Quality of Life Research (2019) 28:3163–3175
https://doi.org/10.1007/s11136-019-02226-5

A hybrid modelling approach for eliciting health state preferences: the Portuguese EQ-5D-5L value set

Pedro L. Ferreira[1,2] · Patrícia Antunes[1] · Lara N. Ferreira[1,3,4] · Luís N. Pereira[1,3,4] · Juan M. Ramos-Goñi[5]

Accepted: 5 June 2019 / Published online: 14 June 2019

Abstract

Background The EQ-5D is a generic preference-based quality of life measure considered useful for supporting clinical and policy decisions by providing utility values that can easily be converted into quality-adjusted life years to be integrated in cost-utility economic evaluations. Although the three-level classification system of the EuroQol questionnaire (EQ-5D-3L) is still the most popular preference-based instrument used worldwide, several studies reported a ceiling effect on this version, especially in healthy and/or young individuals. In 2009, the EuroQol Group introduced a five-level EQ-5D, which expands the descriptive system from three to five levels within the same five dimensions. For this version to be used in health economic evaluation, societal values need to be assigned to the 3125 health states generated by this instrument.

Objectives The aims of this study were to elicit the EQ-5D-5L health state preferences from the general Portuguese population and to derive the Portuguese value set for the EQ-5D-5L.

Methods A representative sample of the Portuguese general population aged above 18 years was stratified by age and gender ($n = 1451$). Between October 2015 and July 2016, 28 interviewers carried out a series of 1-h-long computer-assisted personal interviews following the EuroQol Valuation Technology protocol. Each interview included the valuation of ten health states using the composite time trade-off (cTTO) and seven pairs of discrete choice experiments (DCEs). A standardized tool for quality control was used to assess the quality of the data as well as direct supervision and cross-examination of 10% of the global sample size. Data from both cTTO and DCE valuation tasks were modelled using a censored heteroskedastic hybrid model.

Results Interviewers complied with the quality control protocol in providing high-quality valuation data. The hybrid econometric model had consistent and significant parameters. The derived societal values for the Portuguese population ranged from − 0.603 to 1.

Conclusion This study provided the Portuguese value set for the EQ-5D-5L on the basis of a hybrid econometric model using cTTO and DCE data. These results represent the preferences of the Portuguese population and are recommended to inform health decision-making in Portugal.

Keywords EQ-5D-5L · Preferences · Portuguese value set · Composite time trade-off · Discrete choice experiment · Quality of life

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11136-019-02226-5) contains supplementary material, which is available to authorized users.

✉ Lara N. Ferreira
lnferrei@ualg.pt

[1] Centre for Health Studies and Research of the University of Coimbra (CEISUC), Coimbra, Portugal

[2] Faculty of Economics, University of Coimbra, Coimbra, Portugal

[3] University of the Algarve, ESGHT, Faro, Portugal

[4] Research Centre for Spatial and Organizational Dynamics (CIEO), University of the Algarve, Faro, Portugal

[5] Axentiva Solutions S.L., Tenerife, Spain

Introduction

Within the context of scarce resources and limitless possible uses the decision-making process implemented has relevant consequences to the health of individuals [1, 2]. Preference-based measures are increasingly used in health economic evaluations and provide a valid solution to achieve the maximization of health and the reduction of inequalities. These approaches incorporate utilities for health outcomes that are used in cost-utility analyses (CUAs) to aid resource allocation decisions. CUAs use the quality-adjusted life year (QALY), a measure which combines information on longevity and quality of life. The QALY assumes that individuals, during their life span, move through different health states, and that each health state has a value assigned [3]. Many organizations responsible for reviewing or producing health technology assessments advise or require the use of preference-based measures in QALY calculations in order to present CUAs results using a cost-per-QALY framework [4, 5]. For many of the regulatory agencies, the EQ-5D is the preferred preference-based measure to estimate utilities used in QALY calculations [4].

The EQ-5D was developed by the EuroQol Group for use in economic evaluation in 1988 [6]. This measure derives from an original descriptive system with six dimensions (mobility, self-care, usual activities, social relationships, pain/discomfort and anxiety/depression), and was subsequently revised to include only five dimensions (social relationships were excluded) [6]. Being the most popular generic preference-based instrument worldwide, the three-level classification system of the EQ-5D (EQ-5D-3L) has three levels in each dimension—no problems, some problems and extreme problems. This classification system describes 243 (3^5) unique health states.

In 2009, the EuroQol Group introduced a five-level EQ-5D that expands the descriptive system from three to five levels within each dimension and therefore describes 3125 (5^5) unique health states [7]. These five levels of severity range from no problems (level 1) to extreme problems/unable to (level 5). This enables a health state to be identified by a five-digit number containing the severity level of each dimension. For example, a 11111 health state indicates no problems in any of the five dimensions and 55555 means a state with the highest level of problems in each dimension.

The EQ-5D-5L was developed in order to tackle the ceiling effect on the three-level version reported in several studies, mainly with healthy and/or young individuals (e.g. [8–12]). Recent research has shown that the ceiling effect has been substantially reduced with the EQ-5D-5L (e.g. [13]).

For the EQ-5D-5L to be used in health economic evaluation, societal values need to be assigned to the 3125 health states generated by this measure. Since the demographic characteristics and the cultural aspects of each country can influence the valuation of health states, value sets should be derived at a national level [14–16]. A standardized protocol was developed by the EuroQol Group to ensure a common methodology for use in all EQ-5D-5L valuation studies. The EuroQol Valuation Technology (EQ-VT) protocol, launched in 2012, includes two valuation techniques: the composite time trade-off (cTTO) complemented by a discrete choice experiment (DCE) [17, 18]. The cTTO involves the use of the traditional TTO approach for states better than death and lead-time TTO for states worse than death in a single task [19]. However, since 2012 the EQ-VT has been subjected to improvements meaning that distinct national valuation studies benefited from better methodologies for data collection and control. The improvements were mainly due to major data issues associated to version 1.0 of the EQ-VT. Therefore, version 1.1 benefited from a quality control procedure that introduced training tasks to obtain better elicitations and the assessment of extreme inconsistencies in valuations related with the 55555-health state. On the other hand, version 2.0 added a feedback module to allow the respondents to flag inconsistent answers. Further information on the existing tariffs may be consulted on the EQ-5D website.

In Portugal, the Portuguese value set for the EQ-5D-3L version based on the preferences of a representative sample of the Portuguese general population, as well as the Portuguese norms, were estimated a few years ago [20, 21]. However, a value set for the EQ-5D-5L for Portugal does not exist yet. Therefore, the purposes of this study were to elicit the EQ-5D-5L health state preferences from the general Portuguese population and to derive the Portuguese value set for the EQ-5D-5L.

