was determined using information criteria. We evaluated different number of classes ranging from 2 to  $N$  classes and select the  $N-1$  number of classes conditional on the BIC of the model with  $N$  classes being higher than of the model with  $N-1$  classes.<sup>22</sup> In addition, we estimated the standard multinomial logit (clogit Stata command) and the random coefficient mixed logit models (mixlogit Stata command) to show latent class modeling worked better in this data.

The “best” latent class model candidate was then used to test 2 anchoring methods of DCE latent scale utilities onto the 1-full health 0-dead scale: (1) anchoring on the pits state utility value and (2) a hybrid model where DCE and C-TTO information is estimated in a single model. For convenience, approaches (1) and (2) were estimated using an adapted version of the hyreg command.<sup>23</sup> The standard hyreg was adapted to be able to incorporate weights. We used the class grades of the latent class DCE models adjusted (multiplied) by the population weights for age and gender, therefore estimating a latent class weighted hybrid model.<sup>24</sup>

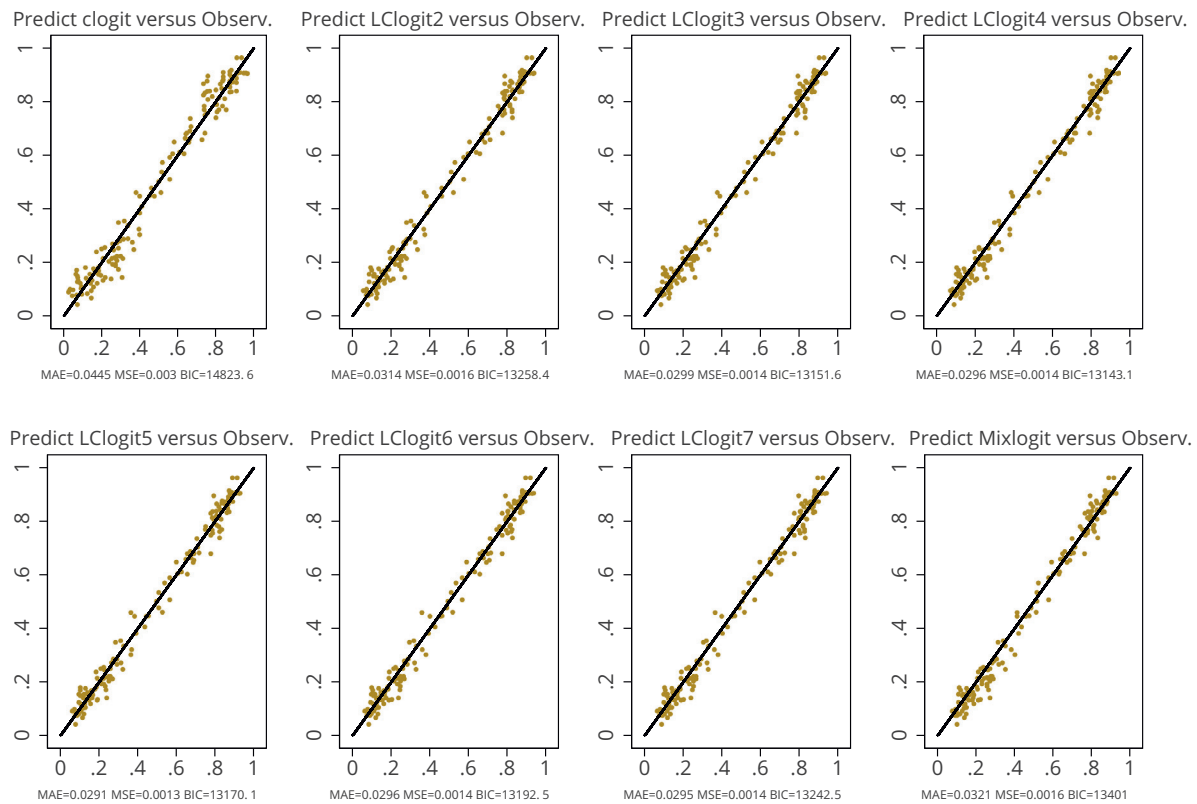
The first anchoring approach “anchoring on the pits state utility value” consisted of estimating a hybrid model based on the DCE latent class analysis and including only the C-TTO observations for the pits state (33333) (referred to below as hybrid with only 33333 C-TTO value). The set of coefficients estimated by this approach is equivalent to that based on a linear rescaling of the DCE scale by using the pits state value, as described by Stolk et al.<sup>25</sup> and Ramos-Goñi et al.<sup>26</sup> We first estimated the best DCE latent class model; this included the model coefficients, but also the class grades associated to each respondent within each class and the class shares proportions. These DCE respondents’ class grades were adjusted by the population weights. The C-TTO information was weighted as the mean class grade of the classes also adjusted by the population weights. Once all data weighted, we estimated a weighted hybrid model per each class censoring C-TTO values at  $-1$ . In summary, we had available population weights and respondent class grades from the latent class models and classes shares. Class grades and population weights are multiplied and used as final weights into the hybrid models’ estimations per class. The class shares were adjusted before being used to weigh each class set of coefficients when calculating the overall hybrid model. This further adjustment is needed because each class model was rescaled within the hybrid models estimation (nonlinear transformation); therefore, their original latent scale was no longer available. Therefore, the weights for the overall model should be the class shares adjusted by the latent scale within class. This adjustment was calculated as latent-pits-value-class-model/latent-pits-value-overall-model per each class. The second anchoring approach was the same but included all C-TTO values for the 10 health states included in the C-TTO design (referred to below as hybrid with all C-TTO values). To compare the anchoring approaches, we have plotted the predictions together with the observed C-TTO censoring-adjusted values in a scatter plot including MSEs.

