

2

An EQ-5D-5L value set based on Uruguayan population preferences

*Augustovski F^{1,2}, Rey-Ares L¹, Irazola V¹, Garay OU¹, Gianneo O³,
Fernández G³, Morales M³, Gibbons L¹, Ramos-Goñi, JM^{4,5}*

ABSTRACT

Purpose

To derive a value set from Uruguayan general population using the EQ-5D-5L questionnaire and report population norms.

Methods

General population individuals were randomly assigned to value 10 health states using composite Time Trade Off and 7 pairs of health states through Discrete Choice Experiments. A stratified sampling with quotas by location, gender, age and socioeconomic status was used to respect the Uruguayan population structure. Trained interviewers conducted face to face interviews. The EuroQol valuation technology was used to administer the protocol as well as to collect the data. OLS and maximum likelihood robust regression models with or without interactions were tested.

Results

We included 794 respondents between 20 and 83 years. Their characteristics were broadly similar to the Uruguayan population. OLS model showed inconsistencies, and the main effects robust model was chosen to derive social values. Values ranged from -0.264 to 1. States with a misery index=6 had a mean predicted value of 0.965. When comparing the prediction for a misery index = 6 in the Uruguayan population with the Argentinian EQ-5D-5L crosswalk value set, the Uruguay values are about

-
1. Instituto de Efectividad Clínica y Sanitaria (IECS), Buenos Aires, Argentina
 2. Escuela de Salud Pública, Facultad de Medicina, Universidad de Buenos Aires, Argentina
 3. Fondo Nacional de Recursos, Montevideo, Uruguay
 4. The EuroQol Research Foundation. Rotterdam, The Netherlands
 5. Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC), Spain

0.05 higher. The mean index values, using the selected Uruguayan EQ-5D-5L value set, for the general population in Uruguay is 0.895. In general, older people had worse values and males have slightly better values than females.

Conclusion

We derived the EQ-5D-5L Uruguayan value set, the first in Latin America. These results will help inform decision making using economic evaluations for resource allocation decisions.

INTRODUCTION

Worldwide, healthcare funders are increasingly using economic evaluations to inform their decisions related to health care and the adoption of new technologies. There are many multi-attribute utility-based instruments (MAUI) for the assessment of Quality of Life (QoL), the Canadian Health Utilities Index (HUI) 3, the Finnish 15D, the SF6D and the EQ-5D, the most widely used MAUI in published cost-utility analyses^[1,2]. The National Institute for Health and Care Excellence (NICE) from England and Wales states that health effects in cost-effectiveness analyses should be expressed in quality adjusted life years (QALYs), utilities should be based on public preferences, and recommend the use of the EQ-5D^[3].

The EQ-5D is a generic instrument commonly used to measure patient-reported QoL. In order to help inform decision-making in economic evaluations, it is used to assign a preference value to the amount of time living on the reported health status. Many countries, some from Latin America (LA), have derived population value sets for the EQ-5D^[4-7].

The classic version of the EQ-5D comprises five dimensions with three severity levels and a visual analogue scale (EQ-VAS)^[8]. Dimensions are mobility, self-care, usual activities, pain/discomfort; and anxiety/depression. Recently, the EuroQol Group developed a new version of the instrument with the same dimensions but five levels of response instead of the original three in the classic version, now called: EQ-5D-3L. The new labels for the level response range from no problems to unable/extreme problems with three intermediate levels (slight, moderate and severe problems). This new version, called EQ-5D-5L, describes 3,125 possible health states and it was designed to improve the measurement properties of the EQ-5D-3L, reducing the ceiling effect and increasing the discriminatory power^[9]. Recent studies confirm the higher discriminatory power and lower ceiling effect for the 5L version as compared to the 3L^[10-13]. A five-digit number (one per dimension, on the same order than in the instrument) usually codifies the EQ-5 D health states; this code is usually called state profile. Each digit varies from 1 to 5 representing no problems level to unable/extreme problems level, respectively. For example, the worst state

(“pits”) is represented by 55555. While being unable to walk but with no problems in the rest of dimensions is represented by 51111. The misery index is a proxy measure of the severity of the state. It is defined as the addition of the five digits of its profile, for example the state 23221 has a misery index equal to 10. The EuroQol have recently developed an EQ-5D-5L valuation study protocol aiming to harmonize 5L valuation studies around the world^[14]. To implement and facilitate this standard protocol, the EuroQol group developed specific software, the EuroQol Valuation Technology (EQ-VT)^[15].

Uruguay is located in the southern cone of South America and according to the last census has 3,286,314 inhabitants^[16]. By the end of 2007 Uruguay started a health care reform, encompassing health care delivery, financing and management, establishing an Integrated National Health System with the objective of improving the quality, access and efficiency of health care services. Since the reform the Ministry of Health regularly uses Health Technology Assessment (HTA) and economic evaluations in decision making process involving the financial coverage and provision of high cost technologies for the whole population, through the National Resource Fund (Fondo Nacional de Recursos)^[17,18]. Uruguay had no previous value set, and it is the first country in LA to undertake a general population valuation study for the EQ-5D-5L instrument. The objectives of our study were to obtain social preferences and derive a value set from Uruguayan general population using the EQ-5D-5L questionnaire, as well as to report the population norms.

METHODS

Protocol

We used a standardized interview protocol developed by the EuroQol group based on the obtained evidence from a set of conducted pilot studies^[19]. The protocol consists in five different sections:

- (i) a general welcome,
- (ii) introduction to the research and completion of background information,
- (iii) the set of composite time trade-off (C-TTO) tasks,
- (iv) the set of discrete choice (DC) experiment tasks, and
- (v) general thank you and goodbye.