Methods

Sample size and data collection

The recruitment of interviewers took place between September 2015 and June 2016. Prior to data collection, interviewers were trained by the Portuguese research team (PLF, PA, LNF, LNP, co-authors of this paper) in 4-h sessions that included a general presentation of the study and its relevance in the context of a decision-making process. Next, the EQ-5D-5L descriptive system and visual analogue scale (VAS) were presented, with the session ending with a full simulation of an interview and a discussion on how to overcome possible issues. As part of the training and testing, a pilot test comprising five test interviews was conducted by each interviewer prior to the real data collection. Only after being

approved through this pilot test could each interviewer start the fieldwork.

Data were collected at respondents' residences between October 2015 and July 2016 through computer-assisted personal interviews using the EQ-VT protocol, version 2.0 [17, 18, 22]. The EQ-VT was developed to administer the protocol and to collect the data. The advantage of using a digital aid is that it forces interviewers to follow exactly the same procedures, thereby reducing the probability of interviewer bias and eliminating data entry and coding errors [17]. This software was specifically designed by the EuroQol Group and previously translated, validated and adapted to Portuguese by the Portuguese research team following the same guidelines defined by the EuroQol Group that are currently used for the translation of the EQ-5D. The interviewer training materials were also standardized and officially translated into Portuguese. The Portuguese EQ-5D-5L descriptive system applied in Ferreira et al. [13] was used.

The random sample stratified by gender and age group was originally designed based on the random route sampling method. The following steps were taken: (1) random selection of a parish; (2) random selection of a starting point (based on maps); (3) from each starting point, definition of a quasi-random path, based on a predefined schema (left or right in respect of the pavement; rules for intersections and blind alleys); (4) selection of buildings (first, third, fifth building, etc.) and (5) selection of the person to interview (based on the next household person's birthday).

The target population was the Portuguese general population aged 18 or over. The population was stratified according to two gender and four age group categories (18–29; 30–49; 50–69; ≥ 70). The age-gender representativeness followed the last Portuguese census [23]. This study granted approval by the (Portuguese) National Data Protection Commission (ref. 1737/2015) to elicit the preferences of the Portuguese general population, preserving the anonymity and confidentiality of the participants. Moreover, before each interview, each respondent was asked if s/he agreed to participate in the study.

Since each participant should value 10 health states in the cTTO task, a sample of 1000 respondents generates 10,000 valuations, which is the minimum required from a power calculation to obtain a standard error of 0.01 of the observed mean of the cTTO [24].

Study design

The design for the cTTO task was created using Fedorov's exchange algorithm. Eighty-six health states were valued using cTTO and assigned to 10 blocks of 10 cTTO tasks: (1) 80 health states were randomly selected from a total of 3125 health states described by the EQ-5D-5L, (2) plus five fixed mild health states (21111, 12111, 11211, 11121 and 11112) and (3) the worst health state (55555) in the EQ-5D-5L descriptive system [17]. From a baseline of 1000 participants, around 100 observations per state were expected for the 80 states, 200 for the very mild states and 1000 for the 55555 health state. The selection of the health states is thoroughly described elsewhere [17].

The design for the DCE task included 196 pairs of health states generated by the experimental design created by Oppe et al. [17] and randomly assigned to 28 blocks of seven pairs. The blocks were balanced in terms of severity of problems in each dimension. None of the health states included in each choice pair was dominant over the other. The pairs were randomly presented within blocks. The levels of randomization included the following criteria: (i) respondent to one block; (ii) pairs within each block and (iii) left/right presentation within a pair. Again, from a baseline of 1000 participants, around 36 observations per pair were expected. Each respondent completed 10 cTTO and seven DCE tasks.

Interview overview

The valuation interview consisted of five different sections:

(1) general welcome of the participant;
(2) introduction to the purpose of the research, followed by the self-completion of the EQ-5D-5L descriptive system and rating of the overall health on the day of the interview on a vertical VAS. Background questions regarding age, gender and previous experience with illness were also asked;
(3) set of cTTO tasks. The respondents were first introduced to the task using a wheelchair example and a valuation of three practice states, respectively a light, a moderate and a severe one. They would then have to carry out the 10 cTTO tasks followed by a feedback module where they would have the option to reconsider their previous answers and three debriefing questions;
(4) set of DCE tasks where participants were asked to complete seven paired comparisons and to answer three debriefing questions regarding the task;
(5) general thank you for the cooperation and goodbye.

In this study, data regarding some local specific questions were also collected. These questions focused mainly on sociodemographic data and which EQ-5D dimensions were more relevant to the participants on each valuation task. These results will not be presented here.

Quality control of the data

During the data collection period, the data were reviewed using the quality control (QC) procedure developed by the EuroQol Group [25].

The QC process initially evaluates interviewer protocol compliance through four key parameters of the cTTO task. If any of the four following criteria was met for an individual interview, the interview was flagged as being of poor quality: (1) no explanation of the worse than death task in the wheelchair example; (2) less than three minutes spent on the wheelchair example; (3) a clear inconsistency in the TTO ratings (that is, 55555 is not the lowest and is at least 0.5 higher than the state with the lowest value) and (4) less than five minutes for the completion of all ten cTTO tasks. DCE data were monitored to detect suspicious response patterns (AAAAAAA, BBBBBBB, ABABABA and BABABAB).

Data were later also evaluated to check the characteristics of the collected data against known issues with other EQ-VT studies, including: few worse than death responses, clustering of responses at certain values (spikes), inconsistencies involving the worst state (pits) represented by 55555 and low values for mild states. This was done every week by the national team and biweekly by the EQ-VT support team. Based on the results, the national team periodically decided whether to retrain or drop interviewers from the team.

In addition, 10% of the sample was randomly cross-questioned by telephone after the interview in order to detect possible fraud. This second interview covered questions regarding the date of the interview, age and gender of the respondent, self-assessment of health status according to the EQ-5D-5L descriptive system and VAS.

Statistical analyses

Descriptive statistics were used to summarize participant characteristics as well as cTTO and DCE valuations, using the Statistical Package for the Social Sciences (SPSS) version 23.

Exclusion criteria

Prior to modelling, exclusion criteria were defined with regard to cTTO responses. These focused on:

(1) participants who gave the same value to all health states (e.g. individuals who did not trade time and were therefore designated as non-traders);
(2) responses flagged by respondents in the feedback module (the feedback module presented participants with the rank ordering of the health states based on their cTTO valuations. When the respondent considered that the health state valuation was not properly ranked, the response was flagged);
(3) participants with a positive slope on the regression between their values and the misery index of the health states indicating that the participant provided higher utility values for poorer health states on average;
(4) inconsistencies related to the "pits" state in the cTTO ratings (the value for 55555 is not the lowest).

