ScienceDirect
Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Themed Section: Applications of Health Preferences Research

Peruvian Valuation of the EQ-5D-5L: A Direct Comparison of Time Trade-Off and Discrete Choice Experiments

Federico Augustovski, MD, MSc, PhD, María Belizán, Luz Gibbons, Nora Reyes, MD, MSc, Elly Stolk, Benjamin M. Craig, Romina A. Tejada, MD, MSc*

ABSTRACT

Objectives: (1) To produce Peruvian general population EQ-5D-5L value sets on a quality-adjusted life-year scale, (2) to investigate the feasibility of a "Lite" protocol less reliant on the composite time trade-off (cTTO), and (3) to compare cTTO and discrete choice experiment (DCE) value sets.

Methods: A random sample of adults (N = 1000) in Lima, Arequipa, and Iquitos did a home interview; 300 were randomly selected to complete 11 cTTOs first. All respondents completed a DCE, including 10 latent-scale pairs (A/B) with 5 EQ-5D-5L attributes, and 12 matched pairs (A/B and B/C) with 5 EQ-5D-5L and one lifespan attributes. We estimated a cTTO heteroscedastic tobit (N = 300) model and 3 DCE Zermelo-Bradley-Terry models (N = 300, 700, and 1000).

Results: Each model produced a consistent value set (20 positive incremental parameters). Nevertheless, their lowest quality-adjusted life-year values differed greatly (cTTO: −1.076 [N = 300]; DCE: −0.984 [300], 0.048 [700], −0.213 [1000]). Compared with the cTTO, the DCE (N = 300) produced different parameters (Pearson's correlation = 0.541), fewer insignificant parameters (0 vs 8), and fewer values less than 0 (26% vs 44%). Compared with the DCE (N = 300), the DCE (N = 700) produced higher values but similar parameters (Pearson's correlation = 0.800).

Conclusions: Besides producing EQ-5D-5L value sets for Peru, the results casts doubt about the feasibility of a Lite protocol like the one in this study. Additionally, fundamental differences between cTTO and DCE—without the existence of a gold standard—need further clarification. The choice between the two rational value sets produced in the current study is a matter of judgment and may have substantial policy implications.

Keywords: discrete choice experiment, EuroQol, preference elicitation methods, time trade off, utility assessment.

VALUE HEALTH. 2020; 23(7):880–888

Introduction

Health systems around the globe increasingly use health technology assessment (HTA) mechanisms that include economic evaluations to assess the value of health technologies under constantly constrained budgets.[1,2] The World Health Organization promotes the institutionalization of HTA to encourage evidence-informed, social value–based decision making, with the aim of reaching and improving universal health coverage. This is particularly the case with Peru, a middle-income country that is part of REDETSA (a Pan-American Health Organization HTA network) and has three public HTA agencies in the Ministry of Health and in the social security system.[3]

To inform the allocation of scarce healthcare resources, public HTA agencies typically rely on a summary measure when assessing the impact of alternative technologies on patients' length and quality of life[4]: quality-adjusted life years, or QALYs. Compared with measures based on expert opinion or global rating surveys, such as disability-adjusted life years,[4] QALYs[5] are either favored or accepted largely for their reliance on preference evidence from nationally representative populations. Because health preferences vary across populations, local valuation studies are conducted to derive local preference weights for different health outcomes, usually using a generic descriptive system of health-related quality of life.[6] Several systems are available for this, such as the 6-item short-form health survey, the Health Utility Index, or the 15D.[7] The EQ-5D family of instruments is the most widely used preference-based measure worldwide and has 18 country-specific value sets for the EQ-5D-5L.[8]

The EuroQol protocol for EQ-5D-5L valuation studies (EuroQol Valuation Technology [EQ-VT] version 2.1) is resource intensive. In its most prevalent form, the protocol entails eliciting health preferences from a sample of 1000 respondents from the general

* Address correspondence to: Romina Tejada Caminiti, MD, MSc, Unidad de Análisis y Generación de Evidencias en Salud Pública, Centro Nacional de Salud Pública, Instituto Nacional de Salud. Av. Defensores del Morro 2268 (Ex Huaylas) - Chorrillos, Lima 9, Lima, Peru. Email: romina.tejada.caminiti@gmail.com

population with face-to-face interviews. Its composite time trade-off (cTTO) task is also known to be cognitively burdensome because each use involves an adaptive series of choices.[9] Specifically, the cTTO asks respondents to choose iteratively between health problems and reduced lifespan until reaching indifference, and then the cardinal response is converted to a scale from –1 to 1 QALY under the constant proportionality assumption (ie, assuming that each additional year of life has the same value). After the cTTO, respondents may also complete a discrete-choice experiment (DCE). Discrete-choice experiments may also be cognitively burdensome but are easier to implement (ie, no iterations). Unlike cTTO responses, DCE responses are ordinal and cannot be converted to the QALY scale directly. Each EQ-VT interview takes, on average, 45 minutes. Highly trained interviewers and daily monitoring are required to ensure sufficient data quality.[10]

As economic evaluations and HTA extend into lower- and middle-income countries, there is an increasing need for lighter or "Lite" versions that rely on fewer respondents (eg, 400 or fewer persons) and that use methods that are potentially easier to implement. With that background, Stolk et al[11] originally recommended adding a DCE to the protocol, similar to combining ranking and TTO responses in the original EQ-5D-3L valuation studies.[12] Stolk et al[11] and Rowen et al[13] both suggested that a hybrid model based on both cardinal and ordinal responses may need fewer interviews under the assumption that the cardinal and ordinal responses produce similar values.[10] Slimming down the cTTO part of EQ-VT protocol was, therefore, introduced as an interesting possibility.

Taking this information into consideration, we undertook this collaborative study under the leadership of the Institute for Clinical Effectiveness and Health Policy and the National Institute of Health of Peru (INS) and funded and supported by the INS and the EuroQol Research Foundation. Our aims were to produce EQ-5D-5L values from the perspective of the Peruvian general population, to demonstrate the feasibility of a "Lite" protocol that relies on fewer respondents, and to directly compare cTTO and DCE value sets under the EQ-VT protocol. Specifically, we hypothesize that the cTTO and DCE value sets are inherently different owing to differences in their assumptions (eg, constant proportionality).

