VALUE IN HEALTH 21 (2018) 596–604

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

PREFERENCE-BASED ASSESSMENTS

Handling Data Quality Issues to Estimate the Spanish EQ-5D-5L Value Set Using a Hybrid Interval Regression Approach

Juan M. Ramos-Goñi, MSc[1,8,], Benjamin M. Craig, PhD[2], Mark Oppe, PhD[1], Yolanda Ramallo-Fariña, MSc[3,8], Jose Luis Pinto-Prades, PhD[4,5], Nan Luo, PhD[6], Oliver Rivero-Arias, PhD[7,8]*

[1]EuroQol Research Foundation, Rotterdam, The Netherlands; [2]University of South Florida, Tampa, FL, USA; [3]HTA Unit of Canary Islands Health Service, Canary Islands, Spain; [4]University of Navarra, Navarra, Spain; [5]Glasgow Caledonian University, Glasgow, UK; [6]National University of Singapore, Singapore; [7]University of Oxford, Oxford, UK; [8]Research Network on Chronic Diseases (REDISSEC), Madrid, Spain

ABSTRACT

Background: The Spanish five-level EuroQol five-dimensional questionnaire (EQ-5D-5L) valuation study was the first to use the EuroQol Valuation Technology protocol, including composite time trade-off (C-TTO) and discrete choice experiments (DCE). In this study, its investigators noticed that some interviewers did not fully explain the C-TTO task to respondents. Evidence from a follow-up study in 2014 confirmed that when interviewers followed the protocol, the distribution of C-TTO responses widened. **Objectives:** To handle the data quality issues in the C-TTO responses by estimating a hybrid interval regression model to produce a Spanish EQ-5D-5L value set. **Methods:** Four different models were tested. Model 0 integrated C-TTO and DCE responses in a hybrid model and models 1 to 3 altered the interpretation of the C-TTO responses: model 1 allowed for censoring of the C-TTO responses, whereas model 2 incorporated interval responses and model 3 included the interviewer-specific protocol violations. For external validation, the predictions of the four models were compared with those of the follow-up study using the Lin's concordance correlation coefficient. **Results:** This stepwise approach to modeling C-TTO and DCE responses improved the concordance between the valuation and follow-up studies (concordance correlation coefficient: 0.948 [model 0], 0.958 [model 1], 0.952 [model 2], and 0.989 [model 3]). We recommend the estimates from model 3, because its hybrid interval regression model addresses the data quality issues found in the valuation study. **Conclusions:** Protocol violations may occur in any valuation study; handling them in the analysis can improve external validity. The resulting EQ-5D-5L value set (model 3) can be applied to inform Spanish health technology assessments.

Keywords: economic, health status index, life valuation, quality of life.

Introduction

In 2012, the EuroQol Group developed a new standardized protocol (version 1.0) to perform country-specific valuation studies for the five-level EuroQol five-dimensional questionnaire (EQ-5D-5L) using EuroQol Valuation Technology (EQ-VT) [1]. The EQ-VT protocol was developed to elicit health preferences through face-to-face interviews using two valuation techniques, the composite time trade-off (C-TTO) [2,3] and a discrete choice experiment (DCE) [4]. Each respondent completed C-TTO tasks for 10 EQ-5D-5L health states and forced-choice pair comparisons for seven pairs of EQ-5D-5L health states without duration. The C-TTO was a modified version of the traditional TTO technique [5,6], which used the traditional TTO technique for health states considered to be better than immediate death (BTD) and a lead-time TTO technique [7–9] for states considered to be worse than immediate death (WTD).

The C-TTO task entailed a series of consecutive and adapted choices terminating when respondents stated indifference. Because of the complexity of the task, the EQ-VT protocol included an example of this task (being in a wheelchair), which was designed to facilitate and standardize interviewers' explanations. In a previous publication, we described the Spanish EQ-5D-5L valuation study [10]. During this initial analysis, interviewer effects were identified, which were attributed to protocol violations by specific interviewers. Some interviewers did not explain the WTD sections of the C-TTO task and respondents may not have been aware of these sections, leading to fewer WTD

* Address correspondence to: Juan M. Ramos-Goñi, EuroQol Research Foundation, Executive Office, C/Calvario 271-B 1IZQ, Canary Islands 38350, Spain.

E-mail: jramos@euroqol.org

https://doi.org/10.1016/j.jval.2017.10.023

values. In fact, evidence from a follow-up study performed in Spain [11], which used an updated protocol version, showed that when interviewers properly explained the WTD sections of the C-TTO task, a higher proportion of negative numbers were observed [12], altering the distribution of the C-TTO responses. In addition, some interviewers did not properly explain the wheelchair example or showed only a few steps from the iterative procedure to respondents [12]. These participants may have responded imprecisely either because they were not aware of the full iterative procedure or to avoid the time and effort needed to reach their accurate indifference points (i.e., satisficing) [13]. We hypothesized in this study that the C-TTO responses in the Spanish EQ-5D-5L valuation study, although not being as precise as we had expected, still contain valuable information about health preferences from the Spanish population. We demonstrate that such information can be retrieved by assessing individual's paths during the iterative procedure when completing each C-TTO task. At this time, we have no reason to believe that the DCE responses in the valuation study were affected by the protocol violations in the C-TTO tasks.

The primary objectives of this article were to introduce an analytical approach based on hybrid interval regression models (jointly incorporating C-TTO and DCE responses), which updates our previous work [10] to handle the data quality issues commented earlier, and to produce an EQ-5D-5L value set for health technology assessments in Spain. Furthermore, we assessed the external validity of the resulting value set by comparing its estimates with those of a follow-up study.

Methods

Data

The Spanish EQ-5D-5L valuation study has been previously reported in the literature [10,12], and therefore we describe it only briefly here. The valuation study included 1000 face-to-face interviews conducted in 2012 following the EQ-VT protocol version 1 [1]. After applying exclusions, the analytical sample included 9730 C-TTO responses on 86 health states and 7000 DCE responses on 196 pairs of health states. The sample was representative of the Spanish general population with respect to age and sex.