No exclusion criteria were applied to DCE data.

Data modelling

The collected data allow different ways for estimating a value set. However, to date, there are not known statistical measures to inform the comparison of models when applying hybrid econometric models versus cTTO or DCE models. The AIC and BIC information criteria cannot be applied as the log-likelihood of a hybrid model is by definition higher than both the log-likelihood of cTTO and the log-likelihood of DCE. The mean square error or the mean absolute error cannot be applied either in a hybrid model as there is no evidence that supports its use [24].

In this case, and following most of recent EQ-5D-5L published value sets, we have decided to use all available information (cTTO and DCE) to be modelled using a hybrid approach, as thoroughly described by Ramos-Goñi et al. [24]. This model can be estimated in Stata using the *hyreg* command and detailed mathematical description of the model can be found elsewhere [26]. The main assumption of the hybrid model is that $\beta' = \beta/\theta$, being β' the coefficients from the DCE side of the model and β the respective from cTTO, in other words, coefficients from cTTO are proportional to those from DCE. In order to test whether this is true in our data, several econometric models were tested to estimate cTTO and DCE data separately to be able to plot cTTO versus DCE. The selected DCE model follows McFaden's approach of using a conditional logit model [27]. For cTTO data a number of decisions were made on basis of data properties: cTTO as used in the EQ-VT protocol censor the observed values at − 1 (as the ratio used in the Lead-Time side was 1:1 [28]), therefore we have used a Tobit model censored at − 1. In addition, cTTO data are clearly heteroskedastic as people tend to better agree when valuing mild health states while it is not the case for severe health states, therefore we have used a heteroskedastic Tobit model. The selected hybrid model combines both sources of data assuming all these theoretical considerations [22].

In order to show that the hybrid model make sense in our case we have plotted DCE versus cTTO using scatter plots and we have plotted the line Y = X. If DCE versus cTTO shows a line it means that both are proportional and hybrid model can be safely estimated. The angle between the Y = X and the line of dots represent the θ.

Sensitivity analysis

Finally, a sensitivity analysis was conducted to assess how the exclusion of the inconsistent responses flagged by

respondents in the feedback module impacted on the modelling of cTTO and hybrid results. In this study, the sensitivity analysis was carried out to assess the robustness of the estimated parameters, when all the valuation responses were included, i.e. when no exclusion criteria were applied.

Comparison of results

The best model was compared with corresponding models of other countries, namely Spain [29] and Ireland [30].

Results

Supervision of data collection

Fifty-four interviewers were trained in the course of this study. Some of these interviewers abandoned the study by their own initiative due to the complexity of the task. Others were invited by the national team to abandon the study due to their poor performance. Training was performed in groups or individually, depending on the number of individuals recruited in each time. Of the 54 trained interviewers, 15 individuals gave up shortly after the training session, meaning that 39 potential interviewers started and concluded the pilot test laid down in the code of conduct. This formative process led to the realization of 253 pilot interviews, with an average number of 6.5 interviews per interviewer.

Of the 39 interviewers who concluded the pilot test, 30 started the fieldwork but only 28 fulfilled the quality criteria with a maximum of 9% of interviews identified as being of poor quality according to the four key parameters of the cTTO task defined earlier. The two remaining interviewers were asked to abandon the study by the national team because of their poor performance. In general, the interviewers have complied with the quality control protocol, providing high-quality valuation data, a reduced number of inconsistencies and no peaks around specific values.

As the quality control process was not within the scope of this study, the results obtained are not described here.

In what concerns the 10% of the sample questioned by telephone, we confirmed that all the interviews have been performed.

Sample

A total of 1451 respondents, representative in terms of age and gender of the general population of mainland Portugal and the islands, participated in this study. The interviews were conducted all over the country. Figure 1 highlights in bold the municipalities where individuals resided. The municipalities with the largest number of participants were Coimbra, with 402 respondents (27.7%), Évora, with 228 participants (15.7%) and Lisbon, with 206 individuals (14.2%).

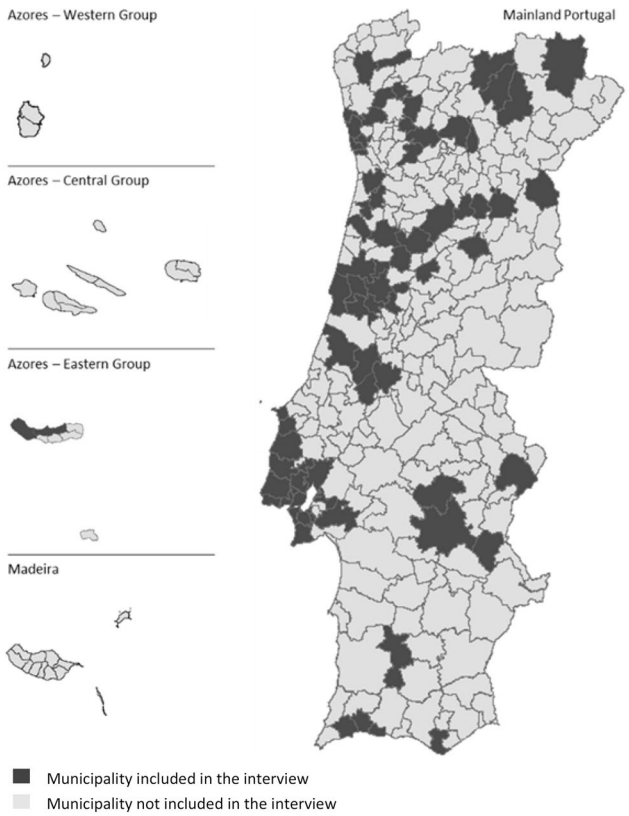

Fig. 1 Sample distribution by municipality of residence. Filled square Municipality included in the interview, open square Municipality not included in the interview

The sociodemographic characteristics of the sample, along with the Portuguese general population are presented in Table 1.

Table 1 shows that about 62% of the respondents had at least 12 years of education and 3.2% did not have any formal education. Moreover, more than half of the respondents were married, 50% were employed and about 30% had a family monthly household income of between 1000€ and 2000€, followed by 22% with an income of between 500€ and 1000€.

Table 2 presents information on the experience with disease and health-related quality of life as measured by the EQ-5D-5L. Almost a third—478 (32.9%)—of the sample stated that they had a disease, which they had been informed about by a health professional. Amongst the most referred to diseases are hypertension, diabetes and heart disease. More than 75% did not have a serious disease but had contact with a relative with a serious disease and most of them had experience as a carer. When asked about their self-assessment of their own health, the majority referred to being in good (5.1%) or fair (34.0%) health.