We also collected socio-demographic information and health literacy measurement using the Short Assessment of Health Literacy-Spanish questionnaire (SAHL-S).

Eliciting preferences methods

Traditionally, EQ-5D-3L valuation studies were mainly based on time trade-off (TTO) methods^[4]. However, the TTO version used in the “Measurement and Value of Health” protocol^[20] had some problems, especially with the transformation of values from states considered to be worse than dead^[21,22]. Lead-Time TTO, Lag-Time TTO^[23] and C-TTO^[24], more recent TTO versions were tested on several pilot studies. Based on the results of those studies the EuroQol Group concluded that the EQ-5D-5L protocol should include the C-TTO version^[14].

The conventional TTO approach has two different tasks, one for states considered to be better than death (BTD) and another one for states considered to be worse than death (WTD). For the valuation of states BTD respondents are required to choose between living 10 years in a specific health state (life B) or X years in full health (life A). The amount of time X in life A varied between 0 and 10 years. For states WTD, the procedure is conceptually and operationally different, participants have to choose between dying immediately or live X years in a specific state followed by 10-X years in full health (life B).

Lead-time TTO and lag-time TTO methods add extra time in full health to the health state to be valued (Life B). The lead-time approach used in this protocol places the extra time before, being the options living 10 years in full health and later 10 years in the given health state or X amount of time (between 0 and 20 years) in full health in life A. The main characteristic of these variants is that the iterative trading process allows the participant to move between negative and positive values without thinking about whether the state is worse or better than being dead.

C-TTO involves the use of traditional TTO approach for states considered to be better than dead, and a lead-time TTO for states considered to be WTD, combined in a unique task. When the participant exhausted the 10 years in full health in the traditional TTO task is asked to inform how many years in full health is willing to resign to avoid it using the lead-TTO task.

The interview protocol included 86 EQ-5D-5L health states (selected using Monte Carlo simulation^[14]) to be evaluated using C-TTO divided in ten blocks with similar representation of all severity levels. All the blocks included one very mild state with only one health state with mild problems (i.e: 21111) and the pits state (55555). Respondents were randomly assigned to one of the ten C-TTO blocks. The presentation order of the states within each block was also randomly generated by the EQ-VT.

The protocol included a DC experiment as a secondary valuation technique^[14]. A DC experiment is an ordinal elicitation technique that has received recent attention for

eliciting EQ-5D-5L values^[25,26]. DC experiments require individuals to make a pairwise comparison between two different scenarios, being in our case two EQ-5D-5L health states. The protocol included 196 DC pairs of EQ-5D-5L health states divided in 28 blocks of seven pairs with similar misery index. Respondents were randomly assigned to one DC block, the order of each pair as well as its position on the screen (i.e. left and right) were also randomly generated by the EQ-VT.

Quality control

The initial EQ-5D-5L valuation studies found some interviewer effects that could affect data quality, as it is reported by Ramos-Goñi et al^[27]. For further valuation studies the EuroQol Group decided to create a quality control tool to monitor interviewer performance. This tool mainly evaluates interviewer protocol compliance through 4 key parameters of the C-TTO task (the explanation and the time spent of the wheel chair example, the time used to complete the 10 C-TTO tasks and the presence of large inconsistencies). The DC experiments section was monitored to detect unusual response patterns (i.e. AAAAAAA, ABABABA). Interview quality was checked weekly. Based on weekly results we decided whether we had to retrain or drop interviewers from the team. After quality control analysis, researchers (FA, UG and LRA) gave feedback and retrained interviewers when necessary.

Sampling and data collection

Uruguay was geographically stratified. The study took place in the following locations: Uruguay capital city, Montevideo; and the departments of Maldonado and Paysandú. Quotas by location, gender, age and socioeconomic status replicated the Uruguayan population structure^[16].

The EuroQol Group recommends including 10,000 C-TTO responses in the valuation studies. Since each participant values 10 health states the initial sample size was 1,000 individuals. The power calculation was based on precision requirements for the estimation of the C-TTO mean^[27].

Twenty-one trained interviewers administered the questionnaire using the EQ-VT. During the face to face interviews, respondents had the control of the computer most of the time and the interviewers were available to assist and monitor the process.

The valuation exercise started explaining the objectives of the research, then the respondents filled out the EQ-5D-5L and rated their current health state using the EQ-VAS. Additionally they gave background information (age, sex, educational level, and their experience with illness). Prior to completion of the C-TTO tasks, participants received an explanation (using as example a life living in wheel chair)

and completed three mock states of different severity in order to verify their understanding. Later on, they completed 10 C-TTO and 7 DC experiments. Upon completion of the tasks, participants answered follow-up questions related to the difficulty and comprehension.

Once both tasks and the follow-up questions were completed, the interviewers asked the respondents to complete the Short Assessment of Health Literacy–Spanish (SAHL-S) instrument^[28]. This questionnaire, evaluates health literacy through 18 multiple-choice questions combining word recognition and comprehension. Low health literacy is defined by identifying 14 or fewer correct items. Respondents were asked to read aloud 18 medical terms (word recognition), and the interviewer assessed comprehension through the multiple-choice question. Health literacy is a construct that reflects the capacity to obtain, process and understand health information and services needed to make appropriate health decisions^[29].