We used C-TTO and DCE responses from a follow-up study conducted also in Spain in 2014 to assess the external validity of the models described later [11]. This follow-up study was performed in only one Spanish region (Canary Islands), and therefore it was not representative of the Spanish population. Nevertheless, it included the quality control process currently recommended by the EuroQol Group to improve data quality. The original aim of the follow-up study was to test the effect of adding a ranking task to the protocol and its results showed that this addition had no significant effect. Therefore, the data from all study arms of the follow-up study were used for external validation.

The C-TTO Iterative Procedure

The C-TTO task used an iterative procedure (Fig. 1) composed of a series of consecutive and adapted choices terminating when respondents stated indifference. Across its four sections, boxes indicate the possible C-TTO responses (i.e., values) and the arrows represent steps from one value to another. Each C-TTO task started (Start box) by asking whether the respondent preferred 10 years in full health or 10 years in the EQ-5D-5L state. If the respondent preferred 10 years in the EQ-5D-5L state (double arrow up from 1 to 1), the same question was asked again to confirm the extreme value. If the respondent preferred 10 years in full health over 10 years in the EQ-5D-5L state (i.e., double arrow down from 1 to 0), the next question was whether the respondent preferred 0 years in full health (i.e., die immediately) or 10 years in the EQ-5D-5L state.

In the iterative procedure (Fig. 1), the "immediate death" question separated the BTD and WTD scenarios (0 at center left). If the respondent preferred 10 years in the EQ-5D-5L state (i.e., BTD state; double arrow up from 0 to 0.5), the next question was whether the respondent preferred 5 years in full health or 10 years in the EQ-5D-5L state. If the respondent preferred to die immediately over 10 years in the EQ-5D-5L state (i.e., WTD state; double-dash arrow from 0 to 0 on the left), the next question was a confirmation of the response but in a lead-time TTO scenario, that is, 10 years in full health versus 10 years in full health followed by 10 years in the EQ-5D-5L state. If the respondent preferred 10 years in full health (double arrow down from 0 to −0.5), the next question was whether the respondent preferred 5 years in full health or 10 years in full health followed by 10 years in the EQ-5D-5L state. If the respondent preferred 10 years in full health followed by 10 years in the EQ-5D-5L state (double arrow horizontal from 0 to 0.05), the iterative procedure changed back to the BTD scenario and the next question asked whether the respondent preferred 0.05 years in full health or 10 years in the EQ-5D-5L state. After these initial steps (double arrows to −0.5, 0.05, 0.5, and 1), the iterative procedure imposed 1-year increments/decrements (i.e., single arrows) followed by half-year corrections (i.e., single-dash arrows) depending on the respondent's preferences. Respondents who visited the BTD scenario after the three initial steps and switched later to the WTD scenario, that is, preferred to die immediately over 10 years in the EQ-5D-5L state (double-dash arrow from 0 to 0 on the right), also had to complete the WTD confirmatory question. This was, however, only once per state.

Although respondents were allowed to go from −0.05 to 0 (immediate death), they were not allowed to go from 0 to −0.05 because of a survey programming error (elbow arrows from 0 to −0.5).

Analysis

Modeling

In a previous publication, we developed and estimated a hybrid model using C-TTO and DCE responses [10]. This initial hybrid model (model 0) assumed normality, homoscedasticity, and that respondents completed the C-TTO tasks accurately. In this study, we followed an analytical approach that relaxed the initial assumption about the accuracy of the C-TTO responses. Specifically, we reconsidered censoring, respondent uncertainty, and protocol violations on the C-TTO tasks [12] as follows.

Censoring of C-TTO responses at −1. The C-TTO task had a minimum TTO value bounded at −1 by design and produced responses in the range [−1, 1]. Nevertheless, feedback from interviewers suggested that some respondents would have responded beyond −1 if allowed, which corroborates the findings of Attema et al. [14]. Because values may be in the range (−∞,1], we relaxed this lower bound assumption and considered responses at the lower bound (−1) to be censored, similar to the open intervals produced by DCE responses ($A > B$) [15].

Inaccuracy of C-TTO responses. The EQ-VT recorded the full path in the C-TTO iterative procedure for each state presented. Using these paths, we built intervals for each state for each respondent. Instead of considering only the final indifference point, this interval assessment used all path information in a conservative manner. Specifically, we observed four response