Table 1 Sample sociodemographic characteristics

Variable	Category	Estimation sample		Portuguese general population aged 18+[a]
		n	%	%
Gender	Female	825	56.9	53
	Male	626	43.1	47
Age	[18–29] years	239	16.5	17
	[30–39] years	236	16.3	18.5
	[40–49] years	200	13.8	17.8
	[50–59] years	233	16.0	16.2
	[60–69] years	155	10.7	13.7
	≥70 years	388	26.7	16.8
Education	No formal education	47	3.2	6.2
	4 years	239	16.5	31.2
	6 years	86	6.0	9.7
	9 years	176	12.1	14.9
	12 years	322	22.2	19.2
	Higher education	581	40.0	18.8
Family status	Married or living in common law	732	50.5	56.9
	Single	409	28.2	27.4
	Widower	173	11.9	8.9
	Divorced or separated	137	9.4	6.8
Labour status	Employed/self-employed	726	50.0	49.4
	Unemployed but able to work	70	4.8	7.4
	Unemployed but unable to work for health reasons	3	0.2	1.6
	Student	146	10.1	5.8
	Housekeeper	43	3.0	4.7
	Retired	463	31.9	31.1
Family net monthly income	< 500€	116	11.4	n.a.
	[500€–1000€]	224	22.0	n.a.
	[1000€–2000€]	303	29.7	n.a.
	[2000€–3000€]	182	17.8	n.a.
	≥ 3000€	107	10.5	n.a.
	Family dependent	88	8.6	n.a.

n.a. not available

[a]*Source* Census 2011 [23]

This table also shows the distribution of each EQ-5D-5L dimension. The dimensions with a high percentage of individuals referring to not having problems were 'self-care', 'usual activities' and 'mobility'. It is also interesting to mention that these percentages decrease as age increases (not shown in this table), especially for mobility. Few severe health problems were reported (0.3%) in the 'usual activities' and 'anxiety/depression' dimensions. As regards the self-reported EQ-5D-VAS, a mean score of 78.9 ± 14.9 was obtained.

cTTO data

In this study, each of the 1451 participants contributed with a valuation of 10 health states in the cTTO task, resulting in a total of 14,510 valuations. Each cTTO valuation was completed, on average, in 78.6 s (1.3 min). The 80 randomly selected health states were assessed on average 145.1 times, the very mild states were assessed on average 290.2 times and finally all 1451 respondents assessed the 'pits' state (55555). Supplementary Table 1 presents the descriptive

Table 2 Health status/experience with disease

	n	%
Disease, informed by a health professional		
Yes	478	32.9
No	904	62.3
Does not know/does not answer	69	4.8
Self-assessment of health		
Very bad	2	0.1
Bad	29	2.0
Fair	494	34.0
Good	727	50.1
Very good	199	13.7
Incidence of severe disease		
Yes	329	22.7
No	1122	77.3
Experience with relative with severe disease		
Yes	1063	77.3
No	388	26.7
Experience as carer of patient with a severe disease		
Yes	496	34.2
No	955	65.8
Mobility		
No problems in walking about	1079	74.4
Slight problems in walking about	210	14.5
Moderate problems in walking about	127	8.8
Severe problems in walking about	33	2.3
Unable to walk about	2	0.1
Self-care		
No problems washing or dressing myself	1269	87.5
Slight problems washing or dressing myself	129	8.9
Moderate problems washing or dressing myself	42	2.9
Severe problems washing or dressing myself	9	0.6
Unable to wash or dress myself	2	0.1
Usual activities		
No problems doing my usual activities	1095	75.5
Slight problems doing my usual activities	205	14.1
Moderate problems doing my usual activities	122	8.4
Severe problems doing my usual activities	24	1.7
Unable to do my usual activities	5	0.3
Pain/discomfort		
No pain or discomfort	780	53.8
Slight pain or discomfort	446	30.7
Moderate pain or discomfort	182	12.5
Severe pain or discomfort	41	2.8
Extreme pain or discomfort	2	0.1
Anxiety/depression		
Not anxious or depressed	978	67.4
Slightly anxious or depressed	346	23.8
Moderately anxious or depressed	103	7.1
Severely anxious or depressed	19	1.3
Extremely anxious or depressed	5	0.3
EQ-VAS		
Mean	78.9	
Standard deviation	14.9	
Median	80	
Percentile 25%	70	
Percentile 75%	90	

statistics for the values obtained in the 86 health states assessed by the cTTO task in this study. Observed values ranged from − 0.45 for the 'pits' state (55555) to 0.95 for the 11112 state.

Prior to cTTO modelling, health states flagged by the respondents in the feedback module were excluded. Thus, the 14,510 valuations of health states obtained by the cTTO task were reduced to 13,391, excluding 1119 valuations considered by respondents to be incorrectly ordered during the course of the interview. In addition, the participant ($n = 1$) with a positive slope on the regression between his values and the misery index of the health states was also excluded. No inconsistencies were observed when modelling cTTO data alone.

DCE data

In this valuation task, each of the 1451 participants valued seven choice pairs, which resulted in a total of 10,157 valuations of health states. The participants spent 54 s answering on average when faced with each choice pair. No exclusion criteria were applied to DCE data. No inconsistencies were observed when modelling DCE data alone.

Modelling cTTO and DCE combined data: the hybrid model

Figure 2 shows the correlation between cTTO, DCE and the hybrid models. The three graphs presented correspond respectively to the correlation between the cTTO and hybrid scores, the DCE and cTTO scores and, finally, between DCE and hybrid scores. The idea of this figure is to show that it is safe to merge TTO and DCE data through a combination of a tobit model with a logit model.

As figure shows a strong correlation, it seems appropriate to combine both sources of information (cTTO and DCE) in a hybrid model.