Statistical analysis

We describe the sample characteristics using means and standard deviations for continuous variables and percentages for discrete variables. The EQ-5D-5L descriptive system from the recruited sample is presented by age group.

Valuation data from C-TTO tasks and data from DC experiments were available from the collected data. We initially tried to follow the hybrid approach reported by Ramos-Goñi et al^[27]. However, in our case the DC models had several logical inconsistencies, leading us to base our analysis on the C-TTO data only. The DC and hybrid analyses are available from the authors upon request.

We started the C-TTO analysis using the classic ordinary least square (OLS) model. However, our data had problems that prevented us from using this approach: heteroscedasticity and significant outliers. Thus, we opted for using robust regression^[30]. This regression method basically applies a different weight to each observation based on how far away it is located from the median of the population sample. In this way the impact of the outliers is reduced, and the heteroscedasticity problem is addressed. We used a tuning constant based on the bi-weight function to calculate each weight in the model. A tuning constant of 7 is usually used, in order to confer similar model efficiency as the OLS model, and assuming no heteroscedasticity or outliers^[31]. We set this value to 8.5 to include a broader range of values for each respondent, without losing the logical consistency of the model. We analysed the distribution of weights according to the misery index value in order to explore the contribution of different groups of observations to the final model estimations.

For the model specification, we started with a 20-parameter main effects model, using the responses values as dependent variables and health states as explanatory variables. We created a dummy variable D_{ij} indicating whether the dimension i is at level j . For example, we created variables MO2, MO3, MO4 and MO5 for mobility dimension, indicating whether the mobility dimension is at level 2 or 3 or 4 or 5 respectively. Similar sets of variables were created for each dimension. In order to explore alternative model specifications and performance, we added interaction terms to the main effects model. We evaluated traditional N_j terms (1 if at least one dimension is at level j) and N_{ij} terms (1 if at least one dimension is at level i or j). We also tested the following interactions terms:

- (i) D1, Number of movements away from full health beyond the first;
- (ii) I2, Number of dimensions at level 2 or 3 beyond the first;
- (iii) C3, Number of dimensions at level 3, 4 or 5 beyond the first;
- (iv) K45, Number of dimensions at level 4 or 5;
- (v) I45, Number of dimensions at level 4 or 5 beyond the first;
- (vi) O2, 1 if all dimensions are at level 1 or 2;
- (vii) Z2, 1 if at least one dimension is at level 2 or 3 and one is at level 4 or 5;
- (viii) Z3, number of dimensions at level 2 or 3 given that at least one dimension was at level 4 or 5.

We used a stepwise approach to decide whether to keep the interaction terms in the model or leave them out.

In this manuscript, we present three models:

- (i) the OLS model for comparison purpose,
- (ii) the main effects robust model and
- (iii) the robust model including best interaction terms.

All statistical analyses were performed on STATA 11 MP[26], using the “regress” command for OLS and “rreg” for the robust regression.

Expanding the modelling exercise, we have performed additional analyses to check differences on preference values by educational level. We have performed an ANOVA test for crude preferences and we have also added dummies by levels of education on the final estimation of the model. None of these results were statistically significant (data not shown).

Exclusion criteria.

Interviewers with low quality performance, fulfilling pre-specified criteria regarding interview quality (i.e too little time spent explaining the task, no explanation of the

lead time section, C-TTO responses with clear inconsistencies or too little time to perform all TTO tasks were excluded during the data collection process. We also excluded respondents meeting two additional criteria:

- (i) having a positive slope on the relationship between their values and the misery index of the health states. This means, that the respondent poorly understood the task, as he/she provided higher utility values for worse health states; and
- (ii) respondents who valued all states at the same value, except non-traders (i.e. subjects who value all states as 1).

Model performance

We used four criteria to evaluate the performance of the model:

- (i) logical consistency of parameters,
- (ii) goodness of fit,
- (iii) prediction accuracy and
- (iv) parsimony.

A set of model parameters is said to be logically consistent if predictions for logically better health states (ex: 12111 is logically better than 13111) is higher than the predictions for logically worse health states. In our models, it means that $MO2 < MO3 < MO4 < MO5$, and so on for SC, UA, AD and PD dimensions. We used Akaike (AIC) and Bayesian information criteria (BIC) to evaluate goodness of fit, adjusted by the number of model parameters. We calculated the mean square error (MSE) and the mean absolute error (MAE) to evaluate prediction accuracy. The principle of parsimony stated: When two competing models are similar in terms of performance parameters, the simplest model should be selected. These four criteria were used to compare different model specifications using different interaction terms.

Comparison between predicted values from different models and to the Argentinian crosswalk 5L value set

In order to compare predictions from different models, we calculated the estimated values from robust models and compared them with the weighted means of the 86 TTO health states included. Those predictions were also compared to the ones from the crosswalk 5L value set from Argentina, (the 3L value set that used a mapping function between the 3L and the 5L versions). Because of the EQ-5D-5L cross-walk value set for Argentina has not been previously estimated^[32], it was calculated specifically for this study following the methodology proposed by van Hout et al at 2012^[33].

RESULTS

Study recruitment took place between October 2013 and June 2014. We started the field work with 21 interviewers, who were trained, and evaluated on a weekly basis at the beginning of the data collection phase. Based on the quality control analysis, we decided to keep 11 interviewers with good performance, which resulted in the exclusion of 220 interviews conducted at this stage by 10 poor quality interviewers. Excluded participants had similar age, gender and educational level than the remaining sample. We periodically analysed the remaining interviews during data collection and decided to stop data collection when our analysis showed the robustness of the results. The study sample had 805 respondents between 20 and 83 years old. Eleven subjects met exclusion criteria, leaving 794 subjects in the final sample (see Table 2.1). Sample characteristics were similar to the Uruguayan population in terms of gender. However, younger as well as higher educated categories were slightly over-represented in our sample, though utility values did not significantly differ by educational level. Nearly 44% of the population had low health literacy despite the fact that more than 80% had educational attainment of at least some secondary level.

