ScienceDirect

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

Preference-Based Assessments

United States Valuation of EQ-5D-5L Health States Using an International Protocol

A. Simon Pickard, PhD,[1,*] Ernest H. Law, PharmD, PhD,[1] Ruixuan Jiang, PharmD,[1] Eleanor Pullenayegum, PhD,[2] James W. Shaw, PharmD, MPH, PhD,[3] Feng Xie, PhD,[4] Mark Oppe, PhD,[5] Kristina S. Boye, PhD,[6] Richard H. Chapman, PhD,[7] Cynthia L. Gong, PharmD, PhD,[8] Alan Balch, PhD,[9] Jan J.V. Busschbach, PhD[10]

[1]Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, University of Illinois at Chicago, Chicago, IL, USA; [2]Hospital for Sick Children and University of Toronto, Toronto, ON, Canada; [3]Bristol-Myers Squibb, Lawrenceville, NJ, USA; [4]McMaster University, Hamilton, ON, Canada; [5]Axentiva Solutions, Santa Cruz de Tenerife, Spain; [6]Eli Lilly and Company, Indianapolis, IN, USA; [7]Institute for Clinical and Economic Review, Boston, MA, USA; [8]University of Southern California, Los Angeles, CA, USA; [9]Patient Advocate Foundation, Hampton, VA, USA; [10]Section of Medical Psychology, Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands

ABSTRACT

Objective: To derive a US-based value set for the EQ-5D-5L questionnaire using an international, standardized protocol developed by the EuroQol Group.

Methods: Respondents from the US adult population were quota-sampled on the basis of age, sex, ethnicity, and race. Trained interviewers guided participants in completing composite time trade-off (cTTO) and discrete choice experiment (DCE) tasks using the EuroQol Valuation Technology software and routine quality control measures. Data were modeled using a Tobit model for cTTO data, a mixed logit model for DCE data, and a hybrid model that combined cTTO and DCE data. Model performance was compared on the basis of logical ordering of coefficients, statistical significance, parsimony, and theoretical considerations.

Results: Of 1134 respondents, 1062, 1099, and 1102 respondents provided useable cTTO, DCE, and cTTO or DCE responses, respectively, on the basis of quality control criteria and interviewer judgment. Respondent demographic characteristics and health status were similar to the 2015 US Census. The Tobit model was selected as the preferred model to generate the value set. Values ranged from −0.573 (55 555) to 1 (11 111), with 20% of all predicted health states scores less than 0 (ie, worse than dead).

Conclusions: A societal value set for the EQ-5D-5L was developed that can be used for economic evaluations and decision making in US health systems. The internationally established, standardized protocol used to develop this US-based value set was recommended by the EuroQol Group and can facilitate cross-country comparisons.

Keywords: cost-utility analysis, EQ-5D, quality-adjusted life-year, stated preference, United States, utility, value.

VALUE HEALTH. 2019; 22(8):931–941

Introduction

The EQ-5D is a measure of health that has been widely used around the world to inform resource allocation and decision making in healthcare.[1] The EQ-5D-3L, which has 3 levels for each of the 5 dimensions of health, was revised and expanded to 5 levels as the EQ-5D-5L.[2] The EQ-5D-5L has shown improved measurement properties, including greater discrimination among known groups and reduced ceiling effects compared with the EQ-5D-3L.[3–5]

In the United States, there is renewed interest in value frameworks that focus on the trade-offs between potential benefits and harms of treatment.[6–9] Despite legislation against it as a consideration in federal payer reimbursement decisions,[10] cost per quality-adjusted life-year (QALY) continues to be an important metric for evaluating the value of healthcare interventions in the United States,[8,11,12] and the EQ-5D features prominently as a preference-based measure of health for such evaluations.[13]

Much of the literature on cost-utility analysis in the United States draws from the EQ-5D-3L value set.[14] Upon publication of

* Address correspondence to: A. Simon Pickard, PhD, Department of Pharmacy Systems, Outcomes and Policy, College of Pharmacy, University of Illinois at Chicago, 833 South Wood St, MC 871, Chicago, IL 60612, USA. Email: pickard1@uic.edu

https://doi.org/10.1016/j.jval.2019.02.009

the EQ-5D-5L descriptive system, an interim crosswalk was developed to facilitate the generation of preference-based summary scores on the basis of EQ-5D-5L responses.[15] The EuroQol Group invested in a scientific program that examined alternative approaches to preference elicitation,[16,17] developed methods to improve data quality,[18] and demonstrated the robustness of these approaches across countries and languages.[19] The accumulation of evidence supported a standardized valuation protocol implemented in software named the EuroQol Valuation Technology (EQ-VT).[20,21] The EQ-VT has been refined in several developmental phases and is recommended by the EuroQol Research Foundation, a not-for-profit research foundation whose membership consists of international, multidisciplinary researchers[21–23] for conduct of EQ-5D-5L valuation studies.[24–28] The EQ-VT has been shown to improve data quality and minimize interviewer effects.[18,29] Building on this evidence to apply a standardized approach to a national value set, the aim of this study was to estimate a value set for the EQ-5D-5L that will support economic evaluation of healthcare interventions in the United States; by using an internationally established protocol, the US value set will also facilitate cross-country comparisons.

