GCU
Glasgow Caledonian University

University for the Common Good

Valuation and modeling of EQ-5D-5L health states using a hybrid approach

Ramos-Goni, Juan M; Pinto-Prades, Jose L.; Oppe, Mark; Cabases, Juan M.; Serrano-Aguilar, Pedro; Rivero-Arias, Oliver

OPEN

Valuation and Modeling of EQ-5D-5L Health States Using a Hybrid Approach

Juan M. Ramos-Goñi, MSc,*†‡ Jose L. Pinto-Prades, PhD,§ Mark Oppe, PhD,‡
Juan M. Cabasés, PhD,‖ Pedro Serrano-Aguilar, PhD,*† and Oliver Rivero-Arias, DPhil†¶

Background: The EQ-5D instrument is the most widely used preference-based health-related quality of life questionnaire in cost-effectiveness analysis of health care technologies. Recently, a version called EQ-5D-5L with 5 levels on each dimension was developed. This manuscript explores the performance of a hybrid approach for the modeling of EQ-5D-5L valuation data.

Methods: Two elicitation techniques, the composite time trade-off, and discrete choice experiments, were applied to a sample of the Spanish population (n = 1000) using a computer-based questionnaire. The sampling process consisted of 2 stages: stratified sampling of geographic area, followed by systematic sampling in each area. A hybrid regression model combining composite time trade-off and discrete choice data was used to estimate the potential value sets using main effects as starting point. The comparison between the models was performed using the criteria of logical consistency, goodness of fit, and parsimony.

Results: Twenty-seven participants from the 1000 were removed following the exclusion criteria. The best-fitted model included 2 significant interaction terms but resulted in marginal improvements in model fit compared to the main effects model. We therefore selected the model results with main effects as a potential value set for this methodological study, based on the parsimony criteria. The results showed that the main effects hybrid model was consistent, with a range of utility values between 1 and −0.224.

Conclusion: This paper shows the feasibility of using a hybrid approach to estimate a value set for EQ-5D-5L valuation data.

Key Words: utility theory, quality of life, maximum likelihood estimation, time trade-off, discrete choice experiment

(*Med Care* 2014;00: 000–000)

The EQ-5D instrument is the most widely used preference-based health-related quality of life questionnaire in cost-effectiveness analysis. Reimbursement agencies such as the UK National Institute for Health and Care Excellence (NICE) recommend the use of the EQ-5D in submissions to the institute and this partly explains the spread use of the instrument in applied studies.[1]

The original EQ-5D (EQ-5D-3L) is a questionnaire with 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and 3 levels in each dimension (no problems, some problems, and extreme problems).[2] Extensive research supports the use of the instrument in many disease areas but recent studies have shown ceiling effects issues, particularly in general population samples.[3,4] In response to this, the EuroQol Group proposed a new version of the instrument: the EQ-5D-5L. This new version increased the number of severity levels from 3 to 5 (no problems, slight, moderate, severe, and unable or extreme) describing 3125 (5^5) possible health states.[3] Each health state is usually represented using a 5-digit number (profile) where 11111 indicates perfect health and 55555 the worst health state or pits state.

Available EQ-5D-3L value sets cannot be used directly with 5-level version responses. As a temporary solution, an interim scoring algorithm needs to be used.[5] Therefore, new valuation studies are necessary to obtain preferences from the general public for EQ-5D-5L health states. The EuroQol

Group has developed a valuation protocol to elicit preferences after a series of pilot studies conducted by research teams worldwide.[6] A group of researchers based in Spain, the UK, and the Netherlands, has been one of the first teams in implementing this protocol. This manuscript explores the feasibility of a hybrid method to estimate a potential value set for EQ-5D-5L valuation data.

METHODS

Protocol

The results obtained from the pilot studies[6] informed the standardized protocol for EQ-5D-5L value sets used in this study.[7] The interview process described in the protocol has 5 sections. First, a general welcome and an introduction to the research were given. Next, respondents were asked to provide background information, including their own health using the EQ-5D-5L, age, sex, and experience with illness. This was followed by the composite time trade-off (C-TTO) task, which was administered after giving an explanation of the task, and included 10 EQ-5D-5L C-TTO valuations. The next part was a discrete choice (DC) experiment, which consisted of 7 paired comparisons. Finally, there was a general thank you and goodbye. After each block of tasks (C-TTO and DC experiments) and at the end of the interview, participants were given the opportunity to clarify whether they found difficulties completing the tasks and the overall survey. The EuroQol Group developed the online system to carry out the survey called EuroQol Valuation Technology (EQ-VT).