Methods

Our study followed the latest EuroQol protocol for EQ-5D-5L valuation studies (EQ-VT version 2.1),[9,11] which showed a significant improvement from version 1 in quality control and feedback to respondents. A random sample of adults (N = 1000) was recruited in Lima, Arequipa, and Iquitos for a household interview survey. Some of the respondents (N = 300) were randomly selected to first complete 11 cTTOs. All of the respondents completed 10 paired comparisons with 5 EQ-5D-5L attributes (ie, the "Lite" protocol). In addition, all the sample (N = 1000) completed 12 matched pairs (ie, A vs B and B vs C) with EQ-5D-5L and lifespan attributes. The matched pairs were added to facilitate the comparison of cTTO and DCE value sets and to assess the performance of a newer preference-elicitation task and alternative approaches to experimental design (generator-developed and D-efficient designs).

Descriptive System

The EQ-5D descriptive system uses 5 attributes to characterize health-related quality of life: mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD). Each attribute has 5 ordinal levels (from the best [level 1], having no problems, to the worst [level 5], having severe problems). The EQ-5D-5L can classify persons into 3125 unique health states, ranging from no problems at all (11111) to extreme problems in all 5 attributes (55555).[14] The descriptions in the cTTO and matched pairs included a sixth attribute describing remaining years of life in half-year increments, from 0 (immediate death) to 20 years. Two outcomes serve as anchors for the QALY scale: 0 QALYs is immediate death; and no health problems (11111) for 1 year followed by death is 1 QALY.

Models in Health Valuation

In this article, the value V of a health outcome (Q,T) is defined as a proportional relationship between the independent values of quality of life (Q) and lifespan (T): $V = v_1(Q) \times v_2(T)$, where $v_1(Q)$ is an additive regression of the 25 EQ-5D-5L attribute levels.

$$v_1(Q) = 1 - \left(\begin{array}{c} \beta_{1,1}MO_1 + \beta_{1,2}MO_2 + \beta_{1,3}MO_3 + \beta_{1,4}MO_4 + \beta_{1,5}MO_5 + \\ \beta_{2,1}SC_1 + \beta_{2,2}SC_2 + \beta_{2,3}SC_3 + \beta_{2,4}SC_4 + \beta_{2,5}SC_5 + \\ \beta_{3,1}UA_1 + \beta_{3,2}UA_2 + \beta_{3,3}UA_3 + \beta_{3,4}UA_4 + \beta_{3,5}UA_5 + \\ \beta_{4,1}PD_1 + \beta_{4,2}PD_2 + \beta_{4,3}PD_3 + \beta_{4,4}PD_4 + \beta_{4,5}PD_5 + \\ \beta_{5,1}AD_1 + \beta_{5,2}AD_2 + \beta_{5,3}AD_3 + \beta_{5,4}AD_4 + \beta_{5,5}AD_5 \end{array} \right)$$

On a QALY scale, the first coefficient of each attribute, $\beta_{.,1}$ is 0 by construction (ie, $v_1[11111] = 1$). Therefore, this regression has 20 main-effect parameters, each representing an incremental difference between levels (eg, $\beta_{1,3}$ represents the difference in value between levels 2 and 3 of the first attribute, mobility [MO]). We hypothesize that each parameter is non-negative (ie, more severe health problems reduce value).

On a QALY scale, v_2(immediate death) = 0 and v_2(1 year) = 1 by construction. The cTTO implies constant proportionality, such that (ie, $v_2[T] = T$). In contrast, multiple studies have found that the value of each additional year of life is decreasing, either due to discounting of future events or to the marginal decreasing utility of lifespan.[15,16] To relax the constant proportionality assumption, 3 temporal functions have been proposed: power (ie, $v_2[T] = T^{\alpha}$), exponential ($[1 - \exp(-r \times T)] / [1 - \exp(-r)]$), and hyperbolic ($\ln[1 + r \times T] / \ln[1 + r]$). In this study, we thus relaxed the constant proportionality assumption in the analysis of the DCE data and hypothesized that the power α is less than 1, and that the exponential and hyperbolic discount rates r are positive.

Preference-Elicitation Tasks

The 3 preference-elicitation tasks are the cTTO, the latent-scale pairs, and the matched pairs.

In this study, 300 respondents started with 11 cTTO tasks. Each cTTO task included an adaptive series of choices between 2 hypothetical lives, one life involving 10 years with health problems (with and without 10 years of lead time) and another life involving the same or fewer life years but without health problems. The iterative process starts with a dominant pair (ie, 10 years with health problems and 10 years without health problems) and typically ends at the point of indifference when the respondent reports that the 2 objects are the same in value (ie, cardinal response). If a respondent prefers "immediate death" to 10 years of lead time followed by 10 years with health problems (ie, exhausting the cTTO response range and censoring the indifference statement), the task ends in an inequality statement (ie, ordinal response). Its cardinal responses can be directly converted to a scale from –1 to 1 QALYs; however, the ordinal response of the cTTO cannot be directly converted. In Appendix 1 (in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.004), we show the screenshots that the person experience while performing the cTTO.

Regardless of their cTTO participation, all respondents completed a DCE with 10 latent-scale pairs and 12 matched pairs. Each latent-scale pair included a single choice between 2 hypothetical alternatives (A vs B). Each matched pair included 2 choices between 3 hypothetical alternatives (A vs B and B vs C) such that an object from the first pair ("B") was in the choice set of the second pair, independently of the initial choice. The alternatives of the latent-scale pairs were described using the 5 EQ-5D-5L attributes only; however, the alternatives of the matched pairs had a sixth lifespan attribute. The new object in the second pair (C) refers to 1 of 2 fixed alternatives: immediate death (for 50% of the respondents), or a life that has a shorter lifespan than A and B and no health problems for the other 50% of respondents.

In this study, the 300 respondents provided 11 cardinal or ordinal cTTO responses, and all 1000 respondents provided 34 ordinal DCE responses (10 latent-scale and 12 matched pair [with 2 responses per matched pair]). In Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.004, we show the screenshots of the person experience while performing a latent scale pair and 1 version of a matched pair.