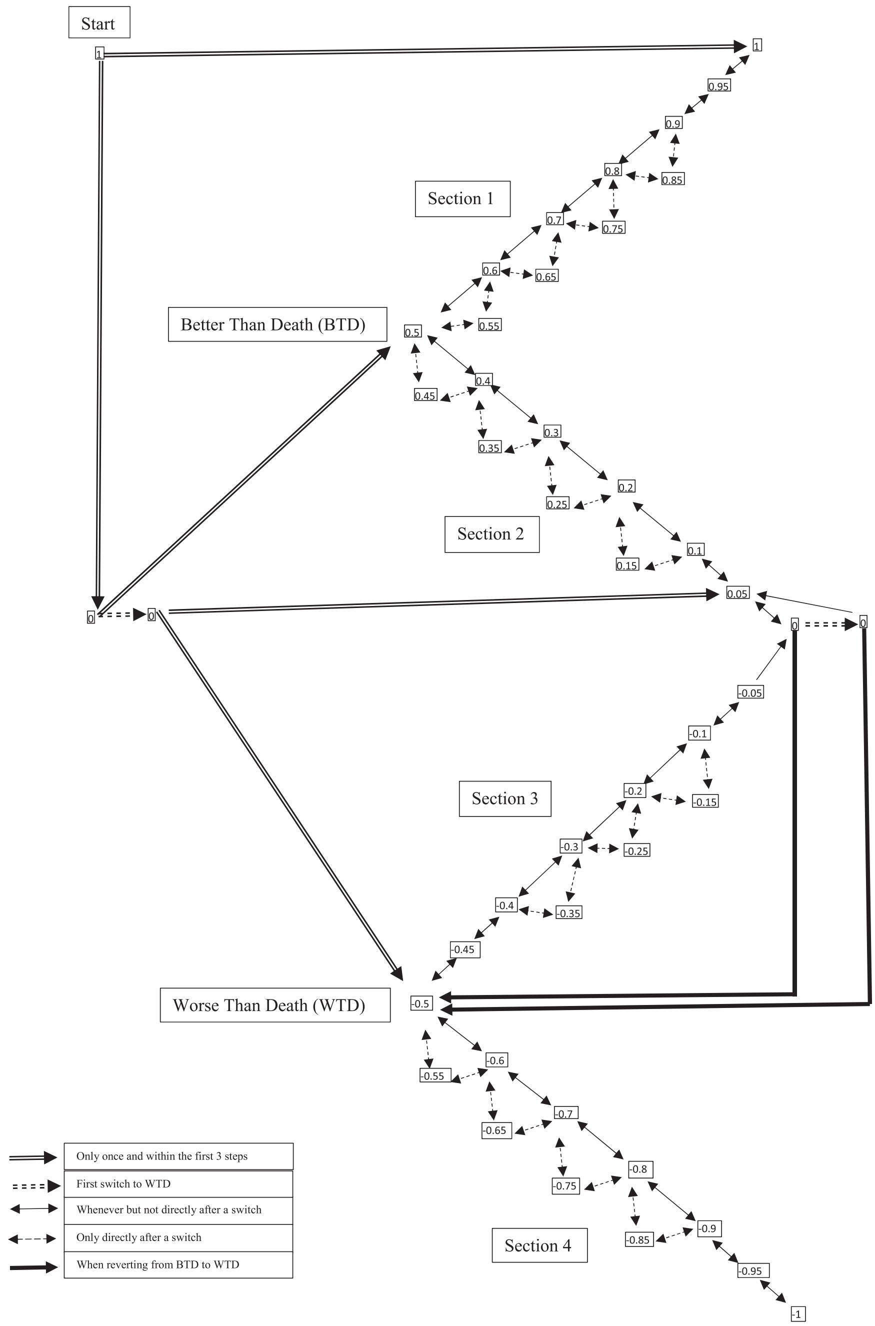

Fig. 1 – C-TTO Iterative procedure. C-TTO, composite time trade-off.

patterns (see examples of each in Supplemental Materials 1 found at http://dx.doi.org/10.1016/j.jval.2017.10.023):

1. *Straight-lining:* This refers to an uninterrupted path, only up or only down, that leads to extreme values of a section, namely, 1, 0.95, 0.05, 0, –0.05, –0.95, and –1, using the minimum number of steps. We refer to this response behavior as straight-lining because it represents repeated choices of the same alternative until the end of a C-TTO section.
2. *Satisficing:* This refers to an uninterrupted path, only up or only down, that leads to nonextreme values of a section using the minimum number of steps. We observed that many interviewers used few number of steps to explain the C-TTO tasks using the wheelchair example, leading us to suspect that some respondents were not trained to perform sufficient steps in the C-TTO iterative procedure to express their values accurately [12]. The literature refers to this lack of engagement as *satisficing* [13] and its prevalence varies by interviewer [12]. For straight-lining and satisficing, we constructed the intervals by section of the iterative procedure (Fig. 1). For example, in section 1, the path 1 → 0 → 0.5 → 0.6 may imply that the respondent had an indifference point of 0.6 or that the respondent was insufficiently engaged to express his or her indifference point accurately within the interval [0.5, 1]. If the path was 1 → 0 → 0.5 → 0.6 → 0.7, then the interval becomes [0.6, 1] and so on. In section 2, the path 1 → 0 → 0.5 → 0.4 → 0.3 may imply an indifference point of 0.3 or an interval of [0, 0.4]. Similar intervals were derived for paths included in sections 3 and 4.

3. *Circling:* This is a path that circles around a value, by going up and down the iterative procedure. For example, the path 1 → 0 → 0.5 → 0.6 → 0.7 → 0.65 → 0.6 → 0.65 → 0.7 may imply a respondent's indifference point of 0.7 or an interval as [0.6, 0.7].
4. *Wandering:* This is a path that wanders up and down in the iterative procedure without any clear pattern, implying that the respondent is having difficulty with the task. For example, the path 1 → 0 → 0.5 → 0.6 → 0.7 → 0.8 → 0.9 → 0.85 → 0.8 → 0.7 → 0.6 → 0.5 → 0.4 → 0.3 → 0.35 → 0.4 → 0.5 may imply an indifference point of 0.5 or an interval using the lowest and highest visited values [0.3, 0.9]. Wandering was like circling in that the respondent takes an inefficient path to the indifference point, but differs from circling because the respondent did not hone in or circle around a value.

In summary, we created an interval for each indifference point. In case of no switches beyond the first three steps (straight-lining and satisficing), the bounds of the interval were defined as the previous visited value and the corner of the iterative procedure section (1 for section 1, 0 for sections 2 and 3, and −1 for section 4). In the case of switches after the first three steps (circling and wandering), the bounds were defined by the minimum and maximum visited values beyond the first three steps. Nevertheless, for section 3, as the iterative procedure forced respondents to go directly from 0 to −0.5 (discussed earlier), the value of −0.5 was not considered the minimum when it was reached from 0.

In addition, we tested several interval definitions for cases with limited information (i.e., fewer than three steps). On the basis of comparing follow-up and other EQ-VT-based valuation studies [12], we decided to define intervals for such paths as follows: 1) path "1 → 0," interval [−0.05, 0.05]; 2) path "1 → 0 → 0.5," interval [0.45, 1]; 3) path "1 → 0 → 0," interval [−0.05, 0.05]; and 4) path "1," interval [−0.995, 1].