Given the nature of these valuation data and as shown in many other valuation studies [31–33], it is expected that heterogeneity occurs associated with the valuation of health states. Using a severity index proxy obtained from adding all five levels of a health state, Supplementary Fig. 1 shows that, as the index of severity of problems increases, the mean

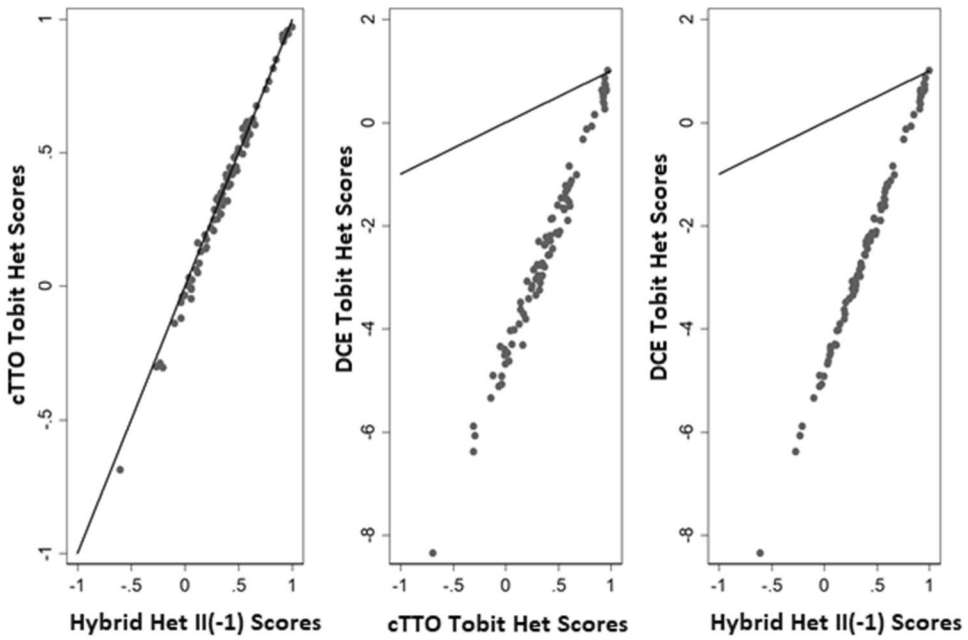

Fig. 2 Comparison of estimated values for the 86 health states included in the cTTO design using the heteroskedastic Tobit model censored at − 1, the logit model and the heteroskedastic hybrid model

Table 3 Heteroskedastic censored hybrid model

Dimension levels	β Incremental dummies	*P* values Incremental dummies	β Regular dummies	SE Regular dummies	[95% confidence interval] Regular dummies
mo2	0.048	0.000	0.048	0.005	[0.0374–0.0577]
mo3	0.044	0.000	0.092	0.008	[0.0767–0.1065]
mo4	0.090	0.000	0.182	0.008	[0.1671–0.1968]
mo5	0.174	0.000	0.356	0.008	[0.3407–0.3710]
sc2	0.048	0.000	0.048	0.005	[0.0381–0.0582]
sc3	0.022	0.003	0.070	0.007	[0.0563–0.0848]
sc4	0.086	0.000	0.156	0.007	[0.1421–0.1714]
sc5	0.138	0.000	0.294	0.007	[0.2810–0.3086]
ua2	0.044	0.000	0.044	0.005	[0.0345–0.0540]
ua3	0.019	0.013	0.063	0.007	[0.0488–0.0769]
ua4	0.072	0.000	0.135	0.007	[0.1213–0.1488]
ua5	0.128	0.000	0.263	0.007	[0.2498–0.2773]
pd2	0.041	0.000	0.041	0.004	[0.0321–0.0495]
pd3	0.060	0.000	0.101	0.008	[0.0857–0.1152]
pd4	0.153	0.000	0.254	0.008	[0.2387–0.2686]
pd5	0.152	0.000	0.406	0.008	[0.3896–0.4219]
ad2	0.036	0.000	0.036	0.005	[0.0274–0.0456]
ad3	0.049	0.000	0.085	0.007	[0.0718–0.0998]
ad4	0.127	0.000	0.212	0.007	[0.1989–0.2268]
ad5	0.072	0.000	0.284	0.007	[0.2703–0.2992]

mo mobility, *sc* self-care, *ua* usual activities, *pd* pain/discomfort, *ad* anxiety/depression

cTTO value decreases, achieving mean negative values from scores greater than 20. On the other hand, the standard deviation increases, particularly from scores greater than 10.

When valuing health states, one option to tackle heterogeneity is using heteroskedastic econometric models. The heteroskedastic censored hybrid model estimated consistently and statistically significant parameters (Table 3).

Based on this model it is possible to assign a utility score to a health state derived by the EQ-5D-5L descriptive system. For instance, the health state 14235 has the following

utility: U = 1 − 0.000 − 0.156 − 0.044 − 0.101 − 0.284 = 0.415. The syntax for the Portuguese EQ-5D-5L value set is provided in the Supplementary Material.

Sensitivity analysis

In this study, the sensitivity analysis focused on the impact of the exclusion of data related to the flagged responses in the feedback module. It consisted in assessing the reliability of this decision considering all valuation data ($n = 14{,}510$) without applying any exclusion. Therefore, we estimated the model for cTTO and no inconsistencies occurred, except for the mo2 coefficient, which presented a negative value (− 0.002) in the censored Tobit model. Nonetheless, a greater number of parameters had no statistical significance when compared with the results obtained from the models associated with the exclusion of data.

Comparison of results

Figure 3 presents a comparison between the Portuguese value set for the EQ-5D-5L and those obtained from Spain [29] and Ireland [30], using a hybrid econometric model that combined continuous and dichotomous data from cTTO and DCE tasks.

Notwithstanding the difference between the value sets, it is interesting to note the similarity between the Portuguese and the Spanish value sets obtained when using the hybrid model. In this case, the Pearson correlation coefficient is very close to 1 ($r = 0.962$). Although the Pearson correlation coefficient with the values from Ireland is 0.919, in general the Irish valued the health states lower.

Discussion

The estimation of QALYs in economic evaluations is usually performed using a value set of a preference-based measure. Previous research suggests that health status assessments may differ from country to country due to differences in demographics, socio-cultural values and other characteristics (e.g. [34, 35]). Although, in the absence of country-specific value sets, another value set derived from a culturally similar country can be used, several countries have derived their own value sets. Therefore, the main objective of this study was to derive the value set for the EQ-5D-5L for the Portuguese general population. This value set can now be used in health economic evaluation studies conducted in Portugal and/or submitted to the Portuguese regulatory agency (INFARMED) for reimbursement purposes when the EQ-5D-5L has been used as a measure of benefit.

The results of this study also suggest that pain/discomfort and mobility are the most relevant EQ-5D dimensions according to the preferences of the general Portuguese population. Usual activities and anxiety/depression are the least relevant dimensions.