Eliciting Preferences Methods

C-TTO

The traditional time trade-off (TTO) has been widely used in the EQ-5D-3L valuation studies conducted so far and it is appropriate to value health states considered better than dead.[8,9] However, using the traditional TTO method for states worse than dead gives negative values that are normally transformed to be bounded to -1, which has been criticized in the literature.[10] Other TTO alternatives to evaluate health states were therefore assessed during the EuroQol pilot studies including lead and lag time.[11,12] In the former, additional trading time is included before the health state, whereas in the latter, trading time is included after the health state to be valued. The pilot studies looked at the potential of using these methods in practice and concluded that the protocol should include a composite TTO method.

This composite approach involved the use of the traditional TTO approach for states better than dead and lead time TTO for states worse than dead in a single task.[13] For the lead time TTO, 10 years lead time and 10 years in the state were used. This lead time method produces a minimum value of -1 and no transformation of negative values is needed. The iterative process used in the original UK valuation exercise[8] was adapted to be used in the C-TTO task. The C-TTO design included 86 health states selected using Monte Carlo simulation. The health states were distributed over 10 blocks and each block contained 1 very mild state (1 dimension at level 2, the remaining dimensions at level 1),

the pits state 55555, and a balanced set of intermediate states. The EQ-VT randomly assigned respondents to one of the blocks and presented the states in random order.

DC Experiment

The use of DC experiments for health state valuation has received recent attention in the literature.[14,15] Modeling ordinal data follows the theoretical foundations of random utility theory.[16] Values obtained with DC models have been shown to have patterns similar to those obtained with TTO models.[17] The values obtained from DC models are expressed on an arbitrary scale and need to be rescaled on the dead (0) full health (1) scale.[17,18] Using DC experiments was also piloted and the results suggested that collecting such information could provide additional useful information to the C-TTO data. Hence, a DC experiment was included as part of the protocol. The DC experiment design included 196 pairs divided in 28 blocks with similar severity representation identified using Bayesian design.[19] The EQ-VT randomly assigned respondents to one of the blocks, presented the pairs in random order, and randomized the location of the states within the pair (ie, left and right).

Sampling and Data Collection

Our power calculations estimated that to obtain a 0.01 SE of the observed mean C-TTO, we needed 9735 C-TTO responses. We therefore recruited 1000 participants that after completing the valuations tasks provided 10,000 C-TTO and 7000 DC responses to estimate the models.

A 2-stage sampling strategy was designed to obtain a representative sample of the Spanish population. In a first stage, we stratified geographically by Spanish provinces, whereas in a second stage we systematically sample individuals from a panel until an accurate age and sex distribution for that province was achieved. We contracted an independent market research company, which identified respondents and arranged interviews at convenient places. Interviews were conducted face-to-face during June and July 2012 by 33 trained interviewers. Respondents did not receive payment for participating in the survey. A different market research company was contracted to call a random subsample of 15% of respondents as quality control of the process.

Statistical Analyses

Descriptive statistics were used to summarize respondent's characteristics and responses to the C-TTO and DC experiments.

Two sources of data were available to estimate the EQ-5D-5L value set: C-TTO and DC data. To maximize the use of the available data, we implemented a hybrid modeling approach that made use of both C-TTO and DC data to estimate the potential value sets. This hybrid method estimated a unique set of coefficients from a likelihood function obtained multiplying the likelihood functions of a normal distribution for the C-TTO data by the likelihood function of a conditional logit distribution for DC data.[20] As the coefficients estimated from a conditional logit are expressed on a latent arbitrary utility scale, we used a rescaled parameter

θ, which assumes that the C-TTO model coefficients are proportional to DC model coefficients. See the Appendix for a full description and analytical derivation of the hybrid method. This method combines the utility values elicited in the C-TTO for the 86 health states with utility values elicited in the DC experiment for 196 pairs of states. The dependent variable in the C-TTO part of the model was defined as 1 minus the C-TTO observed values for a given health state to indicate disutility and therefore coefficients expressed utility decrements. In the DC part of the model, the dependent variable was a binary outcome 0/1 indicating the respondent's choice for each pair of EQ-5D-5L states. We used cluster estimation to acknowledge that for each participant included in the models, 10 C-TTO and 7 DC responses were available.

We also present models to estimate C-TTO and DC data separately, to illustrate how the hybrid model combined both types of data. We analyzed C-TTO data using a linear regression model assuming normal distribution in its errors, as it is the C-TTO part of hybrid model. We analyzed DC data using the standard econometric method for ordinal data conditional logit regression.[16] To make model coefficients comparable, we rescaled the DC model coefficients using the same rescaling parameter θ that was estimated in the hybrid model.