To illustrate the intervals, we used a scatterplot showing each health state included in the C-TTO design with their observed mean values together with the mean upper and lower bounds of the intervals. Further details about each path that we found in the data and its corresponding interval can be found in Supplemental Materials 1.

Protocol violations. For most respondents in the valuation study (76.1%), interviewers did not show and explain the iterative procedure allowing for WTD responses, largely shifting the lower bound up from −1 to 0 [12]. To relax the assumption of no protocol violations, all C-TTO responses that were equal to 0 and were collected without WTD explanation were considered to be censored.

We used hybrid models to sequentially apply these three assumptions to our reference case (model 0), which assumed a normal distribution of the errors for the C-TTO responses (as an ordinary least squares model) and a logistic distribution of the error differences for the DCE responses (as a conditional logit model). Coefficients for model 0 were presented previously, but in this study the constant was removed because it was not statistically significant [10]. Model 1 relaxed the lower bound assumption of the value range (i.e., C-TTO values are censored at −1). Model 2 extended model 1 by replacing C-TTO point responses with intervals (explained earlier), relaxing the interpretation of the C-TTO responses and integrating behavioral data on the path to the indifference point. Finally, model 3 extended model 2 by incorporating protocol violations (i.e., C-TTO values were censored at 0 if the respondent did not receive a WTD explanation). All four models were estimated using a cluster estimation of the standard errors on the basis of the respondent to account for multiple C-TTO and DCE responses from each respondent.

Dependent and independent variables
The dependent variable represented the DCE and C-TTO responses. The DCE responses were codified as a binary variable for all models. The C-TTO were codified either as indifference points (model 0) or as intervals (models 1–3) [15]. Each model included an identical set of 20 dummy variables that represented the incremental differences between the five consecutive levels (1–2, 2–3, 3–4, 4–5) within each of the five dimensions of the EQ-5D-5L (i.e., main effects). To facilitate health technology assessments, the Appendix in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2017.10.023 shows the preferred model with 20 cumulative dummy variables (1–2, 1–3, 1–4, 1–5).

Heteroscedasticity
The observed variability in C-TTO responses was not uniform across health states and the variance of C-TTO responses depended on the severity of the health state being valued [10]. We tested for homoscedasticity of the error term using a separate Tobit model for the C-TTO data with the 20 incremental dummy variables (i.e., main effects) [16]. If the homoscedasticity assumption was rejected, the statistical inference may not have been accurate.

External validation
The follow-up study had fewer protocol violations (5.2%) than did the Spanish valuation study (86.5%); therefore, we considered its data more accurate. Using the follow-up study data, we estimated a heteroscedastic hybrid model in which C-TTO responses were censored at −1. Predictions for the 3125 EQ-5D-5L states (5^5) were compared with the predictions of the four aforementioned models using the Lin's concordance correlation coefficient (CCC) as a measure of agreement. We evaluated model performance using 1) logical consistency of parameters and 2) CCC with the external validation model.

We also compared the predictions for the 86 health states included in the C-TTO design using scatterplots. Scatterplots were also used to compare the four models. We plotted the kernel distribution for model 0, the selected value set, and the external validation model. We plotted the kernel distribution of the 3125 values of the final selected value set, the 243 values of the previous three-level EQ-5D (EQ-5D-3L) value set [17], and the crosswalk value set derived from the EQ-5D-3L value set in Spain [18].

We performed all analyses using Stata 14 MP (StataCorp LP, College Station, TX) [19] and estimated hybrid models using the user-written *hyreg* command [15].

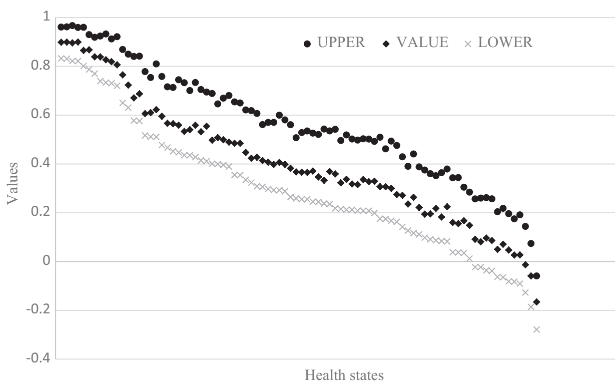

Fig. 2 – Mean of observed values and lower and upper bounds of the intervals based on the iterative procedure.

Table 1 – Hybrid models (C-TTO and DCE) estimations.