Warm-Up Exercise

As a warm-up exercise, the 300 respondents under the Lite protocol completed 5 practice tasks before the 11 cTTO tasks. The first task described a person in a wheelchair for 10 years. The second task adapted its description based on the initial response: if a respondent preferred being in a wheelchair for 10 years over "immediate death," the second outcome was "much worse than being in a wheelchair" (otherwise, the second outcome was "much better than being in a wheelchair"). After giving responses better than and worse than "immediate death," the respondent completed the 3 additional warm-up tasks using the profiles 21121, 35554, and 15411. A key part of the cTTO warm-up and practice tasks is to familiarize respondents with the entire range of attainable values and acknowledge that the transitioning between the better-than-dead and worse-than-dead parts of the scale is notoriously hard for both the participants and the interviewer, potentially giving rise to interviewer effects. For comparison, the latent scale pairs had no warm-up exercise and the matched pairs had 1 practice warm-up task.

Experimental Design

In Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.004, we detail the experimental design of the cTTO, latent scale pairs, and matched pairs. Overall, it included 31 cTTO objects, 260 latent-scale pairs, and 699 matched pairs. The cTTO included 31 objects (25 EQ-5D-5L outcomes taken from an orthogonal array, 5 mild outcomes [21111, 12111, 11211, 11121, 11112], and the worst outcome or "pits" [55555]). The latent-scale pairs included 80 pairs in the cTTO arm (N = 300) and 180 other pairs for the rest (N = 700). The design of the matched pairs had 2 sets of criteria and 2 construction methods. The first set criterion allowed for differential lifespans in the first pair and included "immediate death" as the third object in the second pair.[17] The second set criterion included the same lifespan in the first pair, and the new object in the second pair had a shorter lifespan and no health problems.[18] The 2 construction methods for the matched pairs included a generator-developed design and a Bayesian D-efficient design. This complex design was chosen to allow future investigation of the impact of methodological choices about the design on the results. A comparison between design approaches will be part of a study initiated by the valuation working group of EuroQol (charged with the aim of addressing questions about valuation techniques) in collaboration with Deborah Street and Marcel Jonker. Analyses specific to this component were not included in this article.

Sample and Recruitment

We recruited a population-based sample of 1000 adults, 18 to 75 years old, from 3 major cities located on the coast, in the highlands, and in the jungle of Peru: Lima (the capital, planned N = 600), Arequipa (N = 200), and Iquitos (N = 200). In 2015, more than a third of the country's population (35.7%) lived in these 3 cities.[19] The sample size was determined on the basis of other valuation studies performed around the world.[20–28] Because health preferences are considered likely to vary by age, sex, and geographic area, the sampling method was stratified accordingly to obtain a representative sample. The multistage sampling design consisted of 4 stages stratified by socioeconomic level (census enumeration areas, blocks within each area, households selected from each block using systematic sampling, and 1 member of each household between 18 and 75 years old). The final sampling was aimed at having 1 respondent per household, stratified by sex (50% women and 50% men) and age categories (equal distribution in the following intervals: 18-35, 36-50, 51-64, and 65-75). To improve the representativeness of this sample, we applied differential weights in the analysis according to the most recent census data from Peru (2017).[19] The sampling was carried out by Peru National Statistics Office (Instituto Nacional de Estadística e Informática).

Interviewers and Their Training

The INS hired 11 local medical students who were specifically trained in the EQ-VT protocol. In addition, one of the researchers (R.T.) conducted interviews during site visits. The interviewers completed a rigorous face-to-face 5-day training, which included a go/no-go decision after each of them performed 5 to 10 pilot interviews. The interviewers also completed a 2-day training on ethics and on protocol methodology using standardized materials; mock interviews were carried out as demonstrations and among interviewers. Afterward, the team undertook a 2-day pilot in which health professionals and administrative personnel at the INS were interviewed. The researchers supervised all the pilot interviews and performed a quality-control analysis. Each interviewer received personal and group feedback on their performance. Finally, the researchers held a retraining session highlighting key elements to improving the quality of the interviews. Based on pilot performance, 4 interviewers were selected to collect cTTO and DCE responses (2 in Lima, 1 in Arequipa, and 1 in Iquitos). The remaining 7 interviewers collected DCE responses only (5 in Lima, 1 in Arequipa, and 1 in Iquitos).

Quality Control

Each interviewer was assigned districts with a specific number of houses to visit. Interviews were conducted offline using laptops. Once the data were uploaded, the interviewers were monitored, and feedback given if needed, at least weekly to ensure data quality. Quality-control analyses were conducted to verify that each interviewer followed the adequate methodology for interview and data collection. In brief, the analysis included cTTO process indicators such as time spent on each task, number of moves, number of responses better than and worse than immediate death, and presence of interviewer effects in the distribution of responses. It also indicated the presence of inconsistencies (eg, if respondents evaluated some profiles as worse than the worst possible profile, 55555). Latent-scale pair responses had time indicators and we flagged unlikely patterns (ie, choosing always A, B), but the matched-pair responses had no specific indicators. In addition, one of the researchers accompanied interviewers in the field in approximately 15% of the interviews.

The protocol was approved by Peru INS institutional review board (protocol number OC-0033-17). All respondents signed informed consent. The consent forms were kept at the INS, and only the investigators had access to them. Participant information was deidentified to protect confidentiality.

Analysis

The analyses included the estimation of 4 standard models, each with the 20 incremental parameters: a cTTO heteroscedastic tobit model (N = 300) with censoring at –1[29] and 3 DCE Zermelo-Bradley-Terry models with a power function (N = 300 in the subsample that performed cTTO initially, 700 in the subsample that did not, and 1000 analyzing the whole study sample).[15] Identical models were estimated to produce the recently published EQ-5D-5L value sets for the United States.[29–31]

In each estimation, validity was defined by the sign and significance of the 20 incremental parameters. In the DCE models, we further hypothesized that the power parameter was less than 1 (ie, rejecting constant proportionality and discounting future events). The feasibility of the “Lite” protocol was inferred from the cTTO and DCE value sets. These results also facilitated the direct comparison of the cTTO and DCE value sets (ie, within the same respondents). As a confirmatory analysis, the DCE value sets (N=300 and 700) were compared.

Furthermore, the 20 parameters were applied to create values for the 3125 EQ-5D-5L profiles and examined for the number and proportion of negative values (ie, outcomes worse than immediate death) and the range of values (ie, ceiling and floor effects). The association between the 20 parameters of the DCE models (ie, 300 vs 700 respondents) was compared using Pearson correlations as a confirmatory analysis.