This study was part of the third wave of EQ-5D-5L valuation studies (version 2.0 of the QC protocol), which included some improvements aimed at increasing data quality [22, 36]. These sets of developments included the training tasks prior to the cTTO valuation task, pop-up windows to confirm the participants' responses in the valuation tasks and the feedback module. This last feature allowed the flagged responses to be excluded from the data set prior to modelling. Some authors (e.g. [37, 38]) suggest the use of other criteria to exclude respondents from the analysis, such as non-traders, participants with a limited number of iterations for all health states and respondents who stated that they did not understand the task or that rated mild health states lower than severe health states. These aspects can be minimized by ensuring a thorough training of interviewers and a comprehensive explanation of the tasks to participants. However, a participant who reveals one of these features is possibly not mistaken and may just be highlighting some personal characteristics or the effect of the family and/or social context. Distinct exclusion criteria have been applied in several valuation studies, with no golden rule or methodological standard agreed [39]. Therefore, the application of any exclusion criteria in valuation studies must be well grounded and its consequences evaluated through sensitivity analysis.

Since the feedback module is a recent development of the EQ-VT protocol, introduced in 2015, few valuation studies have used this component. As far as the authors are aware, only one valuation study has been published featuring this exclusion criterion when estimating an EQ-5D-5L value set [33].

In this study, different econometric models were estimated aimed at estimating the Portuguese value set for the EQ-5D-5L, according to the valuation tasks and the type of data obtained (continuous or dichotomous). cTTO data allowed a value set to be obtained from the econometric model with the best performance—the GLS interval regression censored at 1. In this study, DCE data could not be used independently to estimate a tariff because the generated values were not anchored onto the full health-dead 1-0 QALY scale required for use in economic evaluation. This was due to the absence of a method to convert modelled DCE latent values onto the full health-dead scale. Some authors [40–42] suggest anchoring DCE values using TTO values for the best and worst state or including a duration attribute (DCE_{TTO}) in the DCE to generate health utility values. However, this will not prevent the use of dichotomous data in hybrid econometric models obtained from DCE tasks alongside continuous data from cTTO tasks. The combined use of these techniques enables large-scale data collection using a DCE to be undertaken

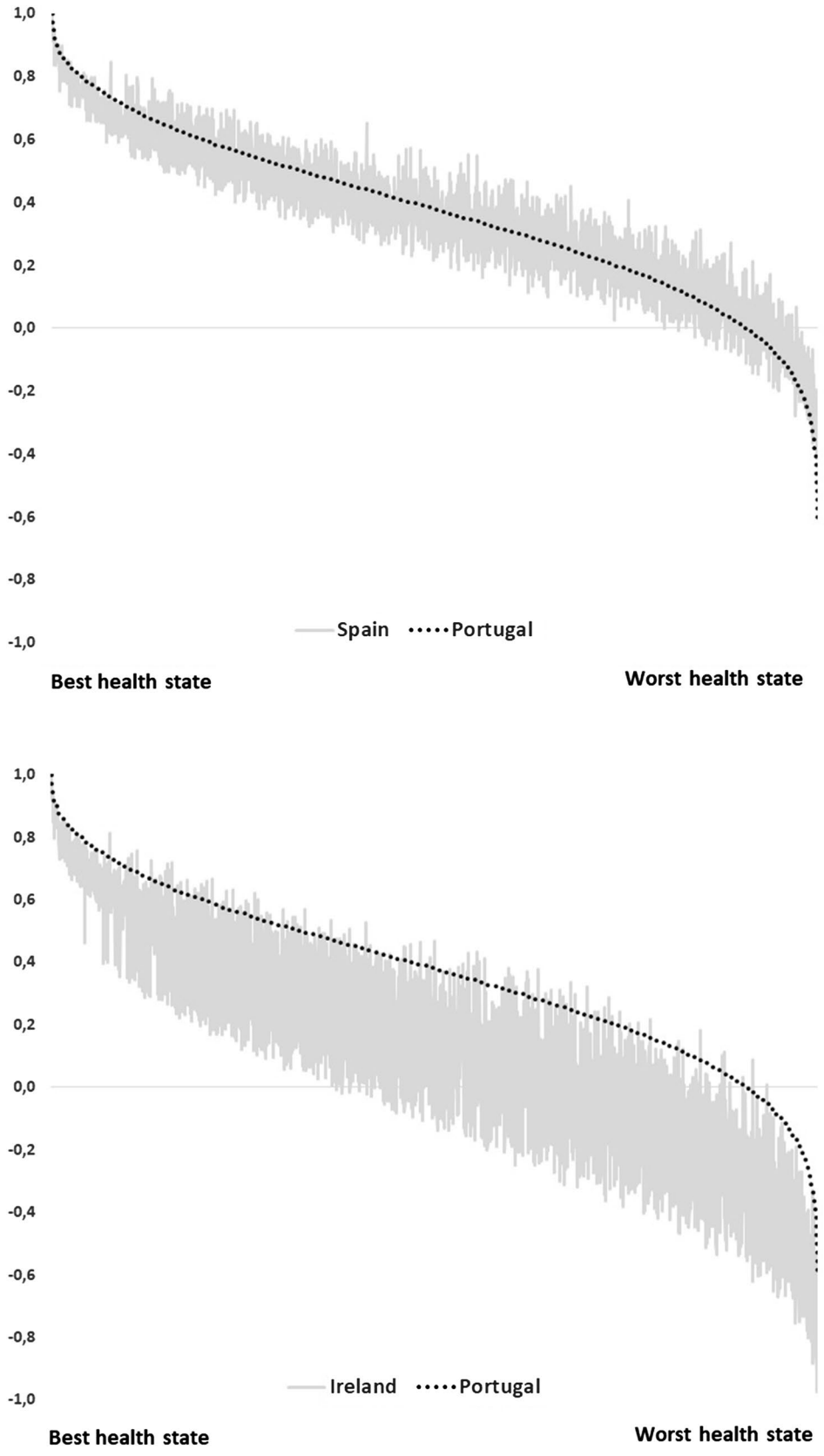

Fig. 3 Comparison between the EQ-5D-5L value sets for Portugal, Spain and Ireland (hybrid models)

inexpensively online with small-scale cTTO data collected through face-to-face interviews [41]. Therefore, we have followed the methodological work of Ramos-Goñi and et al. [24] and have estimated a consistent and well-behaving hybrid model. This methodological approach combines these techniques and enables us to circumvent the disadvantages associated with the cTTO task—mainly time consumption and the high costs associated with the task. On the other hand, previous research [41, 43] demonstrated that there is a need to further explore alternative preference elicitation methods for decision-making, since there are methodological challenges in measuring preferences in particular situations, such as with children, the elderly and specific diseases.

Despite being distinct information, it is considered that both cTTO and DCE data incorporate complementary evidence that can better represent the preferences of the sample. Hence, the hybrid model was chosen to estimate the value set. This decision was reinforced by the ability of this model to estimate consistent and statistically significant coefficients.