To calculate the standard error of the final value set we bootstrapped participants, simulating independently C-TTO and DCE samples and per each pair of samples, we repeated the preferred anchoring approach as explained earlier. The final selected value set is presented using regular dummies instead of incremental ones as it is commonly reported in the literature. All analysis was performed in Stata MP version 14 and 15.

## Results

### Descriptive Statistics

We collected a sample of 1439 participants on the DCE. Of those, 434 participants were removed following the quality control exclusion criteria, 21 because of 2 or more inconsistent responses (to the dominant pairs), 348 consistent speeders, and 65 inconsistent speeders. Therefore, our final DCE sample consisted in 1005 respondents passing the quality control criteria. The C-TTO sample consisted of 200 respondents. No participants were excluded in C-TTO experiment; nevertheless, we excluded from our analysis the responses that respondents flagged in the feedback module. Overall, the DCE and time trade-off (TTO) estimation samples were similar in the distribution of age, gender, employment status, and education to the Spanish population (Table 1). The self-reported health using the EQ-5D-Y of respondents showed that 30% reported problems in pain and discomfort and 53% reported problems in anxiety or depression dimension, which is a larger proportion than the Spanish population norms. In addition, most of respondents (55%) stated that they would not change their responses if the health states were being experienced by themselves instead of by a 10-year-old child and 55.7% of

**Figure 1.** DCE models comparison.

BIC indicates Bayesian information criterion; DCE, discrete choice experiment; MAE, mean absolute error; MSE, mean square error.

respondents stated that in a limited resources context both adults and children should have same priority to receive healthcare (Appendix Table 1 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.10.013>).

The observed TTO mean for the pits health state (33333) was  $-0.389$ , whereas the censoring-adjusted mean reached  $-0.539$  (Table 2). Comparing mean values for the health states 11112 and 11121 and the values for the health states 33233 and 33323 showed that the pain/discomfort dimension was considered more important by the C-TTO participants.

### Modeling Results

All the latent class models with 2 to 7 classes resulted in lower MSE, MAE, and BIC than the multinomial and random parameters mixed logit (Fig. 1). A latent class model with 4 classes exhibited the lowest BIC and was selected as the final model to evaluate the different anchoring techniques. Looking at within-class models structure, the first class was composed of the 48.9% of the population. They were in strong agreement with respect to their preferences and gave most of importance (31.2%) to pain/discomfort dimension followed by being worry/unhappy (24.8%) and usual activities (18.3%) and then similar importance to mobility or looking after yourself (13.3%-12.5%, respectively); the second class was composed of the 29.3% of the population and they focused on the importance in pain/discomfort (31.5%) and mobility (25.8%); the third class was composed of 9% of the population who mainly focused on the importance of mobility, giving 53.2% of the relative importance of all dimensions; finally, the fourth class formed by 13% is composed of people with no

homogeneous preferences (Appendix Fig. 1 in Supplemental Materials found at <https://doi.org/10.1016/j.jval.2021.10.013>).