Dimension	Independent variables of the model	Model 0		Model 1		Model 2		Model 3* (value set)		External validation (follow-up study)	
		Coefficient (SE)	P	Coefficient (SE)	P	Coefficient (SE)	P	Coefficient (SE)	P	Coefficient (SE)	P
Mobility	No to slight problems	0.086 (0.008)	0.000	0.086 (0.008)	0.000	0.084 (0.009)	0.000	0.084 (0.010)	0.000	0.100 (0.015)	0.000
	Slight to moderate problems	0.014 (0.009)	0.120	0.014 (0.009)	0.109	0.012 (0.010)	0.222	0.015 (0.011)	0.183	0.000 (0.016)	0.993
	Moderate to severe problems	0.131 (0.010)	0.000	0.133 (0.011)	0.000	0.134 (0.010)	0.000	0.15 (0.012)	0.000	0.128 (0.016)	0.000
	Severe problems to unable	0.059 (0.010)	0.000	0.062 (0.010)	0.000	0.066 (0.010)	0.000	0.088 (0.011)	0.000	0.106 (0.016)	0.000
Self-care	No to slight problems	0.058 (0.008)	0.000	0.057 (0.008)	0.000	0.052 (0.009)	0.000	0.050 (0.010)	0.000	0.037 (0.013)	0.003
	Slight to moderate problems	0.000 (0.009)	0.975	0.001 (0.010)	0.920	0.003 (0.010)	0.776	0.003 (0.012)	0.780	0.007 (0.017)	0.687
	Moderate to severe problems	0.097 (0.011)	0.000	0.099 (0.011)	0.000	0.098 (0.011)	0.000	0.111 (0.012)	0.000	0.139 (0.017)	0.000
	Severe problems to unable	0.015 (0.009)	0.107	0.017 (0.010)	0.076	0.019 (0.010)	0.048	0.032 (0.011)	0.004	0.030 (0.016)	0.065
Usual activities	No to slight problems	0.055 (0.008)	0.000	0.055 (0.008)	0.000	0.047 (0.009)	0.000	0.044 (0.010)	0.000	0.062 (0.014)	0.000
	Slight to moderate problems	0.005 (0.010)	0.638	0.004 (0.010)	0.670	0.004 (0.010)	0.691	0.005 (0.011)	0.663	0.001 (0.015)	0.943
	Moderate to severe problems	0.072 (0.010)	0.000	0.074 (0.010)	0.000	0.075 (0.010)	0.000	0.086 (0.012)	0.000	0.109 (0.017)	0.000
	Severe problems to unable	0.004 (0.010)	0.685	0.006 (0.010)	0.554	0.009 (0.010)	0.374	0.018 (0.012)	0.122	0.026 (0.020)	0.191
Pain/discomfort	No to slight problems	0.080 (0.008)	0.000	0.080 (0.008)	0.000	0.076 (0.009)	0.000	0.078 (0.010)	0.000	0.069 (0.013)	0.000
	Slight to moderate problems	0.024 (0.009)	0.008	0.024 (0.009)	0.009	0.024 (0.010)	0.022	0.023 (0.012)	0.045	0.019 (0.015)	0.225
	Moderate to severe problems	0.114 (0.011)	0.000	0.118 (0.011)	0.000	0.121 (0.010)	0.000	0.144 (0.012)	0.000	0.136 (0.018)	0.000
	Severe to extreme problems	0.106 (0.010)	0.000	0.109 (0.011)	0.000	0.114 (0.011)	0.000	0.136 (0.012)	0.000	0.164 (0.019)	0.000
Anxiety/depression	No to slight problems	0.088 (0.008)	0.000	0.087 (0.008)	0.000	0.082 (0.009)	0.000	0.081 (0.010)	0.000	0.035 (0.014)	0.010
	Slight to moderate problems	0.043 (0.010)	0.000	0.044 (0.010)	0.000	0.043 (0.010)	0.000	0.047 (0.012)	0.000	0.058 (0.018)	0.001
	Moderate to severe problems	0.123 (0.010)	0.000	0.126 (0.011)	0.000	0.126 (0.010)	0.000	0.143 (0.012)	0.000	0.144 (0.018)	0.000
	Severe to extreme problems	0.052 (0.010)	0.000	0.055 (0.010)	0.000	0.059 (0.010)	0.000	0.077 (0.011)	0.000	0.132 (0.017)	0.000
	Obs. included in the model										
	Cont. uncensored (C-TTO)	9730		9287		–		–		5192	
	Cont. left-censored (C-TTO)	–		443		443		1467		658	
	Cont. interval (C-TTO)	–		–		9287		8263		–	
	Dich. observations (DCE)	7000		7000		7000		7000		4095	
	Estimated values by health state										
	U(21111)	0.914		0.914		0.916		0.916		0.9	
	U(12111)	0.942		0.943		0.948		0.95		0.963	
	U(11211)	0.945		0.945		0.953		0.956		0.938	
	U(11121)	0.92		0.92		0.924		0.922		0.931	
	U(11112)	0.912		0.913		0.918		0.919		0.965	
	U(55555)	–0.225		–0.25		–0.249		–0.416		–0.501	
	Lin's CCC of model 1 vs. external validation data	0.949		0.958		0.951		0.989		NA	

Note. We dropped all constants because of lack of significance.

CCC, concordance correlation coefficient; C-TTO, composite time trade-off; Cont., Continious; Dich, Dichotomous; DCE, discrete choice experiment; NA, not applicable; SE, standard error; WTD, worse than death.

*We have censored all 0 values for the respondents who did not receive the explanation of the WTD element on the wheelchair example.

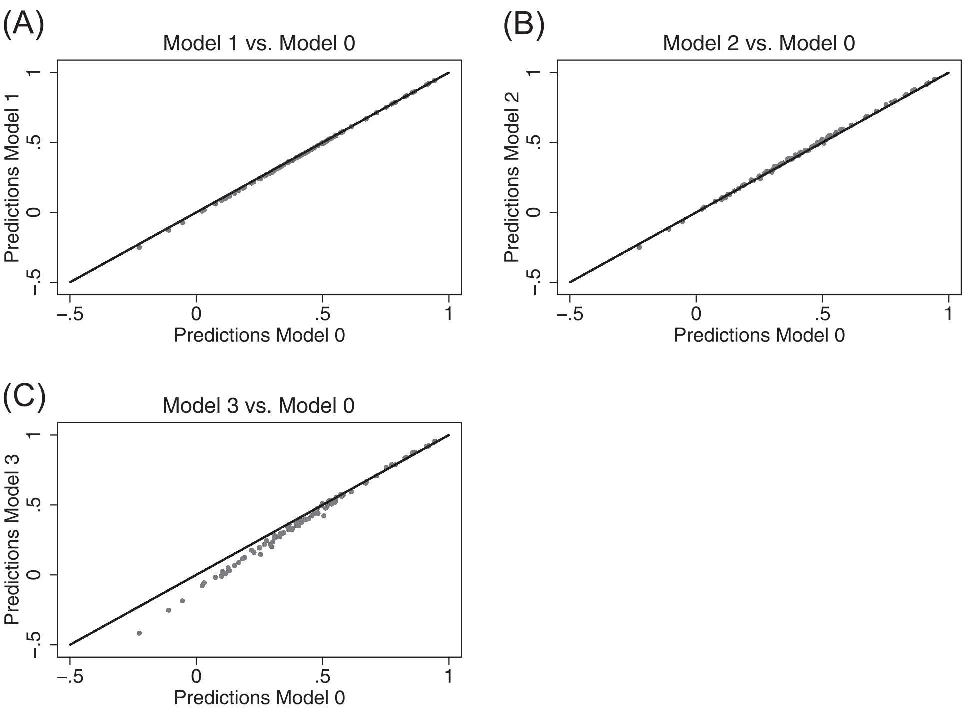

Fig. 3 – Comparison of hybrid model predictions (86 states included in the C-TTO design). C-TTO, composite time trade-off.