We started exploring the hybrid main effects with a 20-parameter model consisting of 4 dummies for each EQ-5D-5L dimensions using level 1 as the reference. We constructed dummies to represent the additional utility decrement of moving from one level to another. For instance for the mobility dimension we created 4 dummies MO1 to MO4 and the coefficient associated to MO1 indicated the utility decrement of moving from no problems (level 1) to slight problems (level 2), MO2 the additional utility decrement of moving from slight (level 2) to moderate (level 3) problems, and so on. Therefore, the overall decrement of moving from no to moderate problems could be calculated as the sum of the coefficients of MO1 plus MO2. The same set of dummy variables was defined for each of the remaining dimensions: self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD). We also estimated the model using the definition of dummies implemented in most previous EQ-5D-3L valuation exercises[21] and such analyses are available from the authors upon request.

Our starting point for the selection of additional covariates for the models was the US valuation study.[9] Several variables were defined. For example, $D1$ as the number of dimensions at levels 2, 3, 4, or 5 beyond the first; IJ as the number of dimensions at level J beyond the first; $K45$ as the number of dimensions at level 4 or 5, and others. Squared of all terms were also introduced to assess nonlinear effects on the dependent variable. We included all terms first, and use a stepwise approach removing nonsignificant terms and ensuring model consistency.

Exclusion Criteria and Interviewer Assessment

We excluded observations using the following 2 criteria: (1) respondents with a positive slope on a regression between his/her values and the severity of the health states indicating that the participant provided higher utility values

for poorer health states on average; and (2) respondents who valued all states equal to death.

We used the Kruskal-Wallis test to assess the differences among mean values by interviewer in the C-TTO responses. We further assess this including dummies that identified interviewers in the main effects model and using an F test among the dummy coefficients.

Evaluation of Model Performance

We evaluated model performance using (1) logical consistency of parameters; (2) goodness of fit; and (3) parsimony. Estimated coefficients are said to be logically consistent if magnitude values from logically worse health states are lower than those from logically better health states. In our estimated results this is translated to all main effects coefficients being positive. Goodness of fit was assessed using the Akaike (AIC) and the Bayesian information criteria (BIC). Finally, the principle of parsimony stated that if competing models were similar in logical consistency and goodness of fit, the model with fewer parameters was preferred. These 3 criteria were used to compare different hybrid model specifications using different interaction terms. However, prediction accuracy evaluated using mean square error or mean absolute error are not appropriate measures in this case, given the lack of an appropriate counterfactual for hybrid model predictions.

We present the results of the regression with the main effects and the best-fitted model with significant terms. Statistical analysis and regression modeling were conducted in Stata MP 11.[22] The hybrid model was not available in any standard package and was programmed in Stata specifically for this study.

Comparison With EQ-5D-3L Value Set

We calculated and compared predictions for the 3125 health states using the final selected EQ-5D-5L value set and the interim solution to calculate EQ-5D-3L values[5] presented for a selected set of health states covering mild, moderate, and severe states. In addition, we compared the kernel density functions for the index values of the 243 states of the Spanish EQ-5D-3L value set[23] and for the 3125 states of the final selected EQ-5D-5L value set.

RESULTS

Descriptive Statistics

Twenty-seven participants from the 1000 were removed following the exclusion criteria: 18 respondents with a positive slope on a regression between his/her values and the severity of the health states and 9 respondents who valued all states equal to death. Overall the excluded observations were older with no studies or primary school studies than the estimation sample (Table 1). The estimation sample was similar in the distribution of employment status; mean age and sex distribution than Spanish population, but the estimation sample had a larger number of respondents in age group 25–34 and fewer participants over 75 (Table 1). The self-reported health using the EQ-5D-5L of respondents showed that 18.90% reported problems in usual activities and 30.8% reported problems in anxiety or depression

TABLE 1. Background Characteristics of Excluded Sample, Estimation Sample, and Comparison Against Spanish General Population