The 2 hybrid anchoring approaches also produced models with consistent parameters results (Table 3). The hybrid model including all C-TTO information predicted a value for state 33333 of  $-0.364$  whereas the hybrid including only the C-TTO pits states predicted the value of state 33333 as the C-TTO observed censoring-adjusted value of 33333 (ie,  $-0.539$ ). The relative importance of the pain dimension in the hybrid model including all C-TTO values is higher than in the hybrid including only 33333 values, meaning that C-TTO information gave even more importance to pain/discomfort than the DCE. Nevertheless, the ranking by importance of dimension was the same for both models. The within-classes estimation of the hybrid including all C-TTO values produced different estimations for 33333 in the different classes (Table 3).

Comparing the predictions between hybrid models and observed C-TTO data, we can see that although the hybrid model including only the C-TTO value for 33333 underestimates all censoring-adjusted C-TTO observations, it matched the estimation for the 33333. Nevertheless, the hybrid model including all C-TTO health states overestimated the worse than death C-TTO values and underestimated the better than death C-TTO values. MSE for the hybrid model including all health states was lower than the MSE for the model anchored on the pits state only. Nevertheless, there were 2 health states causing this (22223 and 22232). When removing those 2 health states from the MSE calculations predictions from the model anchored on the pits state only were better than from the hybrid model using all health states (Fig. 2).

Finally, the preferred value set using both regular and incremental dummies is presented in Table 4. As shown in the table by



**Table 3.** Results from different anchoring methods.

| EQ-5D-Y dimension                | Model parameters                   | Hybrid model including all 10 C-TTO HS (Incremental dummies) |                    |                       |                      |              | Hybrid model including C-TTO 33333 HS only (Incremental dummies) |                    |                       |                      |              |
|----------------------------------|------------------------------------|--------------------------------------------------------------|--------------------|-----------------------|----------------------|--------------|------------------------------------------------------------------|--------------------|-----------------------|----------------------|--------------|
|                                  |                                    | CL1                                                          | CL2                | CL3                   | CL4                  | Overall      | CL1                                                              | CL2                | CL3                   | CL4                  | Overall      |
|                                  | Unadjusted class share             | 48.90%                                                       | 29.30%             | 9.20%                 | 12.60%               |              | 48.50%                                                           | 29.60%             | 9.30%                 | 12.60%               |              |
|                                  | Scale adjusted class share         | 65.16%                                                       | 26.66%             | 5.19%                 | 2.99%                |              | 65.16%                                                           | 26.66%             | 5.19%                 | 2.99%                |              |
|                                  |                                    | Coeff (std. error)                                           | Coeff (std. error) | Coeff (std. error)    | Coeff (std. error)   | Coeff *      | Coeff (std. error)                                               | Coeff (std. error) | Coeff (std. error)    | Coeff (std. error)   | Coeff *      |