Results

Straight-lining and satisficing paths were found in 22.6% and 61.3% of the C-TTO responses, respectively, whereas circling and wandering responses were observed in 10.5% and 5.6% of the responses, respectively. In general, C-TTO intervals were not symmetric around the mean: the distance from the mean C-TTO to the mean of the upper limit was greater than the distance to the mean of the lower limit. Nevertheless, the two distances were more similar for mean C-TTO near 1 than near −1 (Fig. 2).

Estimates of the 20 parameters' coefficients (main effects) were logically consistent for all models (Table 1). As expected, the estimated value of the pits state (55555) in model 0 was higher than in model 1 (in which C-TTO values are censored at −1). Although model 2 (intervals) had no effect on the value of the pits state (55555), its values for the mild health states were slightly higher than in model 0 or model 1. On the contrary, model 3, which addressed censoring due to protocol violations, had a lower value for the pits state (55555) as well as higher values for the mild state, widening the predicted range. In addition, we tested for homoscedasticity and re-estimated the heteroscedastic version of each model accordingly; some estimates were, however, inconsistent (see Supplemental Materials 2 found at http://dx.doi.org/10.1016/j.jval.2017.10.023).

For the 86 states in the C-TTO task, the scatterplots illustrate the relationship between the predictions of model 0 (the reference case) and models 1, 2, and 3. Although censoring increased some coefficients significantly, there appears to be no discernible difference between model 0 and model 1 predictions (Fig. 3A), which may be due to the small proportion of −1 C-TTO responses (4.55%). The differences between model 0 and model 2 predictions (Fig. 3B) appeared across the range of values, and model 2 appeared to have higher values for the mild states. The differences between model 0 and model 3 predictions (Fig. 3C) appeared to increase when less than 0.5, which implied that failure to account for protocol violations increases the values of severe health states and reduces the range of values.

In terms of external validity, model 0 (the reference case) and the follow-up model had a CCC of 0.948 across the 3125 state predictions. Nevertheless, the CCC between model 3 and the follow-up model was 0.989. This result and the fact that 16 of its 20 coefficients are statistically significant led us to recommend model 3 estimates for use in health technology assessments as the Spanish EQ-5D-5L value set (Table 1).

For the 86 states in the C-TTO task, the scatterplots illustrate the relationship between the predictions of external validation model and models 0, 1, 2, and 3 (Figs. 4A,B,C,D, respectively). The comparison with the external validation model showed that model 3 predictions agreed with the predictions of the external validation model (Fig. 4D; CCC 0.989), whereas the other models had worse agreement. Figure 5A further shows that the prediction distribution of model 3 overlapped more closely with the prediction distribution of the external validation model than with the predictions of the reference case (model 0). Figure 5B shows that the prediction distribution of model 3 overlapped with the distribution of crosswalk predictions [18], but not with the distribution of the EQ-5D-3L predictions [17], which is skewed and has more values less than 0 (i.e., die immediately).

Discussion

In this article, we have presented two main findings. The first finding is that our approach to handling data quality issues in the C-TTO responses can be incorporated into the modeling of the EQ-VT data to improve the estimation of EQ-5D-5L value sets. The second finding is the reporting of an EQ-5D-5L value set based on version 1 of the EQ-VT protocol to inform health technology assessment in the Spanish setting.

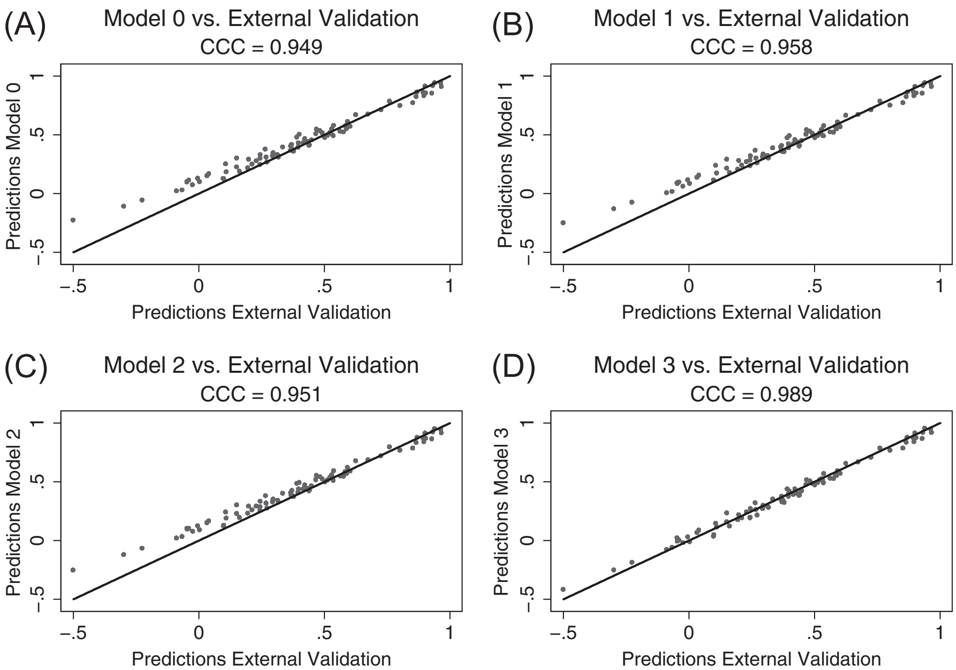

Fig. 4 – Comparison of hybrid model predictions with external validation data from the follow-up study (86 states included in the C-TTO design). CCC, concordance correlation coefficient; C-TTO, composite time trade-off.