	n (%)		
Variables	**Excluded Sample (n = 27)**	**Estimation Sample (n = 973)**	**Spanish General Population (%)***
Age [mean (SD)]	49.26 (18.2)	43.62 (17.2)	40.2 (NA)
Age groups			
18–24	3 (11.2)	114 (11.7)	9.0
25–34	4 (14.8)	270 (27.8)	18.3
35–44	5 (18.5)	170 (17.5)	19.6
45–54	5 (18.5)	148 (15.2)	17.9
55–64	4 (14.8)	111 (11.4)	13.5
65–74	2 (7.4)	108 (11.1)	10.2
75+	4 (14.8)	52 (5.3)	11.0
Sex			
Male	12 (44.4)	463 (47.6)	49.3
Female	15 (55.6)	510 (52.4)	50.7
Employment status			
Housewife/house husband	1 (3.7)	70 (7.2)	10.51
Employed or freelance	11 (40.8)	529 (54.4)	44.98
Student	2 (7.4)	89 (9.1)	6.33
Retired	8 (29.6)	132 (13.6)	20.12
Unemployed	5 (18.5)	139 (14.3)	15.01
Disabled	0 (0)	8 (0.8)	3.03
Missing	—	6 (0.6)	—
Education			
Higher education	10 (37.0)	314 (32.47)	17.70
High school	2 (7.4)	374 (38.68)	53.90
Primary school	10 (37.0)	234 (24.20)	26.30
No studies	5 (18.5)	45 (4.65)	2.10
Missing	—	6 (0.6)	
Experience with illness			
Personal (%yes)	4 (14.8)	140 (14.4)	NA
Relatives (%yes)	17 (62.96)	616 (63.3)	NA
Other (%yes)	9 (33.3)	338 (34.7)	NA
Self-reported EQ-5D-5L			
Mobility			
No problems	22 (81.48)	864 (88.80)	86.1
Slight problems	4 (14.81)	69 (7.09)	6.1
Moderate problems	1 (3.70)	32 (3.29)	4.7
Severe problems	0 (0)	7 (0.72)	2.4
Unable/extreme problems	0 (0)	1 (0.10)	0.8
Self-care			
No problems	24 (88.89)	933 (95.89)	93.9
Slight problems	2 (7.41)	30 (3.08)	2.5
Moderate problems	1 (3.70)	9 (0.92)	1.7
Severe problems	0 (0)	1 (0.10)	0.9
Unable/extreme problems	0 (0)	0 (0)	1.0
Usual activities			
No problems	22 (81.48)	891 (91.57)	89.2
Slight problems	3 (11.11)	57 (5.86)	4.7
Moderate problems	2 (7.41)	20 (2.06)	3.2
Severe problems	0 (0)	4 (0.41)	1.5
Unable/extreme problems	0 (0)	1 (0.10)	1.4
Pain			
No problems	20 (74.07)	772 (79.34)	75.2
Slight problems	5 (18.52)	149 (15.31)	12.3
Moderate problems	1 (3.70)	37 (3.80)	8.7
Severe problems	0 (0)	10 (1.03)	3.5
Unable/extreme problems	1 (3.70)	5 (0.51)	0.4
Anxiety/depression			
No problems	15 (55.56)	673 (69.17)	85.4
Slight problems	8 (29.63)	214 (21.99)	8.4
Moderate problems	2 (7.41)	71 (7.30)	4.2
Severe problems	1 (3.70)	15 (1.54)	1.6
Unable/extreme problems	1 (3.70)	0 (0)	0.4

*Data extracted from the 2012–2013 National Spanish Health Survey.
NA indicates not available.

dimension (Table 1). For the remaining dimensions, proportions of respondents with problems were <10% (Table 1).

The outcome of the quality control reported no incidences, but we observed significant differences between interviewers in the valuations obtained with Kruskal-Wallis ($P < 0.0001$) and F tests ($P < 0.0001$).

We report further descriptive information about the C-TTO and the DC data in the online supplemental digital content (Tables 1 and 2 and SDC Figures 1 and 2, Supplemental Digital Content, http://links.lww.com/MLR/A839).

Modeling Results

The hybrid model with main effects was a consistent model predicting utilities with a range between 1 and -0.224 (Table 2). Both, the C-TTO and the DC models derived logical inconsistencies. It is shown how the hybrid model corrects the inconsistencies in the C-TTO model by using DC information and the DC model inconsistencies with C-TTO information. As described in the Appendix, the log likelihood in the hybrid model was approximately the sum of the log likelihood of both C-TTO and DC models separately.

After exploring many interactions terms, the best-fitted estimation model we found was using the interaction terms $D1^2$ and $K45^2$ (Table 3). The constant term of this model was suppressed as the $D1^2$ term captures the effect of the constant. The reduction of the hybrid log likelihood estimation for those terms inclusion only reduces the AIC and BIC by 0.4%. About 3/4 of this reduction was produced by a reduction in the C-TTO part of the model.

The main effects hybrid model produced a wider range of utility values at the upper and lower end of the scale compared to the hybrid model including the terms $D1^2$ and $K45^2$ (Table 4). Given that the improvement in goodness of fit between the main effects and the best-fitted model was marginal (0.4%), we have selected the estimation results from the hybrid model with main effects as the value set for this methodological study based on the parsimony criteria.

The probability density functions of the Spanish EQ-5D-3L value set and the EQ-5D-5L value set presented here (Fig. 1) show a symmetric distribution for EQ-5D-5L, whereas the EQ-5D-3L has a bimodal distribution. The proportion of states considered worse than death is lower in the EQ-5D-5L value set.