Under the “Lite” protocol, we had originally considered a hybrid modeling that combines the cTTO and DCE responses. Under the assumptions of constant proportionality and that the DCE and cTTO tasks produce similar parameters, the protocol could safely lower the number of cTTO responses and complement them with DCE responses to fill in the gaps.[10,12] Nevertheless, after estimating their values sets separately, the equivalence of their parameters was rejected ($P < .001$). Furthermore, the DCE rejected the constant proportionality assumption implied by the cTTO. Therefore, we opted for a conservative approach and refrained from pooling the cTTO and DCE data. For the sake of exhaustivity and transparency, the hybrid model results are shown in Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.004.

For sensitivity analyses, Appendix 2 (in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.004) includes a wider range of models. Appendix 2 Tables 1 and 2 in Supplemental Materials (found at https://doi.org/10.1016/j.jval.2020.05.004) include a total of 12 models: (no censoring, censoring at –1 QALYs, and censoring at 0 [immediate death]) × (with and without heteroscedasticity) × (with and without 55555 and mild profiles—all domains in “1” and only 1 in “2”). Appendix 2 Tables 3 to 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.004 include 18 DCE models: (N = 300, 700, 1000) × (latent-scale only, matched only, all pairs) × (Zermelo-Bradley-Terry [ZBT], logit). Appendix 2 Table 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.004 includes the 6 ZBT models: (N = 300, 700, 1000) × (exponential × hyperbolic). Overall, these analyses describe the effect of the 55555 and mild profiles on the cTTO results, the associations between the ZBT and logit models, and the association among the 3 temporal functions (power, exponential, and hyperbolic). We summarize these effects using Pearson correlations on the 20 parameters. All analyses were carried out in STATA statistical package.

Results

A sample of 1000 respondents was recruited from April 2018 to February 2019 (Table 1), including 587 respondents from Lima, 208 from Arequipa, and 205 from Iquitos. The response rate was 49%. A total of 77 respondents (7.7%) did not finish the whole interview and dropped out before the matched-pairs section; and 30 respondents (3%) did not complete the 12 matched pairs. According to the predefined quality criteria, 31 cTTO interviews (10%) were flagged and none of the DCE-only interviews were flagged (0%). None of the interviewers had more than 20% of their cTTO interviews flagged (range: 0%-19%), the main reason being short time in wheelchair examples (13 interviews).

Under the “Lite” protocol, the mean interview duration was 11.7 minutes (SD 5.7) for cTTO, 10.9 minutes (SD 6.2) for the latent-scale pairs, and 9.7 (SD 4.3) for the matched pairs. For the remaining respondents, the mean interview duration of the DCE component was 14.2 minutes (SD 6.9) for the latent-scale pairs and 11.5 minutes (SD 4.9) for the matched pairs. Apart from having a longer interview duration, and despite randomization, the respondents under the “Lite” protocol were slightly more likely to be male (50.3% vs 43.6%) and reported slightly more health problems and less self-reported overall health on the visual analog scale (a 3.6-point difference in a 0-100 scale) compared with the remaining respondents (Table 1). No respondents were excluded from the analysis.

The cTTO Warm-Up Exercise and Interviewers’ Effect

In the first practice task, 40% (119 of 300) respondents stated that being in a wheelchair for 10 years was worse than immediate death (aka, worse than dead [WTD] response). The interviewer was not associated with the likelihood of a WTD response ($P = .597$). By the second task, 82% (247) gave at least 1 WTD response (which should have been close to 100% if protocol compliance had been perfect), and this proportion varied among the 5 interviewers (65%, 75%, 84%, 86%, and 93%; $P = 0.014$). If an interviewer had a WTD proportion less than 60%, all of the interviewer’s cases were supposed to be discarded as a form of quality control. By the fifth and last task in the exercise, 89% (267) gave at least 1 WTD response, and this proportion also varied among the 5 interviewers (81%, 83%, 89%, 92%, and 98%; $P = .031$).

In the cTTO tasks after the warm-up exercise, all respondents were assigned the pits profile (55555), and 85% gave a WTD response. This proportion again varied by interviewer (79%, 82%, 83%, 85%, and 98%; $P = .037$), pointing to the potential presence of interviewer effects or regional differences in values. Among the 11 cTTO responses, the number of WTD responses also varied by interviewer (3.05, 3.73, 3.78, 4.60, and 6.07; $P < .001$).

Among the respondents who completed the cTTO (N = 300), 22% gave a WTD response in the matched-pair warm-up exercise, compared with 16% in those who did not complete the cTTO (N = 700). Overall, the WTD response was 45% in all matched pairs among the respondents who completed the cTTO and 28% in those who did not complete the cTTO. In summary, respondents who completed the cTTO (N = 300) had a higher proportion of the WTD responses in the matched-pair warm-up exercise and subsequent tasks ($P < .001$).

Table 1. Sociodemographic characteristics.