There are known issues associated with the cTTO task, namely scale compatibility and loss aversion [44]. DCE also face some issues related to lexicographic responses [45]. Adding up data from these two valuation techniques will neither compensate nor enhance the biases revealed by each one of the tasks. Given that no statistical measure will inform this hybrid econometric model, the decision on its use will be normative.

There are other factors that play a significant role in the quality of data obtained in valuation studies. In this respect, computer-assisted personal interviews, like the ones used in this study, constitute a major advantage over postal or telephone interviews. Face-to-face interviews with trained interviewers enlighten and prepare the participants for the tasks ahead. This method has also been used in valuation studies conducted in other countries (e.g. [46–48]) and in Portugal as well [20], aimed at estimating value sets for the EQ-5D-3L or the SF-6D (e.g. [49–53]). However, this type of interview can be a significant limitation in studies with a limited budget. This constraint can, however, be circumvented if the valuation techniques include a smaller number of more complex tasks such as cTTO and they are replaced with more straightforward and less expensive tasks such as the DCE, which may decrease the time spent in the interviews and therefore reduce the costs of data collection.

On the other hand, the use of computer-assisted personal interviews normalizes the entire interview process, ensuring that every interview is performed in the same way. The constant 24/7 supervision of the interviewer with access to actual responses is also a plus that should be included in similar studies.

Conclusion

This study provided the Portuguese value set for the EQ-5D-5L on the basis of a hybrid econometric model using cTTO and DCE data. These results represent the preferences of the Portuguese population and are recommended for informing decision-making in CUAs to be conducted in Portugal, aimed at supporting decision-making on different alternatives in a context of scarce resources.

Acknowledgements This research was part of an EuroQoL international project aimed at obtaining country-specific EQ-5D-5L value sets. The authors are grateful to the EuroQol Group (Elly Stolk and Bernhard Slaap) for the opportunity to implement this study in Portugal and to the technical team (Kristina Ludwig and Arnd Jan Prause) for their continuous advice and support. Views expressed in the paper are those of the authors alone. The authors are thankful for support from the Centre for Health Studies & Research-University of Coimbra (CEISUC) [FCT-Foundation for Science and Technology Grant Number UID/MULTI/4066/2019] and from the Research Centre for Spatial and Organizational Dynamics-University of the Algarve (CIEO) [FCT Grant Number UID/SOC/04020/2019].

Funding This study was funded by a grant from the EuroQol Research Foundation.

Compliance with ethical standards

Conflict of interest One author discloses that he is a member of the EuroQol Group, an organization focused on the development of instruments that describe and value health outside the for-profit model. The views expressed by the authors in the publication do not necessarily reflect the views of the EuroQol Group. The authors declare that they have no conflicts of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. This study was reviewed and approved by the (Portuguese) National Data Protection Commission (Ref. 1737/2015) to elicit the preferences of the Portuguese general population, preserving the anonymity and confidentiality of the participants.

Informed consent Informed consent was obtained from all individual participants included in the study. Participants were informed about their freedom for refusal.

References

1. Baltussen, R., & Niessen, L. (2006). Priority setting of health interventions: The need for multicriteria decision analysis. *Cost Effectiveness and Resource Allocation, 4,* 14.
2. Devlin, N., & Sussex, J. (2011). *Incorporating multiple criteria in HTA: Methods and processes*. London: Office of Health Economics.
3. Brazier, J., Ratcliffe, J., Salomon, J., & Tsuchiya, A. (2007). *Measuring and valuing health benefits for economic evaluation*. Oxford: Oxford University Press.