DISCUSSION

In this manuscript we have reported the performance of a hybrid approach to estimate a value set for the EQ-5D-5L questionnaire. The choice of the hybrid approach is based on the assumption that subjects have a unique utility function that generates both the sets of responses. If utilities were the same in the C-TTO and DC methods there would be no need of combining them except for having more precise estimates. Our hypothesis is that this disparity is related to the choice versus matching discrepancy as it is one of the most replicated effects in preference elicitation literature.[24–27] Some researchers have tried to find arguments in favor of one method or the other.[28] We believe that neither matching-based (like C-TTO) nor

TABLE 2. Estimation Results for Hybrid Model Using Main Effects Only

	Hybrid (C-TTO + DC) Model			C-TTO Model			Rescaled DC Model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	P
MO1	0.084	0.008	0.000	0.015	0.015	0.293	0.088	0.010	0.000
MO2	0.014	0.009	0.101	0.053	0.015	0.000	0.012	0.011	0.272
MO3	0.130	0.010	0.000	0.152	0.018	0.000	0.115	0.012	0.000
MO4	0.060	0.010	0.000	0.023	0.019	0.230	0.081	0.013	0.000
SC1	0.056	0.008	0.000	0.038	0.015	0.009	0.030	0.011	0.008
SC2	0.000	0.009	0.989	**−0.001**	0.017	0.964	0.017	0.011	0.126
SC3	0.097	0.011	0.000	0.131	0.020	0.000	0.079	0.013	0.000
SC4	0.016	0.009	0.090	0.012	0.018	0.506	0.022	0.011	0.047
UA1	0.053	0.008	0.000	0.040	0.014	0.006	0.037	0.011	0.000
UA2	0.005	0.010	0.572	0.035	0.019	0.069	**−0.008**	0.011	0.485
UA3	0.072	0.010	0.000	0.085	0.021	0.000	0.069	0.011	0.000
UA4	0.004	0.010	0.705	**−0.030**	0.017	0.082	0.024	0.013	0.056
PD1	0.078	0.008	0.000	0.049	0.014	0.000	0.066	0.011	0.000
PD2	0.024	0.009	0.007	0.044	0.019	0.024	0.019	0.012	0.093
PD3	0.115	0.011	0.000	0.100	0.019	0.000	0.130	0.013	0.000
PD4	0.105	0.010	0.000	0.090	0.023	0.000	0.118	0.014	0.000
AD1	0.085	0.008	0.000	0.057	0.018	0.002	0.065	0.011	0.000
AD2	0.044	0.010	0.000	0.038	0.018	0.040	0.049	0.012	0.000
AD3	0.121	0.010	0.000	0.116	0.019	0.000	0.119	0.013	0.000
AD4	0.053	0.010	0.000	0.049	0.017	0.005	0.063	0.013	0.000
Constant	0.007	0.004	0.087	0.098	0.018	0.000			
LogL	−10292.97			−6565.7			−3675.81		
AIC	20631.95			13173.41			7391.62		
BIC	20809.62			13324.25			7528.69		
U(55555)	−0.224			−0.194			−0.196		
Lowest predicted value (state)	−0.224 (55555)			−0.224 (55455)			−0.196 (55555)		

Bold figures indicate logical inconsistencies.

MO, mobility; SC, self-care; UA, usual activities; PD, pain/discomfort; AD, anxiety/depression.

AIC indicates Akaike information criteria; BIC, Bayesian information criteria; C-TTO, composite time trade-off; DC, discrete choice.