| Mobility (walking about)         | No problems to some problems       | 0.055 (0.011)                                                | 0.106 (0.011)      | 0.252 (0.027)         | 0.157 (0.04)         | <b>0.082</b> | 0.076 (0.071)                                                    | 0.125 (0.015)      | 0.283 (0.052)         | 0.212 (0.071)        | <b>0.104</b> |
|                                  | Some problems to a lot of problems | 0.113 (0.009)                                                | 0.248 (0.012)      | 0.471 (0.05)          | <u>0.042 (0.044)</u> | <b>0.166</b> | 0.128 (0.065)                                                    | 0.273 (0.022)      | 0.537 (0.097)         | <u>0.03 (0.065)</u>  | <b>0.185</b> |
| Looking after myself             | No problems to some problems       | 0.029 (0.009)                                                | 0.041 (0.011)      | 0.081 (0.025)         | 0.099 (0.039)        | <b>0.037</b> | 0.044 (0.062)                                                    | 0.052 (0.012)      | 0.089 (0.032)         | 0.143 (0.062)        | <b>0.051</b> |
|                                  | Some problems to a lot of problems | 0.135 (0.009)                                                | 0.122 (0.011)      | 0.162 (0.027)         | <u>0.079 (0.043)</u> | <b>0.131</b> | 0.149 (0.061)                                                    | 0.134 (0.015)      | 0.185 (0.041)         | <u>0.08 (0.061)</u>  | <b>0.145</b> |
| Doing usual activities           | No problems to some problems       | 0.098 (0.008)                                                | 0.047 (0.01)       | <u>-0.049 (0.029)</u> | 0.22 (0.044)         | <b>0.08</b>  | 0.121 (0.095)                                                    | 0.057 (0.011)      | <u>-0.058 (0.036)</u> | 0.309 (0.095)        | <b>0.100</b> |
|                                  | Some problems to a lot of problems | 0.148 (0.008)                                                | 0.148 (0.011)      | 0.193 (0.029)         | <u>0.062 (0.038)</u> | <b>0.148</b> | 0.161 (0.053)                                                    | 0.16 (0.015)       | 0.221 (0.045)         | <u>0.063 (0.053)</u> | <b>0.161</b> |
| Having pain or discomfort        | No to some                         | 0.159 (0.008)                                                | 0.135 (0.01)       | 0.064 (0.024)         | 0.17 (0.04)          | <b>0.148</b> | 0.185 (0.079)                                                    | 0.153 (0.015)      | 0.071 (0.03)          | 0.237 (0.079)        | <b>0.172</b> |
|                                  | Some to a lot of                   | 0.293 (0.016)                                                | 0.312 (0.018)      | 0.133 (0.027)         | 0.142 (0.044)        | <b>0.285</b> | 0.295 (0.066)                                                    | 0.332 (0.028)      | 0.138 (0.035)         | 0.169 (0.066)        | <b>0.293</b> |
| Feeling worried, sad, or unhappy | Not to a bit                       | 0.121 (0.008)                                                | 0.07 (0.01)        | <u>0.03 (0.025)</u>   | <u>0.037 (0.039)</u> | <b>0.101</b> | 0.139 (0.055)                                                    | 0.079 (0.012)      | <u>0.03 (0.03)</u>    | <u>0.046 (0.055)</u> | <b>0.114</b> |
|                                  | A bit to very                      | 0.212 (0.01)                                                 | 0.155 (0.011)      | <u>0.03 (0.026)</u>   | 0.186 (0.045)        | <b>0.186</b> | 0.242 (0.084)                                                    | 0.174 (0.017)      | <u>0.043 (0.03)</u>   | 0.252 (0.084)        | <b>0.214</b> |
|                                  | <b>U(33333)</b>                    | -0.362                                                       | -0.385             | -0.368                | -0.194               | -0.364       | -0.539                                                           | -0.539             | -0.539                | -0.539               | -0.539       |