The approach introduced here was developed from previous work introducing the hybrid model to estimate a value set using C-TTO and DCE responses [10,15,20,21]. The estimation of a hybrid model, although initially feasible, did not address the data quality issues encountered during the valuation study [12]. New evidence from a follow-up study in Spain suggested that the values for severe health states calculated with the reference case (hybrid model 0) were upward biased because of the data quality issues. At that time, we decided to further develop the hybrid model to allow the use of intervals and censored responses. After these post hoc adjustments, the final model (model 3) produced predictions that were closer to follow-up predictions than the reference case (model 0).

In the process of creating our approach, we developed a path analysis of C-TTO responses to produce intervals that may better represent individual preferences than indifference points (i.e.,

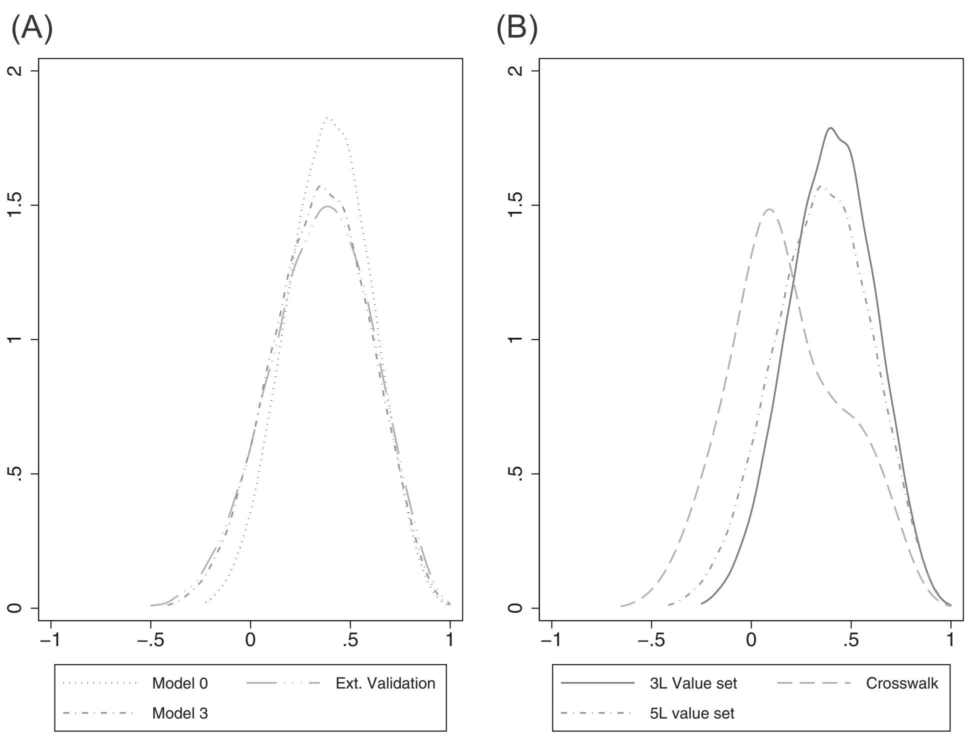

Fig. 5 – Comparison of the distribution of the selected value set (model 3).

indifference point). We recognized that a C-TTO response is not only affected by interviewer behavior but also limited to half-year units by the design. The EQ-VT task generates a total of 41 unique values ranging from −1 to 1 (Fig. 1) and disallows responses less than −1 (censoring), which may not be sufficient to accurately reflect a respondent's value of a health state; more importantly, independently of interviewer's behavior, it is possible that only some respondents are capable of accurately reporting a range of values for a health state. By modeling ranges (i.e., open and closed intervals) as described by C-TTO paths, interval regression analyses benefit from both an improvement in precision (beyond 41 points) and the mitigation of behavioral imprecision (i.e., straight-lining, satisficing, circling, wandering, and censoring) in the iterative procedure. Hence, we encourage further investigation of the interval regression on the basis of pathway analysis for EQ-5D-5L valuation data or similar health preference data (e.g., standard gamble).

The final results (model 3) further suggested some differences between the EQ-5D-3L and the EQ-5D-5L value sets in Spain, which could be due to the instrument or study design. The sample of the valuation study comprised a representative sample of the Spanish population, whereas the EQ-5D-3L value set was estimated using only a representative sample of Catalonians [17]. In the original EQ-5D-3L value set, mobility had the largest value decrement from all EQ-5D dimensions (i.e., confined to bed), but in the EQ-5D-5L, this label was replaced with "unable to walk about" and anxiety and depression had the largest decrement in the EQ-5D-5L value set. In addition to the amendments in labeling, a possible explanation is that preferences of the Spanish population have changed over time because of changes in the socioeconomic environment. For instance, the EQ-5D-3L value set was estimated more than 15 years ago, when the socioeconomic situation in Spain was in a different state than at the time of the EQ-5D-5L valuation study. The current economic situation in Spain has been associated with an increase in the number of people with mental health problems in the country [22–25], and hence the EQ-5D-5L value set reported in this article provides a more realistic representation of the current health preferences in Spain than the original EQ-5D-3L value set.