Table 2.1 Study sample and Uruguayan general population characteristics

Socio-demographic characteristics	Study sample (794)		Uruguayan population (Census 2011**)
	(n)	%	%
Age group (years)			
20-39 [#]	386	48.6%	42.4
40-59	271	34.1%	35.2
60+ [#]	137	17.3%	22.4
Female	439	55.3%	52.0
Educational attainment *			
Primary level	137	17.3%	36.2
Secondary level	406	51.3%	44.5
Tertiary level	249	31.4%	18.5
Low Health literacy *	346	43.7%	
SALHS score – Mean (SD) *	14.3	(2.5)	
Experience with serious illness			
in itself	180	22.7	
in family	450	56.7	
in caring for others	425	53.5	

*Not all responses for these questions complete. We estimate the denominator with the non-missing values.

** Percentages for the age categories were calculated using the population between 20 and 79 years for the ease of comparison with our sample, and the census estimations for educational level comprise inhabitants older than 25 years of age. SALHS: Short Assessment of Health Literacy-Spanish questionnaire; SD: Standard Deviation.

[#] The 20-39 group was slightly overrepresented and the 60+ group was slightly underrepresented (results statistically significant).

Forty four percent of the sample reported no problems on any dimension of self-reported EQ-5D-5L. Older respondents reported more problems in all dimensions and the mean self-reported VAS also decreased with increasing age, and was smaller in women (Table 2.2).

Table 2.2 Self-reported health using EQ-5D-5L descriptive system and EQ VAS

		Age		20-39		40-59		60+		Total	
		n	%	n	%	n	%	n	%		
Mobility	No problems	367	95.10	214	79.00	84	61.31	665	83.80		
	Slight problems	16	4.10	39	14.40	31	22.63	86	10.80		
	Moderate problems	2	0.50	13	4.80	16	11.68	31	3.90		
	Severe problems	1	0.30	5	1.80	6	4.38	12	1.50		
	Unable to walk	0	0.00%	0	0.00%	0	0.00%	0	0.00%		
Self-Care	No problems	383	99.20	257	94.80	124	90.51	764	96.20		
	Slight problems	2	0.50	9	3.30	8	5.84	19	2.40		
	Moderate problems	0	0.00	2	0.70	4	2.92	6	0.80		
	Severe problems	1	0.30	3	1.10	1	0.73	5	0.60		
	Unable to	0	0.00%	0	0.00%	0	0.00%	0	0.00%		
Usual activities	No problems	365	94.60	234	86.40	104	75.91	703	88.50		
	Slight problems	15	3.90	18	6.60	25	18.25	58	7.30		
	Moderate problems	5	1.30	14	5.20	6	4.38	25	3.10		
	Severe problems	0	0.00	2	0.70	1	0.73	3	0.40		
	Unable to	1	0.30%	3	1.10%	1	0.73%	5	0.60%		
Pain/ Discomfort	No problems	275	71.20	166	61.30	73	53.28	514	64.70		
	Slight problems	87	22.50	67	24.70	41	29.93	195	24.60		
	Moderate problems	19	4.90%	20	7.40%	14	10.22	53	6.70%		
	Severe problems	4	1.00%	16	5.90%	9	6.57%	29	3.70%		
	Extreme problems	1	0.30%	2	0.70%	0	0.00%	3	0.40%		
Anxiety/ Depression	No problems	252	65.30	170	62.70	94	68.61	516	65.00		
	Slight problems	91	23.60	63	23.30	27	19.71	181	22.80		
	Moderate problems	35	9.10%	24	8.90%	10	7.30%	69	8.70%		
	Severe problems	7	1.80%	13	4.80%	5	3.65%	25	3.10%		
	Extreme problems	1	0.30%	1	0.40%	1	0.73%	3	0.40%		
Visual Analogue Scale	Mean	83.19		78.49		71.81		79.63			
	Standard error	0.68		1.04		1.66		0.58			
	25 th Percentile	80		70		55		70			
	50 th Percentile	85		80		75		80			
	75 th Percentile	90		90		90		90			

The results from the OLS model showed a logical inconsistency between the coefficients associated with slight and moderate problem in usual-activities dimension (Table 2.3). This inconsistency was not observed when using the same model specification with a robust regression approach. Both robust models estimations reported, the main effects and the main effects with interactions were logically consistent. The goodness of fit of both robust models was similar, gaining only 0.4% of relative improvement in AIC or BIC with the interaction terms. Similarly, the prediction accuracy of the main effect robust model was similar to the main effect model with interaction (MSE=0.002). Based on the similarities of these parameters and taking into account the parsimony criterion, we chose the robust estimation of the main effects model as the most appropriate model for the Uruguayan value set. The technical appendix shows how to estimate an individual state value, as well as the code to do it in Stata. In the electronic supplementary material, we report the utility values for the 3,125 different states.

Table 2.3 Model estimations: comparison of OLS, main effect robust model, and main effect with interaction terms.