	N = 1000	N = 300	N = 700	*P* value
	N (%)	N (%)	N (%)	
Age in years*	45.6 (15.8)	46.8 (16.8)	45.1 (15.3)	.124
Sex[†]				
Female	544 (54.4)	149 (49.7)	395 (56.4)	.049
Male	456 (45.6)	151 (50.3)	305 (43.6)	
Mobility[‡]				
No problems in walking	737 (73.7)	203 (67.7)	534 (76.3)	.005
Slight problems in walking	179 (17.9)	72 (24.0)	107 (15.3)	
Moderate problems in walking	72 (7.2)	22 (7.3)	50 (7.1)	
Severe problems in walking	9 (0.9)	1 (0.3)	8 (1.1)	
Unable to walk	3 (0.3)	2 (0.7)	1 (0.1)	
Self-care[‡]				
No problems washing or dressing myself	923 (92.3)	265 (88.3)	658 (94.0)	.005
Slight problems washing or dressing myself	53 (5.3)	26 (8.7)	27 (3.9)	
Moderate problems washing or dressing myself	18 (1.8)	5 (1.7)	13 (1.9)	
Severe problems washing or dressing myself	3 (0.3)	2 (0.7)	1 (0.1)	
Unable to wash or dress myself	3 (0.3)	2 (0.7)	1 (0.1)	
Usual activities[‡]				
No problems doing my usual activities	815 (81.5)	227 (75.7)	588 (84.0)	.014
Slight problems doing my usual activities	132 (13.2)	53 (17.7)	79 (11.3)	
Moderate problems doing my usual activities	44 (4.4)	15 (5.0)	29 (4.1)	
Severe problems doing my usual activities	4 (0.4)	2 (0.7)	2 (0.3)	
Unable to doing my usual activities	5 (0.5)	3 (1.0)	2 (0.3)	
Pain/discomfort[‡]				
No pain or discomfort	450 (45.0)	119 (39.7)	331 (47.3)	.005
Slight pain or discomfort	374 (37.4)	138 (46.0)	236 (33.7)	
Moderate pain or discomfort	139 (13.9)	34 (11.3)	105 (15.0)	
Severe pain or discomfort	36 (3.6)	9 (3.0)	27 (3.9)	
Extreme pain or discomfort	1 (0.1)	0 (0.0)	1 (0.1)	
Anxiety/depression[‡]				
Not anxious or depressed	626 (62.6)	179 (59.7)	447 (63.9)	.250
Slightly anxious or depressed	238 (23.8)	76 (25.3)	162 (23.1)	
Moderately anxious or depressed	120 (12.0)	37 (12.3)	83 (11.9)	
Severely anxious or depressed	16 (1.6)	8 (2.7)	8 (1.1)	
Extremely anxious or depressed	0 (0.0)	0 (0.0)	0 (0.0)	
Self-reported health[†]				
11111	306 (30.6)	79 (26.3)	227 (32.4)	.055
Any other profile	694 (69.4)	221 (73.7)	473 (67.6)	
Self-rated health using EQ VAS[†]				
<80	390 (39.0)	98 (32.7)	292 (41.7)	.061
80-89	299 (29.9)	97 (32.3)	202 (28.9)	
90-99	235 (23.5)	80 (26.7)	155 (22.1)	
100	76 (7.6)	25 (8.3)	51 (7.3)	
EQ VAS*	76.6 (16.6)	79.1 (15.2)	75.5 (17.0)	.001

EQ indicates EuroQol; VAS, visual analog scale.
*Mean (standard deviation)—*t* test was used to compare the 2 groups.
[†]Chi-squared test was used to compare the 2 groups.
[‡]Fisher exact test was used to compare the 2 groups.

EQ-5D-5L Value Sets Under the "Lite" Protocol (N = 300)

The first 2 models in Table 2 present the cTTO and DCE value sets under the "Lite" protocol (N = 300), confirming that all 20 parameters are non-negative (ie, logically consistent). Each incremental coefficient represents the reduction in value associated with an incremental increase in severity (hence, all cTTO and DCE parameters are positive). Figure 1 illustrates the effects of health problems on a QALY scale and also visually shows the moderate agreement of cTTO and DCE values. Table 3 shows the moderate correlation between the 20 cTTO and DCE parameters (Pearson correlation = 0.54). The DCE results also reject the constant proportionality assumption (ie, power < 1; $P < .001$).

In precision, the cTTO parameters have wider 95% confidence intervals than the DCE parameters, regardless of sample. Among the 20 parameters, 12 cTTO parameters are significant ($P < .05$) and all of the DCE parameters are significant.

Figure 2 shows the distribution of the 3125 EQ-5D-5L values estimated with the 4 different models and portions of the data. Under the "Lite" protocol, the cTTO and DCE approaches produced values with a similar range but differ in the percentage of negative values (43.6% of the cTTO values and 25.9% of the DCE values).

Table 2. cTTO and DCE value sets.

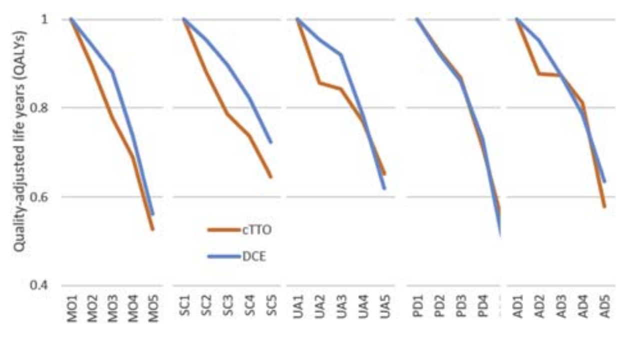
	Heteroscedastic model with censoring at -1			Zermelo-Bradley-Terry model with a power function								
	cTTO (N = 300)			DCE (N = 300)			DCE (N = 700)			DCE (N = 1000)		
	Coef.	95% CI		Coef.	95% CI		Coef.	95% CI		Coef.	95% CI	
Mobility												
Level 1-2	0.104	0.039	0.168	0.057	0.031	0.082	0.042	0.031	0.053	0.048	0.038	0.059
Level 2-3	0.119	0.049	0.190	0.061	0.033	0.090	0.025	0.017	0.033	0.034	0.025	0.043
Level 3-4	0.089	−0.003	0.182	0.146	0.091	0.200	0.075	0.056	0.094	0.093	0.074	0.112
Level 4-5	0.161	0.079	0.242	0.174	0.105	0.243	0.107	0.077	0.138	0.129	0.099	0.158
Self-care												
Level 1-2	0.117	0.066	0.168	0.045	0.018	0.072	0.033	0.022	0.043	0.038	0.028	0.048
Level 2-3	0.097	−0.057	0.252	0.058	0.031	0.086	0.023	0.015	0.031	0.030	0.022	0.038
Level 3-4	0.050	−0.071	0.171	0.074	0.040	0.107	0.037	0.026	0.048	0.048	0.037	0.059
Level 4-5	0.091	−0.020	0.203	0.101	0.064	0.138	0.049	0.034	0.064	0.062	0.048	0.076
Usual activities												
Level 1-2	0.143	0.093	0.194	0.046	0.026	0.067	0.023	0.015	0.031	0.028	0.020	0.036
Level 2-3	0.014	−0.067	0.096	0.034	0.019	0.049	0.021	0.014	0.029	0.025	0.018	0.032
Level 3-4	0.074	0.005	0.143	0.138	0.091	0.185	0.049	0.035	0.063	0.067	0.052	0.082
Level 4-5	0.116	0.041	0.192	0.163	0.101	0.226	0.106	0.079	0.133	0.127	0.100	0.153
Pain/discomfort												
Level 1-2	0.072	0.027	0.117	0.076	0.051	0.101	0.023	0.015	0.030	0.036	0.027	0.044
Level 2-3	0.060	−0.018	0.139	0.063	0.042	0.084	0.035	0.026	0.044	0.043	0.035	0.052
Level 3-4	0.155	0.073	0.238	0.132	0.084	0.180	0.051	0.037	0.065	0.070	0.055	0.085
Level 4-5	0.189	0.084	0.293	0.250	0.169	0.331	0.069	0.050	0.089	0.108	0.085	0.131
Anxiety/depression												
Level 1-2	0.123	0.060	0.186	0.047	0.022	0.072	0.034	0.024	0.044	0.040	0.030	0.049
Level 2-3	0.003	−0.090	0.096	0.079	0.045	0.114	0.028	0.019	0.037	0.039	0.029	0.049
Level 3-4	0.062	−0.014	0.138	0.087	0.050	0.124	0.053	0.037	0.069	0.063	0.048	0.078
Level 4-5	0.234	0.107	0.361	0.152	0.088	0.217	0.067	0.048	0.087	0.087	0.067	0.107
Lifespan in years												
Power	N/A	N/A	N/A	0.444	0.274	0.614	0.285	0.193	0.377	0.340	0.253	0.427
QALY predictions												
Range		−1.076	0.928		−0.984	0.955		0.048	0.977		−0.213	0.972
%WTD	43.616			25.920			0.000			0.640		