TABLE 3. Estimation Results Using Best-fitted Model

	Hybrid (C-TTO + DC) Model			C-TTO Model			Rescaled DC Model		
	Coefficient	SE	P	Coefficient	SE	P	Coefficient	SE	P
MO1	0.112	0.009	0.000	0.065	0.015	0.000	0.119	0.017	0.000
MO2	0.020	0.008	0.018	0.036	0.015	0.014	0.016	0.011	0.138
MO3	0.143	0.014	0.000	0.199	0.022	0.000	0.134	0.018	0.000
MO4	0.070	0.009	0.000	0.035	0.020	0.079	0.076	0.012	0.000
SC1	0.079	0.009	0.000	0.090	0.013	0.000	0.064	0.019	0.001
SC2	0.006	0.009	0.518	**−0.007**	0.017	0.680	0.018	0.011	0.093
SC3	0.115	0.013	0.000	0.180	0.024	0.000	0.101	0.019	0.000
SC4	0.024	0.009	0.008	0.032	0.020	0.097	0.021	0.010	0.038
UA1	0.081	0.009	0.000	0.102	0.015	0.000	0.073	0.021	0.000
UA2	0.005	0.009	0.607	0.021	0.020	0.297	**−0.005**	0.011	0.622
UA3	0.095	0.013	0.000	0.125	0.025	0.000	0.090	0.017	0.000
UA4	0.012	0.009	0.198	**−0.008**	0.017	0.620	0.024	0.012	0.037
PD1	0.104	0.009	0.000	0.095	0.011	0.000	0.100	0.019	0.000
PD2	0.022	0.008	0.008	0.040	0.020	0.045	0.016	0.011	0.130
PD3	0.146	0.014	0.000	0.159	0.022	0.000	0.153	0.020	0.000
PD4	0.106	0.010	0.000	0.097	0.025	0.000	0.109	0.012	0.000
AD1	0.105	0.008	0.000	0.110	0.012	0.000	0.100	0.020	0.000
AD2	0.043	0.009	0.000	0.016	0.018	0.376	0.044	0.011	0.000
AD3	0.133	0.013	0.000	0.153	0.020	0.000	0.139	0.019	0.000
AD4	0.061	0.010	0.000	0.070	0.019	0.000	0.056	0.012	0.000
$D1^2$	−0.009	0.001	0.000	−0.013	0.003	0.000	−0.007	0.003	0.020
$K45^2$	−0.006	0.002	0.001	−0.009	0.003	0.001	−0.007	0.003	0.031
LogL	−10247.2			−6550.7			−3670.1		
AIC	20542.4			13145.3			7384.3		
BIC	20727.8			13303.4			7535.1		
Lowest predicted value (state)	−0.175 (55555)			−0.181 (55455)			−0.160 (55555)		

Bold figures indicate logical inconsistencies.
MO, mobility; SC, self-care; UA, usual activities; PD, pain/discomfort; AD, anxiety/depression.
AIC indicates Akaike information criteria; BIC, Bayesian information criteria; C-TTO, composite time trade-off; DC, discrete choice.

TABLE 4. Predicted Utility Values for Selected Health States for Estimated Models and From the Interim EQ-5D-3L Solution

| State	Hybrid ME	Hybrid ME + $D1^2$ + $K45^2$	Interim Solution From EQ-5D-3L
11112	0.9072	0.8945	0.9320
21111	0.9081	0.8877	0.8930
11121	0.9146	0.8958	0.9100
12111	0.9363	0.9207	0.8680
11211	0.9392	0.9190	0.9240
42114	0.4574	0.4259	0.4790
33511	0.7030	0.6342	0.2500
25331	0.5788	0.5418	0.1290
35411	0.5942	0.5265	0.1580
34511	0.6064	0.5383	0.2080
35412	0.5088	0.4676	0.1090
33531	0.6012	0.5541	0.1610
55512	0.3160	0.2747	−0.2980
52533	0.2829	0.2822	−0.2030
34544	0.1389	0.1716	−0.0530
34553	0.1551	0.1541	−0.1160
55433	0.1740	0.1817	−0.3170
35552	0.1831	0.1738	−0.2350
54454	−0.1517	−0.0777	−0.3970
55444	−0.0625	0.0045	−0.4260
55552	−0.0060	0.0057	−0.5590
54455	−0.2046	−0.1391	−0.4380
55554	−0.1712	−0.1134	−0.6120
55545	−0.1192	−0.0689	−0.5510
55555	−0.2242	−0.1748	−0.654

ME indicates main effects.
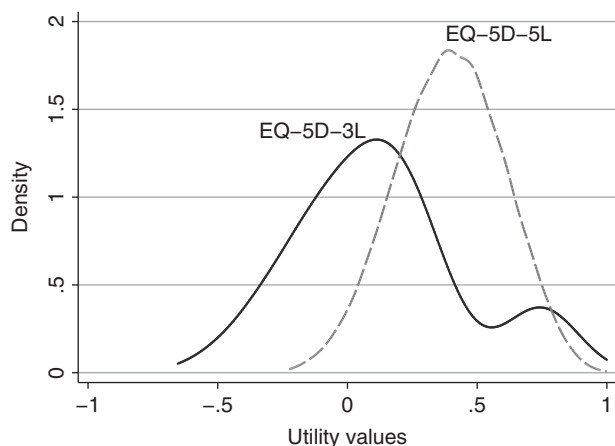
FIGURE 1. Probability density function of EQ-5D-3L and EQ-5D-5L value sets.