	OLS			(Final Value set)			Robust estimation model		
	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value
MO2	0.0767	0.019	0.00	0.0140	0.016	0.37	0.0514	0.017	0.00
MO3	0.1019	0.020	0.00	0.0322	0.016	0.05	0.0857	0.020	0.00
MO4	0.1906	0.022	0.00	0.1077	0.018	0.00	0.0930	0.020	0.00
MO5	0.3435	0.020	0.00	0.2987	0.016	0.00	0.2688	0.021	0.00
SC2	0.0249	0.019	0.18	0.0256	0.015	0.09	0.0529	0.015	0.00
SC3	0.0820	0.021	0.00	0.0609	0.017	0.00	0.0875	0.021	0.00
SC4	0.1282	0.021	0.00	0.1169	0.017	0.00	0.0801	0.021	0.00
SC5	0.2616	0.019	0.00	0.2734	0.016	0.00	0.2120	0.023	0.00
UA2	0.0710	0.020	0.00	0.0424	0.016	0.01	0.0691	0.016	0.00
UA3	0.0512	0.021	0.01	0.0455	0.017	0.01	0.0714	0.018	0.00
UA4	0.1303	0.021	0.00	0.1183	0.017	0.00	0.0905	0.019	0.00
UA5	0.2101	0.019	0.00	0.2315	0.016	0.00	0.1775	0.022	0.00
PD2	0.0260	0.018	0.14	0.0171	0.014	0.23	0.0450	0.015	0.00
PD3	0.0820	0.021	0.00	0.0607	0.017	0.00	0.0974	0.022	0.00
PD4	0.2160	0.019	0.00	0.1870	0.015	0.00	0.1511	0.022	0.00
PD5	0.2833	0.021	0.00	0.2705	0.017	0.00	0.2184	0.022	0.00
AD2	0.0320	0.020	0.11	0.0095	0.016	0.55	0.0329	0.014	0.00
AD3	0.0884	0.022	0.00	0.0435	0.018	0.01	0.0885	0.019	0.02
AD4	0.1509	0.021	0.00	0.1043	0.017	0.00	0.0832	0.018	0.00
AD5	0.1809	0.019	0.00	0.1771	0.016	0.00	0.1381	0.019	0.00
Const.	0.0104	0.020	0.61	0.0126	0.016	0.44	-	-	-
D1							-0.0192	0.015	0.00
I45^2							0.0140	0.002	0.19

Table 2.3 Model estimations: comparison of OLS, main effect robust model, and main effect with interaction terms. (Continued)

	OLS			(Final Value set)			Robust estimation model		
	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value	Coefficient	Std. error	P-value
LogL	-5707.448			-3290.65			-3274.3		
AIC	11456.9			6623.3			6592.6		
BIC	11603.5			6769.9			6746.1		
MSE	0.003			0.002			0.002		
MAE	0.04			0.03			0.03		
U(55555)	-0.28			-0.26			-0.28		

Bold values indicate logical inconsistencies. Parameter abbreviations are described in the text. LogL: Log likelihood; AIC: Akaike information criterion; BIC: Bayesian information criterion; MSE: mean standard error; MAE: mean absolute error; U(55555): utility value of the pits state.

The weights applied to the responses in the robust model range from 0 to 1. The distribution of weights, according to different values of the misery index, showed how the respondents' opinions have more discrepancies as the misery index increases (Table 2.4). For example, for a misery index=6, the 10th percentile is 0.992, meaning that 90% of the responses have almost the same impact in the model estimations (i.e. weight close to 1). However, the 10th percentile for those states with a misery index= 21 is 0.449, meaning that at least 10% of the responses for these states will be considered as half important in the model estimations (i.e. weight close to 0.5). The selected model predicts index values that range from -0.264 to 1. States with a misery index=6 had a mean predicted value of 0.965 (Table 2.4). The predictions from the model with interactions are lower at the top and bottom of the scale than predictions from the main effects model; i.e., it has a lowest prediction of -0.288 for the pits state and has a mean prediction for misery index 6 states of 0.95. When comparing the prediction for a misery index = 6 in the Uruguayan population with the Argentinian EQ-5D-5L crosswalk value set, the Uruguay values are about 0.05 higher, and they are also higher than the Argentinian across the misery index spectrum.

Table 2.4 Model predictions comparison, and with the crosswalk 3L Argentinian value set

Misery index	Percentile 10% for Weights	Weighted observed TTO values	Robust Estimation Model	Robust estimation model with interactions	3L Argentinian value set by cross-walk mapping
6	0.992	0.953	0.965	0.950	0.906
7	0.987	0.927	0.947	0.925	0.845
9	0.966	0.855	0.878	0.871	0.750
10	0.942	0.796	0.827	0.837	0.732
11	0.938	0.761	0.801	0.790	0.689
12	0.943	0.744	0.732	0.741	0.517
13	0.895	0.666	0.650	0.667	0.454

Table 2.4 Model predictions comparison, and with the crosswalk 3L Argentinean value set (Continued)

Misery index	Percentile 10% for Weights	Weighted observed TTO values	Robust Estimation Model	Robust estimation model with interactions	3L Argentinean value set by cross-walk mapping
14	0.819	0.600	0.593	0.608	0.407
15	0.860	0.557	0.544	0.558	0.403
16	0.843	0.532	0.514	0.520	0.376
17	0.848	0.449	0.462	0.457	0.421
18	0.730	0.358	0.340	0.351	0.229
19	0.758	0.354	0.285	0.296	0.164
20	0.735	0.215	0.252	0.224	0.175
21	0.449	0.095	0.161	0.143	0.082
22	0.613	0.150	0.140	0.138	0.069
25	0.707	-0.300	-0.264	-0.288	-0.376

The mean index values, using the selected Uruguayan EQ-5D-5L value set, for the general population in Uruguay is 0.954. Older people have worse health related quality of life for all paired comparisons (highest P-value < 0.001). Males had slightly higher values than females (Table 2.5), but this difference was not significant.