N/A = cTTO assumed. MAE = 0.464.
CI indicates confidence interval; cTTO, composite time trade-off; DCE, discrete choice experiment; MAE, mean absolute error; N/A, not applicable QALY, quality-adjusted life-years; WTD, worse than death.

Confirmatory Analysis of the DCE Value Set

As a confirmatory analysis, we analyzed the DCE results for the 700 people who did not complete the cTTO task first compared with the "Lite" DCE values; its values have a narrower range (eg, no negative values). Its parameters were highly correlated with the "Lite" DCE values (Pearson's correlation 0.80; Table 3). Also, the DCE parameters based on the latent-scale pairs and the matched pair (N = 700) were highly correlated (Pearson's

Figure 1. The effect of health problems on a quality-adjusted life year scale (N = 300).* For example, the effect of being unable to walk (5111) reduces the cTTO and DCE values from 1 to 0.527 and 0.562, respectively.



correlation 0.76). More generally, the cTTO parameters were, at best, moderately correlated with any of the DCE parameters (Pearson's correlation <0.60), which suggests that the parameter differences are only partially explained by the scale.

In Appendix 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.004, we present the values of the 3125 EQ-5D-5L profiles using the 20 parameter models for the cTTO (N = 300) and DCE (N = 1000) value sets.

Sensitivity Analysis of the DCE Value Set

We summarize the sensitivity analyses in Table 3 and give additional details of model results in Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2020.05.004. The first sensitivity analysis examined the effects of dropping the worst (55555) and 5 mild profiles (all domains in "1" and only 1 in "2") from cTTO analysis. The Pearson correlation between the cTTO and DCE parameters increased in every model after dropping these profiles (increase in Pearson's correlation from +0.09 to +0.17).

The second sensitivity analysis examined the effects of replacing the ZBT with a multinomial logit model. The logit models produced more negative and insignificant parameters, but their parameters were highly correlated with the ZBT parameters (Pearson correlations >0.8).

The third sensitivity analysis examined the effects of replacing the power function with exponential or hyperbolic functions. Regardless of function, the parameters were nearly identical, except for a few differences in range (Pearson correlation >0.99). Unlike the power and exponential functions, the rates under the hyperbolic function were not significant ($P > .05$).

Discussion

Our study examined a population-based sample of 3 of the largest Peruvian cities and provides the first EQ-5D-5L value sets to inform health policy in Peru, and the second in Latin America (besides Uruguay).[32] Because cTTO and DCE data were not poolable and the cTTO has been used to produce EQ-5D-5L value sets in many other countries, and after discussing this issue with the EuroQol Executive Committee, the recommended EQ-5D-5L value set is the one based on the cTTO-only sample (N = 300). This study produced 2 value sets for Peru using 2 different types of preference evidence: cTTO (N = 300) and DCE (N = 1000). The models used to estimate these EQ-5D-5L value sets are the same as those reported in recent US studies[29,31] and may be compared in future work.[15,29,31]

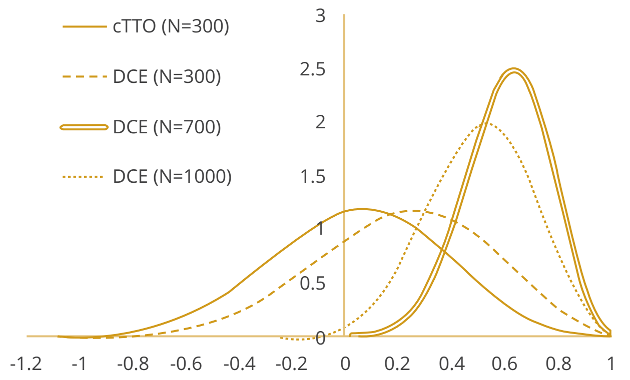
Although this "Lite" protocol produced a viable value set based on only the cTTO responses, it did not work as expected. We had anticipated the merge of the cTTO and DCE responses but refrained from doing so because they produced different parameters (Pearson correlation = 0.541). Perfect agreement was not expected because their analyses make different distributional and temporal assumptions. Also, because neither of the methods is the gold standard to capture preferences, it is importance to advance research regarding how much divergence is reasonable and to better understand the systematic and random differences introduced by the methods. Nevertheless, the discordance in this case

Table 3. Pearson correlation of the 20 parameters by model, whole sample and cTTO and DCE subsamples, cumulative density function (CDF), and discount function.

	N = 300	N = 700	N = 1000
Correlations between samples and models*			
cTTO (N = 300) vs DCE	0.541	0.461	0.503
cTTO (N = 300) vs latent-scale pairs	0.331	0.386	0.420
cTTO (N = 300) vs matched pairs	0.534	0.485	0.519
DCE (N = 300) vs DCE		0.800	0.895
Latent-scale pairs vs matched pairs	0.484	0.760	0.810
Sensitivity analysis 1: correlations after removing the worst and 5 mild cTTO profiles			
cTTO (N = 300) vs DCE	0.713	0.608	0.664
cTTO (N = 300) vs latent-scale pairs	0.470	0.480	0.547
cTTO (N = 300) vs matched pairs	0.688	0.655	0.691
Sensitivity analysis 2: correlations under logit and ZBT			
Latent-scale pairs (logit vs ZBT)	0.820	0.907	0.872
Matched pairs (logit vs ZBT)	0.901	0.926	0.945
All pairs (logit vs ZBT)	0.901	0.895	0.897
Sensitivity analysis 3: correlation under alternative discount functions			
Exponential vs power	1.000	1.000	1.000
Hyperbolic vs power	1.000	1.000	1.000

cTTO indicates composite time trade-off; DCE, discrete choice experiment; ZBT, Zermelo-Bradley-Terry model.
*Correlations measure relative agreement in the 20 incremental parameters and allow for differences in scale. To better understand differences in scale, compare the ranges (next to last row of Table 2).