Note. Presented are the models for each of the 4 latent classes, and the aggregate model. In addition, the share of each class and the predicted utility for the pits state U(33333). Overall model coefficients were boldfaced. Underlined results were nonsignificant at 95% confidence level.

Coeff indicates coefficient; C-TTO, composite time trade-off; HS, health state; std., standard.

\*std error and P values of the overall model are computed for the final value set using bootstrap procedures (see Table 4).

the incremental dummies, all movements between consecutive levels are consistent and significantly different from 0. The most important dimension is pain/discomfort, followed by anxiety/depression, mobility, usual activities, and self-care.

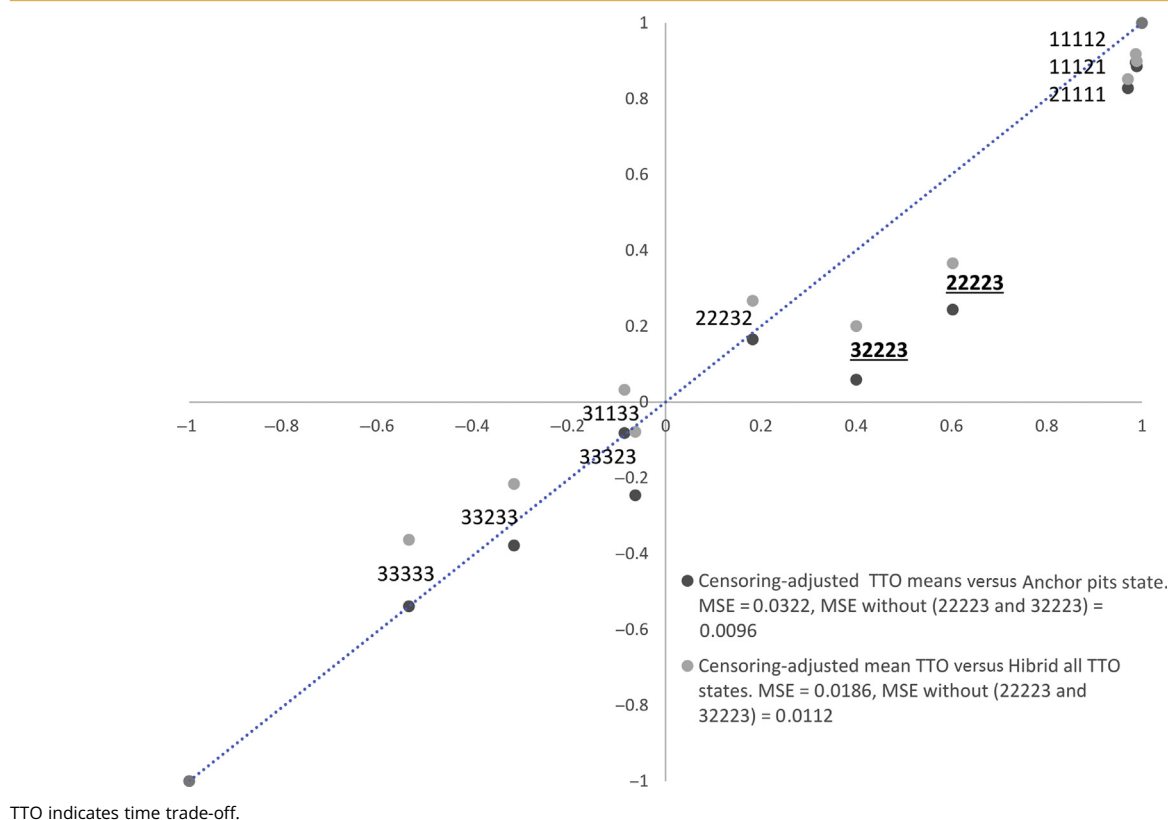
## Discussion

In this study, we applied the recent international valuation protocol for the EQ-5D-Y<sup>1</sup> to obtain a value set for the EQ-5D-Y in Spain. As specified in the protocol, we included a DCE to obtain the relative importance of the dimensions and levels and a C-TTO task to allow anchoring of the DCE on the QALY scale. We found that a latent class model with 4 classes had the best performance. The structure and composition of these 4 classes, as explained in the Results section, clearly indicate that population preferences regarding child health are not homogeneous, at least for the Spanish population, but likely in other countries as well. The use of the latent class analysis allowed us to understand the heterogeneity in our data. In addition, our results showed that the latent class model outperformed the mixed logit model with respect to model predictions, as measured by MAE and MSE and goodness of fit based on BIC. Therefore, we recommend using latent class models for this type of modeling exercise. Nevertheless, whether or not the latent class models outperform other types of models should be confirmed, rather than assumed.

We believe that both a hybrid approach and an anchoring on the pits state approach as described in this article, independently

of the QALY scale anchored valuation technique (TTO or visual analog scale [VAS]), are all feasible techniques to generate a value set in general.<sup>6</sup> When focusing on child health valuation, the use of TTO alone has been previously put into question,<sup>4</sup> directing the research into a combination of techniques as finally suggested by the EQ-5D-Y protocol article.<sup>1</sup> Among the potential external anchors, several alternatives were evaluated including anchoring in pits VAS values as described by Webb et al,<sup>27</sup> anchoring with TTO as described by Stolk et al,<sup>25</sup> and hybrid model.<sup>23</sup> The published EQ-5D-Y protocol discarded the VAS method as external anchor as choice-based methods have been recommended by decision makers in some settings including the National Institute for Health and Care Excellence in England.<sup>28</sup> Nevertheless, the protocol did not provide recommendation about the choice of modeling technique.

Although technically possible, the use of a hybrid model to anchor the potential EQ-5D-Y value set should be taken with caution. The set of 10 health states included in the C-TTO task were not selected for estimation of a model. They were selected to reduce biases in the data collection, such as framing effects, that would cause scaling issues.<sup>29</sup> The standard set of health states used in the C-TTO of the EQ-5D-Y protocol will result in estimates of the values at the bottom of the scale that are too high and will therefore provide an anchoring that is upwardly biased, leading to a compressed QALY scale for the value set, which may have a significant impact on allocating resources decision making. We have identified 2 health states (22223 and 32223), which may have caused issues with the design of the TTO. If the use of a hybrid model is desired for anchoring in a future valuation study,

**Figure 2.** Anchor approaches comparison.

then the states selected for the C-TTO should be amended by selecting the health states specifically optimized for estimating a C-TTO (or hybrid) model. These issues led us to select the anchor in pits state model as the final value set because it better represents the QALY scale of the Spanish population regarding child health.