choices (DC) methods are unbiased.[29] Matching methods are influenced by scale compatibility and, in the case of C-TTO, by loss aversion.[30] Choices are also subject to problems as it has been shown that responses are more lexicographic in choice than in matching. Evidence on the prominence effect suggests that in choices, subjects tend to choose the alternative that is better with respect to the more important attribute without paying enough attention to how much better the option is.[31] Finally, it has also been observed that subjects perceive the distances between outcomes differently when comparisons are conducted in a separate or in a joint model, again without clear evidence that one method is better than another.[32] We then do not think that the "true" values can be inferred from 1 single method and for this reason we suggest that it can make sense to use a hybrid approach. We are not claiming that the biases present in 1 method compensate the biases present in the other so that adding up the 2 methods we get unbiased results. There is no evidence to suggest this is the case. Even in the absence of such empirical evidence, we think that there are reasons to suspect that, at least, the potential biases present in the C-TTO are not enhanced by choices of the DC experiment, rather the opposite.

In our results, introducing the $D1^2$ and $K45^2$ terms provided a better fit to the data suggesting that the selected value set should have included such effect. However, the improvement in fit is mostly captured by the C-TTO part of the model. Given that the improvement in the goodness of fit of using $D1^2$ and $K45^2$ variables was marginal as suggested by the AIC and the BIC, we selected the main effects model using the parsimony criterion.

As far as we are aware there is no EQ-5D-5L value set available in the literature for direct comparison. Given the lack of such information we compared our model with the Spanish value set for the 3L version of EQ-5D.[23] Our model has higher values in the upper scale compared to the EQ-5D-3L valuation study conducted in Spain. This was expected as the label for level 2 in the 3L version is "some problems" and the label for level 2 in the 5L version is "slight problems." However, the utility decrement of level 2 for AD dimension is higher in our study than in the 3L study. A possible ex-

planation for this is the fact that the self-reported health results in our sample showed a high rate of people reporting problems in the AD dimension, causing them to put more weight on this dimension in the valuation tasks. On the other side of the scale the pits state prediction was higher in our study. Something expected as well, as the change in the wording of the mobility level "confined to bed" in EQ-5D-3L to "unable to walk about" in EQ-5D-5L has changed the definition of the worst possible health state. Given that this new level is not as severe as "confined to bed" (which had the largest decrement of all dimensions in the Spanish 3L study) it is expected to obtain higher valuations for 55555 than for 33333. We observed a lower proportion of negative values in our study in comparison with the Spanish EQ-5D-3L value set. The number of nonextreme health states has increased >10-fold in the EQ-5D-5L compared with the 3L version reducing the proportion of the extreme health states, and partly explaining why the kernel density distribution of the 5L value set shows a smaller area below 0 than the 3L value set.

The hybrid model is not exempt of limitations. The assumption of normal distribution for the errors in the C-TTO part suffers from problems related to the robustness of the estimation of SE and related to the violation of the homoscedasticity condition. In addition, the use of conditional logit model for DC data does not explicitly consider within respondents correlations. We try to limit the impact of these limitations by using cluster estimations of the SEs of the estimated coefficients. However, further exploration of more sophisticated hybrid models for both types of data is needed. For example, the use of random coefficient models for the C-TTO part and mixed (conditional) logit models for the DC part of the model.

We have observed significant differences in the valuations observed by interviewers that lead us to be cautious about suggesting a final value set to use in practice in Spain. We are now trying to understand the nature of these differences, which could be attributable to several factors including issues with the EQ-VT software, the use of C-TTO, or noncompliance of the protocol by the interviewer.

We present here a novel methodological approach to obtain an EQ-5D-5L value set. Our results show the feasibility of using a hybrid model to estimate a value set for EQ-5D-5L valuation data.

APPENDIX: THE HYBRID MODEL

There are several methods that enable the combination of both sets of data in a single model. The hybrid model we present here uses a maximum likelihood approach. It builds on the notion that both linear regression (as applied to the C-TTO data) and logistic regression (as applied to the DC data) can be obtained by maximum likelihood estimation and that

both models contain a similar linear component βx underlying the values and choices. If one assumes that this component, which reflects the weight given to the dimensions and labels, is identical between both approaches, one can find the optimal parameters for the combination of the data. This is done by creating a single likelihood function for the joined data by multiplying the likelihoods of the C-TTO data and the DC data (or—equivalently—by adding the log likelihoods).

However, we know that the C-TTO model and the DC model are anchored on a different scale. We can take this into account in the combined likelihood by including an additional parameter relating both linear functions with each other. In the model presented here, we assume that the weights (ie, the β's) in both models differ up to a monotonic transformation θ.