Figure 2. Distribution of 3,125 EQ-5D-5L values by method and sample.



was significant and substantive, particularly the number of EQ-5D scenarios worse than immediate death (0 QALYs). Despite producing EQ-5D-5L value sets for Peru, the results casts doubt about the feasibility of a "Lite" protocol like the one in this study. Future research that can be developed in parallel with the full EQ-VT protocol is needed on other "Lite" protocols, such as pairing the reduction of sample size with a larger expansion of the task. Moreover, our understanding of the differences between the cTTO and DCE values is limited by the different designs in the 2 subgroups of responses (N = 300 and N = 700) that included different interviewers.

Comparing the cTTO and DCE value sets, the analysis shows 3 fundamental differences. First, the DCE value sets are more precise (smaller confidence intervals, fewer insignificant parameters) than the cTTO value set. Second, each DCE model rejected the constant proportionality assumption implicit to the cTTO value set. This assumption has also been rejected by several previous studies.[15,16] Third, the range of the DCE values is about twice the size when the DCE is after the cTTO (n = 300) than in those who did not undergo cTTO (n = 700) (–0.984 to 0.955 vs 0.048-0.977), which may be attributable to the cTTO warm-up exercise or the interviewers, and begs for further exploration. To illustrate this pattern more clearly: respondents who completed the cTTO first had twice the odds of choosing immediate death in the matched pairs (44.9% in the cTTO group vs 27.5% in the DCE only group; odds ratio = 2.15; $P < .001$).

One possible explanation for this doubling is that the cTTO improves understanding of immediate death and that if not warmed up properly for questions involving life and death, respondents may avoid choosing severe outcomes like immediate death. Another explanation is that the cTTO warm-up exercise required respondents to choose immediate death almost by design, and induced a bias. In practice, however, some interviewers initially struggled more with this aspect of the EQ-VT protocol than others, leading to interviewer (learning) effects that the quality control process sought to deter. Instead of cTTO and DCE, some theoretical arguments favor the standard gamble as the gold standard[5] in preference elicitation, but the standard gamble is rarely used in health preference research owing to mounting theoretical, empirical, and practical considerations (eg, prospect theory, numeracy).

Comparing the cTTO and DCE value sets, a reader may ask the following questions: Is it ever acceptable to assume constant proportionality (ie, zero discounting)? Which range and values are the most valid and representative of the Peruvian population? Do the cTTO warm-up exercises improve understanding of immediate death or induce a bias (ie, framing effects)? These questions may be answered in future studies. Likewise, the sensitivity analyses showed that dropping the extreme and mild profiles from the cTTO design has a substantial effect on the cTTO parameters, that the logit results are similar but less precise than the ZBT results, that the results are largely invariant to temporal specification, and that latent-scale and matched pairs produced similar parameters after rescaling. These primary and secondary results may be confirmed in the ongoing EQ-VT study in Denmark that used a similar design.

Study limitations included the 49% response rate, 10% dropout rate in the matched pairs, and small differences between the 300- and 700-person samples' baseline characteristics, which may be attributed to the regional difference in the quota sampling by metropolitan area. The order of the DCE and cTTO tasks could also have been randomized to test for sequence effects. In the cTTO design, the profiles for the worst and the 5 mild profiles may have been excluded to allow for a larger orthogonal design and to identify interaction effects. The DCE design could have been improved by including more pairs with 1 year in full health (1 QALY), which serves as an anchor for the QALY scale. The interviewer effects in the cTTO warm-up should be addressed and taken care of in future studies, particularly in the quality control process. A future study may examine preference heterogeneity, particularly in the power or discount rates.[15,29,31]

In conclusion, the study achieved its aims by producing Peruvian EQ-5D-5L value sets (one of which was based on cTTO that relied on a smaller sample compared to usual valuation protocols, and the other one using a DCE-only set). Nevertheless, these results cast doubt about the feasibility of a "Lite" protocol like the one we undertook. We also directly compared cTTO and DCE values to be used on a QALY scale. Although the cTTO value set under this "Lite" protocol is comparable to those of other countries and may serve as the official value set, the DCE (N = 1000) is an alternative way of representing the values of the Peruvian general population. Choosing between them not only may have policy implications for resource allocation in Peru but also could have significant methodological implications for the future of health valuation in the EuroQol family of instruments and in other valuation studies alike.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2020.05.004.

Article and Author Information

Accepted for Publication: May 21, 2020

Published Online: July 18, 2020

doi: https://doi.org/10.1016/j.jval.2020.05.004

Author Affiliations: Institute for Clinical Effectiveness and Health Policy, Buenos Aires, Argentina (Augustovski, Belizán, Gibbons); Unidad de Análisis y Generación de Evidencias en Salud Pública, Instituto Nacional de la Salud, Lima, Perú (Reyes, Tejada); EuroQol Research Foundation, Rotterdam, The Netherlands (Stolk); University of South Florida, Tampa, FL, USA (Craig).

Author Contributions: *Concept and design:* Augustovski, Belizán, Gibbons, Reyes, Stolk, Craig, Tejada
Acquisition of data: Belizán, Stolk, Tejada
Analysis and interpretations of data: Augustovski, Belizán, Gibbons, Stolk, Craig, Tejada
Drafting of the manuscript: Augustovski, Belizán, Gibbons, Reyes, Stolk, Craig, Tejada

Critical revision of the paper for important intellectual content: Augustovski, Belizán, Reyes, Stolk, Craig, Tejada
Statistical analysis: Augustovski, Gibbons, Craig
Provision of study materials or patients: Reyes, Stolk, Tejada
Obtaining funding: Augustovski, Stolk, Craig, Tejada
Administrative, technical, or logistic support: Reyes, Stolk, Tejada
Supervision: Augustovski, Belizan, Stolk, Tejada

Conflict of Interest Disclosures: Drs Augustovski and Craig reported receiving grants from the EuroQol Research Foundation and from the Instituto Nacional de Salud (Perú) during the conduct of the study. Dr. Stolk is an employed by the EuroQol Research Foundation. EuroQol holds the copyrights of the EQ-5D-5L. The availability of national value sets supports dissemination of the EQ-5D-5L instrument. EuroQol is a charity. Most licenses are granted for free. No other disclosures were reported.