It should be noted that this study is not free of limitations. The data were collected during the COVID-19 outbreak, which may have affected our results. Some DCE interviews were conducted before and some during the lockdown in Spain. In addition, some of the C-TTO interviews were conducted in a face-to-face setting (before the lockdown), whereas others were conducted via

**Table 4.** Spanish EQ-5D-Y value set.

| Dimension                        | Regular dummies                  | Coefficient (value set) | Std. error | Incremental dummies                | Coefficient | Std. error | P-value |
|----------------------------------|----------------------------------|-------------------------|------------|------------------------------------|-------------|------------|---------|
| Mobility (walking about)         | No problems to some problems     | 0.1040                  | 0.0109     | No problems to some problems       | 0.1040      | 0.0109     | .0024   |
|                                  | No problems to a lot of problems | 0.2892                  | 0.0160     | Some problems to a lot of problems | 0.1851      | 0.0124     | .0006   |
| Looking after myself             | No problems to some problems     | 0.0513                  | 0.0125     | No problems to some problems       | 0.0513      | 0.0125     | .0268   |
|                                  | No problems to a lot of problems | 0.1959                  | 0.0159     | Some problems to a lot of problems | 0.1446      | 0.0185     | .0043   |
| Doing usual activities           | No problems to some problems     | 0.1002                  | 0.0130     | No problems to some problems       | 0.1002      | 0.0130     | .0045   |
|                                  | No problems to a lot of problems | 0.2609                  | 0.0065     | Some problems to a lot of problems | 0.1607      | 0.0110     | .0007   |
| Having pain or discomfort        | No to some                       | 0.1719                  | 0.0141     | No to some                         | 0.1719      | 0.0115     | .0006   |
|                                  | No to a lot of                   | 0.4647                  | 0.0115     | Some to a lot of                   | 0.2928      | 0.0207     | .0008   |
| Feeling worried, sad, or unhappy | Not to a bit                     | 0.1144                  | 0.0267     | Not to a bit                       | 0.1144      | 0.0110     | .0019   |
|                                  | Not to very                      | 0.3285                  | 0.0110     | A bit to very                      | 0.2141      | 0.0137     | .0006   |

Std. indicates standard.

videoconference (after the lockdown was lifted 2.5 months later). We are not able to determine whether the pandemic and lockdown might have affected the health preferences of the population. There is some evidence suggesting that preferences may have changed because of the pandemic in the United Kingdom,<sup>30</sup> and it would be expected that similar changes have occurred in other countries. Future research should explore whether this is the case.

A second limitation was the fact that we collected the DCE data online. It has been shown that online administration of DCE for health preference research has limitations when it comes to engagement with the task of the respondents.<sup>31</sup> This can be reflected by speeding through the task and by providing random responses. To reduce the impact of this limitation, we included some proxy measures for the respondent engagement, namely, 3 dominant pairs and timings between choices made.

### Further Research

The small difference in values for the pit state of the EQ-5D-5L (−0.416) and the EQ-5D-Y (−0.539) we found, together with the fact that the majority of respondents (>55%) indicated that adults and children should have equal priority when it comes to decisions on healthcare resource allocation, might make a case for anchoring the DCE data of the EQ-5D-Y on the pit state of the value set in use, that is, the EQ-5D-5L in the Spanish context. Using national adult value sets, such as EQ-5D-5L or 3-level EQ-5D value sets to provide the anchors for the QALY scale would make it much easier to obtain EQ-5D-Y value sets, given that there would be no need for collecting C-TTO data using interviewers, but only conducting an online DCE. This would be especially beneficial for developing countries, but only for those with enough information technology penetration and for developed countries in the age of COVID-19. Of course, our results will have to be confirmed by studies in other countries before such an approach should be adopted.

### Conclusions

In this article, we have presented an EQ-5D-Y value set that can be used for cost-utility analysis in the Spanish setting. The used international EQ-5D-Y valuation protocol should be updated to include a different set of health states for the C-TTO experiment if researchers wish to move away from “pits state” anchoring approach.