Table 2.5 Population norms for EQ-5D-5L in Uruguay

EQ-5D-5L (Index values)		Age	20-39	40-59	60+	Total
Total sample	Mean		0.972	0.942	0.93	0.954
	Standard error		0.003	0.006	0.009	0.003
	25th Percentile		0.961	0.927	0.904	0.945
	50th Percentile (median)		1	0.973	0.961	0.978
	75th Percentile		1	1	1	1
Males	Mean		0.974	0.951	0.957	0.963
	Standard error		0.003	0.01	0.009	0.004
	25th Percentile		0.967	0.963	0.94	0.961
	50th Percentile (median)		1	0.978	0.978	0.978
	75th Percentile		1	1	1	1
Females	Mean		0.97	0.935	0.907	0.947
	Standard error		0.004	0.007	0.014	0.004
	25th Percentile		0.961	0.916	0.867	0.935
	50th Percentile (median)		1	0.97	0.947	0.978
	75th Percentile		1	1	0.989	1

DISCUSSION

The EQ-5D-5L is a recently developed instrument. Only a few countries conducted valuation studies, and only the Spanish study was recently published, though authors did not recommend the use of the reported value set^[27]. This is to our knowledge the first study that provides a population-based EQ-5D-5L value set in Latin America.

The choice of a robust estimation for modelling the C-TTO data is also a novel approach in EQ-5D valuation exercises. While other studies present models mainly based in OLS or random effects estimations^[4], we selected a robust estimation based on the observed between respondents variability. Some extreme differences in opinions in our sample made some of the OLS or random effects coefficients to be logically inconsistent, due to the significant heteroscedasticity and the presence of outliers. Robust regression tries to solve these issues by weighting opinions less strongly if they are extreme. Extreme opinions have less impact depending on how extreme they are, and the values close to the majority (median) have the greatest weight. In our estimation, we relaxed the robust condition of the estimation as much as possible, stopping when inconsistencies in coefficients were found. We included all responses from the sample, as there is no 0 weight for any response. This was a balance point we chose in order to have logical results but also to incorporate everybody's opinions.

Based on the parsimony criteria, we think that a more complex model should be preferred only when the improvement compared to a simpler one is large enough to overcome the complexity. In our estimation, the improvement of the best model with interactions terms tested was marginal compared to the main effects model. That led us to prefer the main effects model instead of the model with the interaction terms.

Given the fact that there is not any EQ-5D-5L value set currently recommended in the literature, and no previous EQ-5D-3L value set was available for Uruguay, we decided to compare our results with the EQ-5D-5L crosswalk value set derived from the original 3L set for Argentina^[5], being a close country with similar socio-economic characteristics. In our selected model, we had slightly higher values in Uruguay in the entire severity spectrum. Taking into account the changes on the descriptive system of the 3L and 5L versions of EQ-5D^[9], it was something expected. For example, the levels for the misery index 6 states on the 5L version are by definition less severe ("slight") compared to the same levels in the 3L version ("some"). Also, the higher observed index value for the pits state can be explained, as the level for the mobility dimension has changed from "confined to bed" in the 3L version, to "unable to walk" in the 5L version, making the description of the pits state (55555) in the 5L version, better than the corresponding state (33333) in the 3L version. As both anchor values have been moving up in our estimations, it is expected that the whole scale move up according these anchors. Populations norms

derived for Uruguay showed to be consistent and similar to international population norms previously published^[34].

One limitation of this study is the fact that young and higher educated individuals were slightly over-represented in our sample in comparison to the population distribution of Uruguay. However, this age difference was small, and utility values did not significantly differ according to educational level. Therefore, we think that the values obtained in this study can be considered representative of the whole population. Another limitation is the fact that we have not used the information from the DC experiment. The recently published study from Spain reported the feasibility of obtaining an EQ-5D-5L value set using a hybrid approach, combining the C-TTO and DC data^[27]. However in our initial analysis, when we tested our DC models, they showed several inconsistencies, and these could not be solved through a hybrid modelling approach. In addition, we have seen in our C-TTO responses some extreme differences in opinions, mainly regarding severe health states. This fact could also explain the inconsistencies found in the DC models. As far as we are aware, there is no available “robust” estimation method for analysing DC data, limiting our capacity to include this information in our estimations. Another limitation of this study is the fact that we have not performed an internal or an external validation of our predictions. Given the final sample size of our study, which was somewhat smaller than originally intended, and that the requirements for internal validation reduce statistical power (i.e. randomly splitting the sample and evaluate how the model derived in the estimation set applies to the validation set), we included all responses in the model estimation. Finally, our sample was broadly representative of the socio-demographic characteristics of Uruguayan population, but had a somewhat higher education level.

We obtained the EQ-5D-5L value set that will be implemented in Uruguay, which is the first country in Latin America to undertake such a study. The use of these values will help researchers, in Uruguay and eventually in other similar socio-economic countries, in conducting cost-utility studies based on the specific preferences of the general population to inform decision makers’ resource allocation decisions.