Funding/Support: This work was supported by grants from the Instituto Nacional de la Salud, Perú and the EuroQol Research Foundation under EQ Project 20170170.

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

Acknowledgments: We acknowledge the collaboration of the EuroQol team (Kristina Ludwig and Aureliano Finch) in undertaking this protocol and IECS librarian Daniel Comandé for his help with bibliographic matters. We also appreciate the feedback received at the 2019 EuroQol Plenary Meeting from John D. Hartman, Juan Manuel Ramos Goñi, Mark Oppe, and Kim Rand on the analysis of the time trade-off and discrete-choice experiment responses. The views expressed by the authors of this publication do not necessarily reflect the views of the EuroQol Group or the Instituto Nacional de Salud.

REFERENCES

1. World Health Organization. 2015 Global Survey on Health Technology Assessment by National Authorities: main findings. https://www.who.int/health-technology-assessment/MD_HTA_oct2015_final_web2.pdf. Accessed February 20, 2020.
2. Poder TG, Bellemare CA. Importance of contextual data in producing health technology assessment recommendations: a case study. *Int J Technol Assess Health Care*. 2018;34(1):63–67.
3. Red de Evaluación de Tecnologías en Salud de las Américas (RedETSA). Miembros de RedETSA. http://redetsa.org/wp/?page_id=324. Accessed June 11, 2019.
4. International Society for Health Economics and Outcomes Research. Pharmacoeconomic guidelines around the world. https://tools.ispor.org/peguidelines/. Accessed June 11, 2019.
5. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. *Value Health*. 2009;12(suppl 1):S5–S9.
6. Norman R. The impact of duration on EQ-5D-5L value sets derived from a discrete choice experiment. In: *The EuroQol Meeting*. The Netherlands: Noordwijk; 2016.
7. Rasanen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynanen O-P, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: a systematic literature review. *Int J Technol Assess Health Care*. 2006;22(2):235–241.
8. EuroQol Research Foundation. Valuation of EQ-5D. https://euroqol.org/eq-5d-instruments/valuation-of-eq-5d/. Accessed June 11, 2019.
9. Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goni JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics*. 2016;34(10):993–1004.
10. Ramos-Goni JM, Craig BM, Oppe M, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value Health*. 2018;21(5):596–604.
11. Stolk E, Ludwig K, Rand K, van Hout B, Ramos-Goni JM. Overview, update, and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. *Value Health*. 2019;22(1):23–30.
12. Craig BM, Busschbach JJ. The episodic random utility model unifies time trade-off and discrete choice approaches in health state valuation. *Popul Health Metr*. 2009;7:3.
13. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Med Decis Making*. 2015;35(3):328–340.
14. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
15. Craig BM, Rand K, Bailey H, Stalmeier PFM. Quality-adjusted life-years without constant proportionality. *Value Health*. 2018;21(9):1124–1131.
16. Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. *Value Health*. 2018;21(8):993–1001.
17. Viney R, Norman R, Brazier J, et al. An Australian discrete choice experiment to value EQ-5D health states. *Health Econ*. 2014;23(6):729–742.
18. Jonker MF, Attema AE, Donkers B, Stolk EA, Versteegh MM. Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. *Health Econ*. 2017;26(12):1534–1547.
19. Instituto Nacional de Estadística e Informática. Población censada urbana. Series nacionales 2019. http://webapp.inei.gob.pe:8080/sirtod-series/. Accessed June 11, 2019.
20. Augustovski F, Rey-Ares L, Irazola V, et al. An EQ-5D-5L value set based on Uruguayan population preferences. *Qual Life Res*. 2016;25(2):323–333.
21. Gandhi M, Xu Y, Luo N, Cheung YB. Sample size determination for EQ-5D-5L value set studies. *Qual Life Res*. 2017;26(12):3365–3376.
22. Kim S-H, Ahn J, Ock M, et al. The EQ-5D-5L valuation study in Korea. *Qual Life Res*. 2016;25(7):1845–1852.
23. Luo N, Liu G, Li M, Guan H, Jin X, Rand-Hendriksen K. Estimating an EQ-5D-5L value set for China. *Value Health*. 2017;20(4):662–669.
24. Pattanaphesaj J, Thavorncharoensap M, Ramos-Goni JM, Tongsiri S, Ingsrisawang L, Teerawattananon Y. The EQ-5D-5L valuation study in Thailand. *Expert Rev Pharmacoecon Outcomes Res*. 2018;18(5):551–558.
25. Xie F, Pullenayegum E, Gaebel K, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care*. 2016;54(1):98–105.
26. Ludwig K, Graf von der Schulenburg J-M, Greiner W. German value set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36(6):663–674.
27. Szende A, Oppe M, Devlin N. *EQ-5D Value Sets: Inventory, Comparative Review and User Guide*. Rotterdam, The Netherlands: Springer; 2007.
28. Oppe M, van Hout B. The "power" of eliciting EQ-5D-5L values: the experimental design of the EQ-VT. https://euroqol.org/wp-content/uploads/2016/10/EuroQol-Working-Paper-Series-Manuscript-17003-Mark-Oppe.pdf. Accessed June 11, 2019.
29. Pickard AS, Law EH, Jiang R, et al. United States valuation of EQ-5D-5L health states using an international protocol. *Value Health*. 2019;22(8):931–941.
30. Jakubczyk M, Craig BM, Barra M, et al. Choice defines value: a predictive modeling competition in health preference research. *Value Health*. 2018;21(2):229–238.
31. Craig BM, Rand K. Choice defines QALYs: a US valuation of the EQ-5D-5L. *Med Care*. 2018;56(6):529–536.
32. Augustovski F, Rey-Ares L, Irazola V, et al. An EQ-5D-5L value set based on Uruguayan population preferences: report of the first experience in Latin America. *Value Health*. 2015;18(7):A810–A811.