### Supplemental Material

Supplementary data associated with this article can be found in the online version at <https://doi.org/10.1016/j.jval.2021.10.013>.

### Article and Author Information

**Accepted for Publication:** October 11, 2021

**Published Online:** xxxx

doi: <https://doi.org/10.1016/j.jval.2021.10.013>

**Author Affiliations:** Maths in Health, Rotterdam, The Netherlands (Ramos-Goñi, Oppe, Estévez-Carrillo, Rivero-Arias); Department of Health Economics and Health Care Management, School of Public Health, Bielefeld University, Bielefeld, Germany (Ramos-Goñi); National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, England, UK (Rivero-Arias).

**Correspondence:** Juan M. Ramos-Goñi, PhD, Maths in Health, Librijestee 307, 3011HN, Rotterdam, the Netherlands. Email: [juanmanuel.ramosgoni@gmail.com](mailto:juanmanuel.ramosgoni@gmail.com)

**Author Contributions:** *Concept and design:* Ramos-Goñi

*Acquisition of data:* Ramos-Goñi, Estévez-Carrillo

*Analysis and interpretation of data:* Ramos-Goñi, Oppe, Estévez-Carrillo, Rivero-Arias

*Drafting of the manuscript:* Ramos-Goñi, Oppe, Estévez-Carrillo, Rivero-Arias  
*Critical revision of the paper for important intellectual content:* Ramos-Goñi, Oppe, Estévez-Carrillo, Rivero-Arias

*Statistical analysis:* Ramos-Goñi, Oppe, Estévez-Carrillo, Rivero-Arias

*Provision of study materials or patients:* Ramos-Goñi, Estévez-Carrillo

*Obtaining funding:* Ramos-Goñi

*Administrative, technical, or logistic support:* Ramos-Goñi

*Supervision:* Ramos-Goñi

**Conflict of Interest Disclosures:** All authors reported receiving grants from the EuroQol Research Foundation during the conduct of the study. Drs Ramos-Goñi and Rivero-Arias reported receiving grants from the European Commission during the conduct of this study. Drs Ramos-Goñi, Oppe, and Rivero-Arias reported being members of the EuroQol Group. No other disclosures were reported. The views expressed are those of the authors and not necessarily those of the European Commission or the EuroQol Group.

**Funding/Support:** This research was funded by the European Union's Horizon 2020 research and innovation program (H2020) and undertaken under the auspices of IMPACT HTA (grant number 779312; <http://www.impact-hta.eu>) and EuroQol Research Foundation (EQ project number 20180510). The ethical approvals for the study were obtained from the Comité Ético de Investigación Clínica, Hospital Nuestra Señora del Prado, number Dictamen 32/19.

**Role of the Funder/Sponsor:** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

**Acknowledgment:** IMPACT HTA HRQoL Group: Greiner Wolfgang (Department of Health Economics and Health Care Management, School of Public Health, Bielefeld University, Bielefeld, Germany); Kreimeier Simone (Department of Health Economics and Health Care Management, School of Public Health, Bielefeld University, Bielefeld, Germany); Ludwig Kristina (Department of Health Economics and Health Care Management, School of Public Health, Bielefeld University, Bielefeld, Germany); Rupel Valentina (Institute for Economic Research, Ljubljana, Slovenia). The authors want to express our gratitude to our funders (European Commission and EuroQol Research foundation); to Carlos Herrera-Carballo for his contribution in developing the surveys used to collect the data and to our C-TTO interviewers who conducted all interviews with the most rigor to produce good quality C-TTO data. Finally, our biggest gratitude goes to the respondents who participated in either the discrete choice or the composite time trade-off experiments.

### REFERENCES

1. Ramos-Goñi JM, Oppe M, Stolk E, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics*. 2020;38(7):653–663.
2. Prevolnik Rupel V, Ogorevc M, IMPACT HTA HRQoL Group. EQ-5D-Y value set for Slovenia. *Pharmacoeconomics*. 2021;39(4):463–471.
3. 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.
4. Kreimeier S, Oppe M, Ramos-Goñi JM, 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):1291–1298.
5. Mott DJ, Shah KK, Ramos-Goñi JM, Devlin NJ, Rivero-Arias O. Valuing EQ-5D-Y-3L health states using a discrete choice experiment: do adult and adolescent preferences differ? *Med Decis Making*. 2021;41(5):584–596.
6. Shah KK, Ramos-Goñi JM, Kreimeier S, Devlin NJ. An exploration of methods for obtaining 0 = dead anchors for latent scale EQ-5D-Y values. *Eur J Health Econ*. 2020;21(7):1091–1103.
7. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, ed. *Frontiers in Econometrics*. New York, NY: Academic Press; 1973:105–142.