4. Longworth, L., & Rowen, D. (2013). Mapping to obtain EQ-5D utility values for use in NICE health technologies assessment. *Value in Health, 16,* 202–210.
5. Ferreira, L., & Ferreira, P. (2014). Health state values and country-specific value sets. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 2744–2749). Dordrecht: Springer.
6. EuroQol Group. (1990). EuroQol—A new facility for the measurement of health-related quality of life. *Health Policy, 16*(3), 199–208.
7. Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., et al. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research, 20*(10), 1727–1736.
8. Bharmal, M., & Thomas, J. (2006). Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value in Health, 9*(4), 262–271.
9. Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics, 13*(9), 873–884.
10. Cunillera, O., Tresserras, R., Rajmil, L., Vilagut, G., Brugulat, P., Herdman, M., et al. (2010). Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey. *Quality of Life Research, 19*(6), 853–864.
11. Ferreira, L., Ferreira, P., & Pereira, L. (2014). Comparing the performance of the SF-6D and the EQ-5D in different patient groups. *Acta Médica Portuguesa, 27*(2), 236–245.
12. Petrou, S., & Hockley, C. (2005). An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Economics, 14*(11), 1169–1189.
13. Ferreira, L. N., Ferreira, P. L., Ribeiro, F. P., & Pereira, L. N. (2016). Comparing the performance of the EQ-5D-3L and the EQ-5D-5L in young Portuguese adults. *Health and Quality of Life Outcomes, 14*(1), 89.
14. Tsuchiya, A., Ikeda, S., Ikegami, N., Nishimura, S., Sakai, I., Fukuda, T., et al. (2002). Estimating an EQ-5D population value set: The case of Japan. *Health Economics, 11*(4), 341–353.
15. Johnson, J., Luo, N., Shaw, J., Kind, P., & Coons, S. (2005). Valuations of the EQ-5D health states: Are the United States and United Kingdom different? *Medical Care, 43*(3), 221–228.
16. Yusof, F., Goh, A., & Azmi, S. (2012). Estimating an EQ-5D value set for Malaysia using time trade-off and visual analogue scale methods. *Value in Health, 15,* S85–S90.
17. Oppe, M., Devlin, N., van Hout, B., Krabbe, P., & de Charro, F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health, 17*(4), 445–453.
18. Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goñi, J., & Luo, N. (2016). EuroQol protocols for time trade-off valuation of health outcomes. *PharmacoEconomics, 34*(10), 993–1004.
19. Janssen, B., Oppe, M., Versteegh, M., & Stolk, E. (2013). Introducing the composite time trade-off: A test of feasibility and face validity. *The European Journal of Health Economics, 14,* 5–13.
20. Ferreira, L. N., Ferreira, P. L., Pereira, L. N., & Oppe, M. (2014). The valuation of the EQ-5D in Portugal. *Quality of Life Research, 23*(2), 413–423.
21. Ferreira, L. N., Ferreira, P. L., Pereira, L. N., & Oppe, M. (2014). EQ-5D Portuguese population norms. *Quality of Life Research, 23*(2), 425–430.
22. Stolk, E., Ludwig, K., Rand, K., van Hout, B., & Ramos-Goñi, J. (2019). Overview, update and lessons learned from the international EQ-5D-5L valuation work: Version 2 of the EQ-5D-5L valuation protocol. *Value in Health, 22*(1), 23–30.
23. Portuguese Statistical Office. (2012). *Census 2011*. Lisbon: Portuguese Statistical Office.
24. Ramos-Goñi, J., Pinto-Prades, J., Oppe, M., Cabasks, J., Serrano-Aguilar, P., & Rivero-Arias, O. (2017). Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Medical Care, 55*(7), e51–e58.
25. Ramos-Goñi, J., Oppe, M., Slaap, B., Busschbach, J., & Stolk, E. (2017). Quality control process for EQ-5D-5L valuation studies. *Value in Health, 20*(3), 466–473.
26. Ramos-Goñi, J. M., Craig, B., Oppe, M., & van Hout, B. (2016). *Combining continuous and dichotomous responses in a hybrid model*. Rotterdam: EuroQol Research Foundation.
27. McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.
28. Devlin, N., Buckingham, K., Shah, K., Tsuchiya, A., Tilling, C., Wilkinson, G., et al. (2013). A comparison of alternative variants of the lead and lag time TTO. *Health Economics, 22,* 517–532.
29. Ramos-Goñi, J., Craig, B., Oppe, M., Ramallo-Fariña, Y., Pinto-Prades, J., Luo, N., et al. (2018). Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value in Health, 21*(5), 596–604.
30. Hobbins, A., Barry, L., Kelleher, D., Shah, K., Devlin, N., Ramos-Goñi, J., et al. (2018). Utility values for health states in Ireland: A value set for the EQ-5D-5L. *PharmacoEconomics, 36*(11), 1345–1353.
31. Purba, F., Hunfeld, J., Iskandarsyah, A., Fitriana, T., Sadarjoen, S., Ramos-Goñi, J., et al. (2017). The Indonesian EQ-5D-5L value set. *PharmacoEconomics, 35*(11), 1153–1165.
32. Ludwig, K., von der Schulenburg, J., & Greiner, W. (2018). German value set for the EQ-5D-5L. *PharmacoEconomics, 36*(6), 663–674.
33. Wong, E., Ramos-Goñi, J., Cheung, A., Wong, A., & Rivero-Arias, O. (2018). Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. *Patient, 11*(2), 235–247.
34. Huang, I.-C., Willke, R., Atkinson, M., Lenderking, W., Frangakis, C., et al. (2007). US and UK version of the EQ-5D preference weights: Does choice of preference weights make a difference? *Quality of Life Research, 16,* 1065–1072.
35. Ferreira, L. N., Ferreira, P. L., Rowen, D., & Brazier, J. E. (2011). Do Portuguese and UK health state values differ across valuation methods? *Quality of Life Research, 20*(4), 609–619.
36. Yang, Z., Luo, N., Bonsel, G., Busschbach, J., & Stolk, E. (2018). The effect of health state sampling methods on model predictions of EQ-5D-5L values: Small designs can suffice. *Value in Health, 22,* 38–44.
37. Attema, A., Edelaar-Peeters, Y., Versteegh, M., & Stolk, E. (2013). Time trade-off: One methodology, different methods. *The European Journal of Health Economics, 14,* 53–64.
38. Engel, L., Bansback, N., Bryan, S., Doyle-Waters, M., & Whitehurst, D. (2016). Exclusion criteria in national health state valuation studies: A systematic review. *Medical Decision Making, 36*(7), 798–810.
39. Devlin, N., Hansen, P., Kind, P., & Williams, A. (2003). Logical inconsistencies in survey respondents' health state valuations—A methodological challenge for estimating social tariffs. *Health Economics, 12*(7), 529–544.
40. Bansback, N., Hole, A., Mulhern, B., & Tsuchiya, A. (2014). Testing a discrete choice experiment including duration to value health states for large descriptive systems: Addressing design and sampling issues. *Social Science and Medicine, 114,* 38–48.
41. Rowen, D., Brazier, J., & van Hout, B. (2015). A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Medical Decision Making, 35*(3), 328–340.
42. Stolk, E., Oppe, M., Scalone, L., & Krabbe, P. (2010). Discrete choice modeling for the quantification of health states: The case of the EQ-5D. *Value in Health, 13*(8), 1005–1013.

43. Brazier, J., Rowen, D., Yang, Y., & Tsuchiya, A. (2012). Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. *The European Journal of Health Economics, 13*(5), 575–587.
44. van Osch, S., Wakker, P., van den Hout, W., & Stiggelbout, A. (2004). Correcting biases in standard gamble and time trade-off utilities. *Medical Decision Making, 24*(5), 511–517.
45. Lancsar, E., & Louviere, J. (2006). Deleting, "irrational" responses from discrete choice experiments: A case of investigating or imposing preferences? *Health Economics, 15*(8), 797–811.
46. Chevalier, J., & Pouvourville, G. (2013). Valuing EQ-5D using time trade-off in France. *The European Journal Health Economics, 4*(1), 57–66.
47. Greiner, W., Claes, C., Busschbach, J., & Schulenburg, J.-M. (2005). Validating the EQ-5D with time trade-off for the German population. *The European Journal of Health Economics, 6,* 124–130.
48. Shaw, J., Johnson, J., & Coons, S. (2005). US valuation of the EQ-5D health states: Development and testing of the D1 model. *Medical Care, 43*(3), 203–220.
49. Brazier, J., Fukahara, S., Roberts, J., Kharroubi, S., Yamamoto, Y., Ikeda, S., et al. (2009). Estimating a preference-based index from the Japanese SF-36. *Journal of Clinical Epidemiology, 62*(12), 1323–1331.
50. Ferreira, L. N., Ferreira, P. L., Pereira, L. N., Brazier, J., & Rowen, D. (2010). A portuguese value set for the SF-6D. *Value in Health, 13*(5), 624–630.
51. Lam, C., Brazier, J., & McGhee, S. (2008). Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. *Value in Health, 11*(2), 295–303.
52. Cruz, L., Camey, S., Hoffmann, J., Rowen, D., Brazier, J., Fleck, M., et al. (2011). Estimating the SF-6D value set for a population-based sample of Brazilians. *Value in Health, 14*(5), S108–S114.
53. Abellán-Perpiñán, J., Sánchez-Martínez, F., Martínez-Pérez, J., & Méndez, I. (2012). Lowering the "floor" of the SF-6D scoring algorithm using a lottery equivalent method. *Health Economics, 21*(11), 1271–1285.