ACKNOWLEDGEMENTS

This Project was funded by Uruguay's National Resources Fund (Fondo Nacional de Recursos) and the EuroQol Research Foundation. Special thanks to Arnd Jan Prause for the support and Elka Pérez and Gastón Díaz from “Equipos Mori” for their great work and commitment.?

REFERENCES

- [1] Hawthorne, G., Richardson, J., & Day, N. A. (2001). A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Ann Med*, 33(5), 358-370.
- [2] Brauer, C. A., Rosen, A. B., Greenberg, D., & Neumann, P. J. (2006). Trends in the measurement of health utilities in published cost-utility analyses. *Value Health*, 9(4), 213-218.
- [3] Guide to the methods of technology appraisal 2013. (2013). Process and methods guides.: National Institute for Health and Care Excellence (NICE).
- [4] Szende, A., Oppe, M., & Devlin, N. (2007). EQ-5D Value Sets: Inventory, Comparative Review and User Guide (Vol. 2). Dordrecht: Springer.
- [5] Augustovski, F. A., Irazola, V. E., Velazquez, A. P., Gibbons, L., & Craig, B. M. (2009). Argentine valuation of the EQ-5D health states. *Value Health*, 12(4), 587-596.
- [6] Zarate, V., Kind, P., Valenzuela, P., Vignau, A., Olivares-Tirado, P., & Munoz, A. (2011). Social valuation of EQ-5D health states: the Chilean case. *Value Health*, 14(8), 1135-1141.
- [7] Viegas Andrade, M., Noronha, K., Kind, P., Maia, A. C., Miranda de Menezes, R., De Barros Reis, C., Nepomuceno Souza, M., Martins, D., Gomes, L., Nichele, D., Calazans, J., Mascarenhas, T., Carvalho, L., & Lins, C. (2013). Societal Preferences for EQ-5D Health States from a Brazilian Population Survey. *Value in Health Regional Issues*, 2(3), 405-412.
- [8] EuroQol, G. (1990). EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy*, 16(3), 199-208.
- [9] Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonse, G., & Badia, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*, 20(10), 1727-1736.
- [10] Pattanaphesaj, J., & Thavorncharoensap, M. (2015). Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in the Thai diabetes patients. *Health Qual Life Outcomes*, 13(1), 14.
- [11] Golicki, D., Niewada, M., Buczek, J., Karlinska, A., Kobayashi, A., Janssen, M. F., & Pickard, A. S. (2015). Validity of EQ-5D-5L in stroke. *Qual Life Res*, 24(4), 845-850.

- [12] Conner-Spady, B. L., Marshall, D. A., Bohm, E., Dunbar, M. J., Loucks, L., Khudairy, A. A., & Noseworthy, T. W. (2015). Reliability and validity of the EQ-5D-5L compared to the EQ-5D-3L in patients with osteoarthritis referred for hip and knee replacement. *Qual Life Res.*
- [13] Greene, M. E., Rader, K. A., Garellick, G., Malchau, H., Freiberg, A. A., & Rolfson, O. (2014). The EQ-5D-5L Improves on the EQ-5D-3L for Health-related Quality-of-life Assessment in Patients Undergoing Total Hip Arthroplasty. *Clin Orthop Relat Res.*
- [14] Oppe, M., Devlin, N. J., van Hout, B., Krabbe, P. F., & de Charro, F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health, 17*(4), 445-453.
- [15] EuroQol Valuation Technology. (2011). Retrieved January 2015, 2015, from <https://www.valuationstudy.org>
- [16] Censo 2011. (2011). Retrieved September 2014, 2014, from <http://www.ine.gub.uy/censos2011/index.html>
- [17] Modificación de la integración de la comisión asesora del Formulario Terapéutico Nacional. (2010). Decreto N° 04/010 (January 7, 2010 ed.). Montevideo: Presidencia República Oriental del Uruguay.
- [18] Fondo Nacional de Recursos. Retrieved January 2015, 2015, from <http://www.fnr.gub.uy>
- [19] Devlin, N. J., & Krabbe, P. F. (2013). The development of new research methods for the valuation of EQ-5D-5L. *Eur J Health Econ, 14* Suppl 1, S1-3.
- [20] Dolan, P. (1997). Modeling valuations for EuroQol health states. *Med Care, 35*(11), 1095-1108.
- [21] Tilling, C., Devlin, N., Tsuchiya, A., & Buckingham, K. (2010). Protocols for time tradeoff valuations of health states worse than dead: a literature review. *Med Decis Making, 30*(5), 610-619.
- [22] Lamers, L. M. (2007). The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. *Med Care, 45*(3), 238-244.
- [23] Augustovski, F., Rey-Ares, L., Irazola, V., Oppe, M., & Devlin, N. J. (2013). Lead versus lag-time trade-off variants: does it make any difference? *Eur J Health Econ, 14* Suppl 1, S25-31.