8. Train KE. *Discrete Choice Methods With Simulation*. Cambridge, United Kingdom: Cambridge University Press; 2003.
9. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis*. Cambridge, United Kingdom: Cambridge University Press; 2015.
10. Hess S, Rose JM. Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation*. 2012;39(6):1225–1239.
11. McFadden D, Train K. Mixed MNL models for discrete response. *J Appl Econ*. 2000;15(5):447–470.
12. Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transp Res B Methodol*. 2003;37(8):681–698.
13. Keane M, Wasi N. Comparing alternative models of heterogeneity in consumer choice behavior. *J Appl Econ*. 2013;28:1018–1045.
14. Norman R, Mulhern B, Viney R. The impact of different DCE-based approaches when anchoring utility scores. *Pharmacoeconomics*. 2016;34(8):805–814.
15. 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):328–340.
16. Kreimeier S, Åström M, Burström K, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. *Qual Life Res*. 2019;28(7):1951–1961.
17. Wong ELY, Ramos-Goñi JM, Cheung AWL, Wong AYK, 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):235–247.
18. Data by subject: population by sex, autonomous communities and age (up to 100 and over). National Statistics Institute. <https://www.ine.es/>. Accessed September 2, 2021.
19. Ramos-Goñi JM, Rand-Hendriksen K, Pinto-Prades JL. Does the introduction of the ranking task in valuation studies improve data quality and reduce inconsistencies? The case of the EQ-5D-5L. *Value Health*. 2016;19(4):478–486.
20. Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJV, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value Health*. 2017;20(3):466–473.
21. Ramos-Goñi JM, Craig BM, Oppe M, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value Health*. 2018;21(5):596–604.
22. Pacifico D, Yoo H il. Lclogit: a Stata command for fitting latent-class conditional logit models via the expectation-maximization algorithm. *Stata J*. 2013;13(3):625–639.
23. Ramos-Goñi JM, Craig BM, Oppe M, Van Hout B. Combining continuous and dichotomous responses in a hybrid model. EuroQol Working Paper Series. [https://euroqol.org/wp-content/uploads/working\\_paper\\_series/EuroQol\\_Working\\_Paper\\_Series\\_Manuscript\\_16002\\_-\\_Juan\\_Ramos-Goni.pdf](https://euroqol.org/wp-content/uploads/working_paper_series/EuroQol_Working_Paper_Series_Manuscript_16002_-_Juan_Ramos-Goni.pdf). Accessed September 2, 2021.
24. Andrade LF, Ludwig K, Goni JMR, Oppe M, de Povourville G. A French value set for the EQ-5D-5L. *Pharmacoeconomics*. 2020;38(4):413–425.
25. Stolk EA, Oppe M, Scalone L, Krabbe PFM. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value Health*. 2010;13(8):1005–1013.
26. Ramos-Goñi JM, Rivero-Arias O, Errea M, Stolk EA, Herdman M, Cabasés JM. Dealing with the health state 'dead' when using discrete choice experiments to obtain values for EQ-5D-5L health states. *Eur J Heal Econ*. 2013;14(suppl 1):33–42.
27. Webb E, O'Dwyer J, Meads D, Kind P, Wright P. Transforming discrete choice experiment latent scale values for EQ-5D-3L using the visual analogue scale. *Eur J Heal Econ*. 2020;21(5):787–800.
28. Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence. <https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781>. Accessed September 2, 2021.
29. Morrison GC, Neilson A, Malek M. Improving the sensitivity of the time trade-off method: results of an experiment using chained TTO questions. *Health Care Manag Sci*. 2002;5(1):53–61.
30. Webb E, Kind P, Meads D, Martin A. Does a health crisis change how we value health? *Health Econ*. 2021;30(10):2547–2560.
31. Mulhern B, Longworth L, Brazier J, et al. Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value Health*. 2013;16(1):104–113.