The likelihood of the C-TTO data is expressed as follows:

$$\text{like}_{\text{C-TTO}} = \text{Normal\_pdf}\left(\sum_{j=1}^{J} \beta_j d_{ij}\right),$$

where β_j is the vector of C-TTO regression coefficients, d_{ij} the vector of dummy variables for state i, J the number of dummies, and pdf the gaussian probability density function.

The likelihood of the DC data is defined as:

$$\text{like}_{\text{DC}} = \text{Logit\_cdf}\left(\sum_{j=1}^{J} \beta_j' \left(d_{ij}^A - d_{ij}^B\right)\right),$$

where β_j' is the vector of DC regression coefficients, d_{ij}^A the vector of dummy variables for state A of pair i, d_{ij}^B the vector of dummy variables for state B of pair i, and cdf the Logit cumulative density function.

Finally: $\text{Loglike}_{\text{hybrid}} = \text{Log}\left(\text{like}_{\text{C-TTO}} \times \text{like}_{\text{DC}}\right)$

$$= \text{Loglike}_{\text{C-TTO}} + \text{Loglike}_{\text{DC}},$$

and the relation between β and β′ is assumed in the estimation to be: $\theta\beta' = \beta$.

REFERENCES

1. National Institute for Health and Care Excellence. *Guide to the Methods of Technology Appraisal*. London: National Institute for Health and Care Excellence; 2013.
2. EuroQol G. EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16:199–208.
3. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20:1727–1736.
4. Bharmal M, Thomas J. Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value Health*. 2006;9:262–271.
5. 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;15:708–715.
6. Devlin N, Krabbe P. The development of new research methods for the valuation of EQ-5D-5L. *Eur J Health Econ*. 2013;14(suppl):1–3.
7. 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;17:445–453.
8. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35:1095–1108.
9. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43:203–220.
10. Craig BM, Oppe M. From a different angle: a novel approach to health valuation. *Soc Sci Med*. 2010;70:169–174.
11. 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–S31.
12. Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. *Health Econ*. 2006;15:393–402.
13. Janssen BM, Oppe M, Versteegh MM, et al. Introducing the composite time trade-off: a test of feasibility and face validity. *Eur J Health Econ*. 2013;14(suppl 1):S5–S13.
14. Salomon J. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. *Popul Health Metr*. 2003;1:12.
15. McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. *J Health Econ*. 2006;25:418–431.
16. Hensher DA, Rose JM, Greene WH. *Applied Choice Analysis: A Primer*. Cambridge: Cambridge University Press; 2005.
17. Stolk EA, Oppe M, Scalone L, et al. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value Health*. 2010;13:1005–1013.
18. Ramos-Goni JM, Rivero-Arias O, Errea M, et al. Dealing with the health state 'dead' when using discrete choice experiments to obtain values for EQ-5D-5L heath states. *Eur J Health Econ*. 2013;14(suppl 1):33–42.
19. Bliemer MCJ, Rose JM, Hess S. Approximation of bayesian efficiency in experimental choice designs. *J Choice Model*. 2008;1:98–126.
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 at: 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. EQ-5D value sets: inventory, comparative review and user guide. In: Szende A, Oppe M, Devlin N, eds. *EuroQol Group Monographs*. Vol. 2. Dordrecht: Springer; 2007.
22. StataCorp. *Stata Statistical Software*. College Station, TX: StataCorp LP; 2011.
23. Badia X, Roset M, Herdman M, et al. 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.
24. Hsee CK, Dube J-P, Zhang Y. The prominence effect in Shanghai apartment prices. *J Market Res*. 2008;45:133–144.
25. Fischer GW, Hawkins SA. Strategy compatibility, scale compatibility, and the prominence effect. *J Exp Psychol*. 1993;19:580–597.
26. Tversky A, Sattath S, Slovic P. Contingent weighting in judgment and choice. *Psychol Rev*. 1988;95:371–384.
27. Carmon Z, Simonson I. Price-quality trade-offs in choice versus matching: new insights into the prominence effect. 1998. *J Consum Psychol*. 1998;7:323–343.
28. Oliver A. Further evidence of preference reversals: choice, valuation and ranking over distributions of life expectancy. *J Health Econ*. 2006;25:803–820.
29. Sumner W II, Nease RF Jr. Choice-matching preference reversals in health outcome assessments. *Med Decis Making*. 2001;21:208–218.
30. Bleichrodt H, Pinto JL. Loss aversion and scale compatibility in two-attribute trade-offs. *J Math Psychol*. 2002;46:315–337.
31. Hawkins SA. Information processing strategies in riskless preference reversals: the prominence effect. *Organ Behav Hum Decis Proces*. 1994;59:1–26.
32. Hsee CK, Loewenstein GF, Blount S, et al. Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. *Psychol Bull*. 1999;125:576–590.