- [24] Janssen, B. M., Oppe, M., Versteegh, M. M., & Stolk, E. A. (2013). Introducing the composite time trade-off: a test of feasibility and face validity. *Eur J Health Econ*, 14 Suppl 1, S5-13.
- [25] Krabbe, P. F., Devlin, N. J., Stolk, E. A., Shah, K. K., Oppe, M., van Hout, B., Quik, E. H., Pickard, A. S., & Xie, F. (2014). Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. *Med Care*, 52(11), 935-943.
- [26] Ramos-Goni, J. M., Rivero-Arias, O., Errea, M., Stolk, E. A., Herdman, M., & Cabases, J. M. (2013). Dealing with the health state 'dead' when using discrete choice experiments to obtain values for EQ-5D-5L health states. *Eur J Health Econ*, 14 Suppl 1, S33-42.
- [27] Ramos-Goni, J. M., Pinto-Prades, J. L., Oppe, M., Cabases, J. M., Serrano-Aguilar, P., & Rivero-Arias, O. (2014). Valuation and Modeling of EQ-5D-5L Health States Using a Hybrid Approach. *Med Care*.
- [28] Lee, S. Y., Stucky, B. D., Lee, J. Y., Rozier, R. G., & Bender, D. E. (2010). Short Assessment of Health Literacy-Spanish and English: a comparable test of health literacy for Spanish and English speakers. *Health Serv Res*, 45(4), 1105-1120.
- [29] Selden, C. R., Zorn, M., Ratzan, S. C., & Parker, R. M. (2000). *Health Literacy, Current Bibliographies in Medicine*. Bethesda, MD: National Institutes of Health.
- [30] Huber, P. J. (1964). Robust Estimation of a Location Parameter. 73-101.
- [31] Hamilton, L. (1991). *srd1: How robust is robust regression?* Stata Technical Bulletin Reprints 1, 169-175.
- [32] EQ-5D-5L Crosswalk value sets. Retrieved January 2015, from <http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d/eq-5d-5l-value-sets.html>
- [33] van Hout, B., Janssen, M. F., Feng, Y. S., Kohlmann, T., Busschbach, J., Golicki, D., Lloyd, A., Scalone, L., Kind, P., & Pickard, A. S. (2012). Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*, 15(5), 708-715.
- [34] Szende, A., Janssen, B. M., & Cabasés, J. M. (2014). *Self-Reported Population Health: An International Perspective based on EQ-5D*: Springer Dordrecht Heidelberg New York London.

APPENDIX 2A.

In this manuscript the value set for Uruguay has been presented (see Table 2.3). This appendix describes how to obtain the utility value for a specific health state. Notice that the model coefficients should be interpreted as the disutility of moving from having no problems in that particular domain (level 1) to the specific level of response of each domain.

Given the profile of a specific health state, LMOLSCLUALPDLAD, and given the final model and coefficients to derive them, the formula to obtain the utility value for each health state is as follow:

$$U(\text{LMOLSCLUALPDLAD}) = 1 - \text{MO}(\text{LMO}) - \text{SC}(\text{LSC}) - \text{UA}(\text{LUA}) - \text{PD}(\text{LPD}) - \text{AD}(\text{LAD}) - \text{Deviation from full health}$$

Where $U(\text{LMOLSCLUALPDLAD})$ denotes the utility for the state LMOLSCLUALPDLAD, LMO denotes the response level on mobility domain, $\text{MO}(\text{LMO})$ denotes the coefficient of the level LMO on mobility domain (and the same for rest of domains), and Deviation from full health is the model constant. When the level of a given domain is no problems (1), the coefficient of that domain is 0. As there is no movement from no problems, no disutility is associated.

Example 1:

$$U(25413) = 1 - \text{MO}2 - \text{SC}5 - \text{UA}4 - \text{PD}1 (=0) - \text{AD}3 - \text{Deviation from full health} = 1 - 0.0140 - 0.2734 - 0.1183 - 0 - 0.0435 - 0.0126 = 0.5382$$

Example 2:

$$U(31412) = 1 - \text{MO}3 - \text{SC}1 (=0) - \text{UA}4 - \text{PD}1 (=1) - \text{AD}2 - \text{Deviation from full health} = 1 - 0.0322 - 0 - 0.1183 - 0 - 0.0095 - 0.0126 = 0.8274$$

Example 3:

$$U(11111) = 1 - \text{MO}1 (=0) - 0 (\text{SC}1) - \text{UA}1 (=0) - \text{PD}1 (=0) - \text{AD}1 (=0) = 1 \text{ (Notice that 11111 represents full health, so the application of deviation from full health is not applicable here).}$$

Stata code:

```
//This code calculates the utility values for a given data set
//The variable representing mobility domain has to be named MO, SC for self-care,
UA for usual activities, PD for pain/discomfort and AD for anxiety/depression
```

```
gen Utility = 1.
```

```
recast double Utility
```

//MO

replace Utility = Utility - 0.0140 if MO == 2

replace Utility = Utility - 0.0322 if MO == 3

replace Utility = Utility - 0.1077 if MO == 4

replace Utility = Utility - 0.2987 if MO == 5

//SC

replace Utility = Utility - 0.0256 if SC == 2

replace Utility = Utility - 0.0609 if SC == 3

replace Utility = Utility - 0.1169 if SC == 4

replace Utility = Utility - 0.2734 if SC == 5

//UA

replace Utility = Utility - 0.0424 if UA == 2

replace Utility = Utility - 0.0455 if UA == 3

replace Utility = Utility - 0.1183 if UA == 4

replace Utility = Utility - 0.2315 if UA == 5

//PD

replace Utility = Utility - 0.0171 if PD == 2

replace Utility = Utility - 0.0607 if PD == 3

replace Utility = Utility - 0.1870 if PD == 4

replace Utility = Utility - 0.2705 if PD == 5

//AD

replace Utility = Utility - 0.0095 if AD == 2

replace Utility = Utility - 0.0435 if AD == 3

replace Utility = Utility - 0.1043 if AD == 4

replace Utility = Utility - 0.1771 if AD == 5

//Deviation from full health

replace ` Utility = Utility - 0.0126 if (MO != 1 | SC != 1 | UA != 1 | PD != 1 | AD != 1)