Study Limitations

This study is subject to some limitations. Definitions of intervals on the basis of limited information such as fewer than three steps could have impacted on modeling results; we have, however, shown that the final model basically replicated results of the follow-up study. The fact that we were not able to estimate a consistent heteroscedastic model made us carefully interpret the resulting *P* values for the model's coefficients. Nevertheless, we tried to limit the impact of this limitation by making a cluster estimation of the standard errors. In addition, the data used for external validation are not representative of the Spanish population, but only from one province. Narrow intervals were more informative than wide ones; therefore, the interval analyses naturally emphasized the C-TTO responses of persons who knew their health preferences and understood the C-TTO task, which may affect the generalizability of the results. Finally, since the completion of this study, the EuroQol international valuation protocol has been updated to version 2, which incorporates new features to improve data quality including a quality control process and a feedback module [26] for C-TTO responses. Evidence from a new wave of studies using this updated protocol suggests that some of the problems encountered in this study during the original data collection are no longer present [27]. Future research should assess the robustness of the value set presented in this article with models using data from the new version of the protocol. The authors encourage such research to understand the implications of using the recommended EQ-5D-5L value set in this article in health technology assessment decisions in Spain.

Conclusions

We explored statistical methods for handling data quality issues in the C-TTO task of the EQ-VT. On the basis of our findings, these analytical adjustments improved external validity and led to the development of a novel interval approach for the analysis of C-TTO responses. Given that the impact of data quality issues is predictable and not unique to the Spanish valuation study, we think that the lessons we learned can be useful to other health preference researchers. Furthermore, we recommend that, in future analysis of EQ-5D-5L valuation data, researchers consider including a similar approach to modeling C-TTO and DCE responses, particularly the examination of intervals on the basis of respondent behaviors. This article also provides a Spanish EQ-5D-5L value set that is recommended for use in health technology assessment in Spain.

Source of financial support: This study was financially supported by the *Instituto de Salud Carlos III*, *Subdirección General de Evaluación y Fomento de la Investigación*, *Plan Estatal de Investigación Científica y Técnica y de Innovación* 2013–2016, and the *Fondo Europeo de Desarrollo Regional* (grant no. PI12/02103). The EuroQol Research Foundation partially funded this work.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2017.10.023 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Oppe M, Devlin NJ, van Hout B, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health 2014;17:445–53.
[2] Janssen BM, Oppe M, Versteegh MM, Stolk EA. Introducing the composite time trade-off: a test of feasibility and face validity. Eur J Health Econ 2013;14(Suppl. 1):S5–13.
[3] Oppe M, Rand-Hendriksen K, Shah K, et al. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics 2016;34:993–1004.
[4] Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics 2008;26:661–77.
[5] Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. Health Serv Res 1972;7:118–33.
[6] Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
[7] Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. Health Econ 2006;15:393–402.
[8] Augustovski F, Rey-Ares L, Irazola V, et al. Lead versus lag-time trade-off variants: Does it make any difference? Eur J Health Econ 2013;14 (Suppl. 1):S25–31.
[9] Luo N, Li M, Stolk EA, Devlin NJ. The effects of lead time and visual aids in TTO valuation: a study of the EQ-VT framework. Eur J Health Econ 2013;14(Suppl. 1):S15–24.
[10] 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: e51–8.
[11] Ramos-Goñi JM, Rand-Hendriksen K, Pinto-Prades JL. Reintroduction of the ranking task in valuation studies: Improved data quality and reduced level of inconsistencies? The case for EQ-5D-5L. Value Health 2016;19:478–86.
[12] Ramos-Goñi JM, Oppe M, Slaap B, et al. Quality control process on EQ-5D-5L valuation studies. Value Health 2017;20:466–73.
[13] Craig BM, Runge SK, Rand-Hendriksen K, et al. Learning and satisficing: an analysis of sequence effects in health valuation. Value Health 2015;18:217–23.

[14] Attema AE, Versteegh MM, Oppe M, et al. Lead time TTO: Leading to better health state valuations? Health Econ 2013;22:376–92.

[15] Ramos-Goñi J, Craig AM, Oppe M, Van Hout B. Combining continuous and dichotomous responses in a hybrid model. Available from: https://euroqol.org/wp-content/uploads/working_paper_series/EuroQol_Working_Paper_Series_Manuscript_16002_-_Juan_Ramos-Goni.pdf. [Accessed September 16, 2017].

[16] Cameron AC, Trivedi PK. Microeconometrics Using Stata. College Station, TX: Stata Press, 2009.

[17] Badia X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making 2001;21:7–16.

[18] van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 2012;15:708–15.

[19] StataCorp. Stata Statistical Software. College Station, TX: StataCorp LP, 2011.

[20] Oppe M, van Hout B. The optimal hybrid: experimental design and modeling of a combination of TTO and DCE. EuroQol Group Proceedings. 2013. Available from: http://www.euroqol.org/uploads/media/EQ2010_-_CH03_-_Oppe_-_The_optimal_hybrid_-_Experimental_design_and_modeling_of_a_combination_of_TTO_and_DCE.pdf. [Accessed October 11, 2014].

[21] Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Med Decis Making 2015;35:328–40.

[22] Córdoba-Doña JA, Escolar-Pujolar A, San Sebastián M, Gustafsson PE. How are the employed and unemployed affected by the economic crisis in Spain? Educational inequalities, life conditions and mental health in a context of high unemployment. BMC Public Health 2016;16:267.

[23] Bartoll X, Palència L, Malmusi D, et al. The evolution of mental health in Spain during the economic crisis. Eur J Public Health 2014;24:415–8.

[24] Roca M, Gili M, Garcia-Campayo J, García-Toro M. Economic crisis and mental health in Spain. Lancet 2013;382:1977–8.

[25] Gili M, Roca M, Basu S, et al. The mental health risks of economic crisis in Spain: evidence from primary care centres, 2006 and 2010. Eur J Public Health 2013;23:103–8.

[26] Wong E, Ramos-Goñi JM, Cheung A, et al. Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong [published online ahead of print October 10, 2017]. Patient. http://dx.doi.org/%2010.1007/s40271-017-0278-0.

[27] Purba FD, Hunfeld JAM, Iskandarsyah A, et al. The Indonesian EQ-5D-5L value set. Pharmacoeconomics 2017;35:1153–65.