Methods

Study Design

The research design and data collection followed a research program developed by the EuroQol Group. The study protocol was approved by the institutional review board at the University of Illinois at Chicago (IRB#2017-0289). We followed the CREATE checklist for reporting key elements of valuation studies for multiattribute utility instruments.[30]

Descriptive System

Health states were described using the EQ-5D-5L, which includes 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). Each dimension describes 5 levels of problems: "no," "slight," "moderate," "severe," and "extreme" problems (or "unable to").[2] Each of the 3125 (5^5) health state profiles described by the EQ-5D-5L can be represented by a 5-digit number that ranges from 11 111 (no problems in any dimension) to 55 555 (extreme problems or unable to in all dimensions). A "misery score" can be calculated by summing all the digits of the profile to broadly categorize health states by overall severity and is not a substitute for a preference-based score.

Preference Elicitation Methods

Both composite time trade-off (cTTO) and discrete choice experiment (DCE) methods were used in the valuation protocol. The cTTO method[22,31–33] compared a suboptimal EQ-5D-5L health state to full health, using the conventional TTO to elicit "better-than-dead" values and the lead-time TTO to elicit "worse-than-dead" (WTD) values (see Appendix A in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009). If the respondent indicated that a health state was WTD, an additional 10 years of time (lead time) was granted to elicit WTD values; thus, the most negative elicited cTTO value was −1. The smallest unit of time traded was 6 months. Full health and death were used as anchors in the cTTO tasks. DCE tasks asked respondents to state their preference between 2 different EQ-5D-5L health state profiles in a pairwise comparison (see Appendix B in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009). Opt-out options (eg, "neither") were not provided.

Health State Selection

The experimental design was informed by multinational EQ-5D-5L cTTO and DCE pilot studies, which included US data.[20,22] For the cTTO tasks, 86 EQ-5D-5L health states were selected to represent a wide range of health problems.[20] Health states were grouped into 10 blocks with 10 health states per block. All 10 blocks contained the worst possible health state (55 555), 1 mild state (mild problems in 1 dimension only), and 8 health states unique to each block that varied in severity. For the DCE tasks, 196 paired comparisons constituted the experimental design.[20,22] Ten of the 196 pairs consisted of very mild health state pairs and the rest were drawn to produce a *D*-error minimized design.[20] These paired comparisons were subdivided into 28 blocks of 7 choice sets.

Sampling and Recruitment

Data were collected between May and September 2017 from 6 US metropolitan areas (Chicago, Philadelphia, Seattle, Birmingham, Phoenix, and Denver). Locations were chosen to ensure representativeness of the US population and sampling in all census regions.[34–36]

Noninstitutionalized English- and Spanish-speaking adult members of the US general population were eligible to participate. The inclusion criteria were as follows: (1) 18 years of age or older; (2) current US residency; (3) ability to understand the tasks, as judged by interviewer; (4) ability to provide informed consent; and (5) ability to complete the tasks in English or Spanish. Quota sampling was based on age, sex, ethnicity, and race to obtain a sample representative of the US general population.

Three recruitment strategies were used. First, ResearchMatch, a national, web-based recruitment tool (https://www.researchmatch.org), was used to contact potential respondents who lived near the recruitment areas. Second, the International Society of Pharmacoeconomics and Outcomes Research student chapters near recruitment sites were contacted to promote study awareness. Third, community platforms (eg, public flyers, online platforms, and community centers) were used to generate local interest.

During data collection at a given location, the study team continued to recruit participants. This multipronged approach to recruitment was guided by recent valuation study experiences and adapted to the US context.[37] Interviews were scheduled throughout each metropolitan area, and respondents were given $30 as cash incentive for study participation.

Survey Administration

All surveys were completed through interviewer-assisted data collection using laptops with the latest version of the EQ-VT software,[19] version 2.0, in Spanish or English. To maintain consistency, the same interviewers were involved throughout data collection. All interactions followed the same format. First, participants were given study information, and verbal informed consent was obtained. Second, participants reported their current level of health as described by the EQ-5D-5L and rated their health using the EuroQol visual analogue scale.[38] Third, respondents completed valuations using the cTTO method. Respondents began with 5 practice cTTO exercises. Respondents then valued 10 EQ-5D-5L health states, presented in random order. Afterwards, a feedback module was completed, where all 10 EQ-5D-5L health

states were presented in rank order on the basis of respondent-assigned cTTO values. Respondents flagged any health states that were "out of order" to ensure the unflagged responses accurately reflected their health preferences. Flagged health states could not be revalued.

Next, respondents completed 7 DCE tasks, presented in random order. The left/right display of the health state pairs was randomized at the respondent level to avoid framing effects. For all preference tasks, respondents read each health state description aloud and were encouraged to think aloud to ensure engagement and help interviewers assess comprehension. Finally, respondents completed sociodemographic and health-related questions.

Interviewer Training, Role of Interviewer, and Quality Control Process

All interviewers were trained by EuroQol scientific staff using EQ-VT 2.0.[19] Interviewers were trained to instruct, motivate, and probe respondents for understanding of valuation tasks. During data collection, quality control reports were reviewed daily to ensure cTTO data quality on the basis of several criteria[22]: (1) the interviewer used at least 3 minutes to explain the first practice cTTO task; (2) the interviewer demonstrated WTD task (ie, lead-time portion of cTTO) for the first or second practice cTTO task; (3) the respondent avoided assigning the worst health state (55 555) a cTTO value of 0.5 or more points higher than values assigned to other health states; and (4) the respondent used at least 5 minutes to complete all 10 EQ-5D-5L cTTO tasks. Interviews that did not meet all 4 criteria were discussed during quality control debriefings. These interviews were not necessarily excluded from analysis if the respondent demonstrated comprehension of the preference tasks.

Data Analysis and Modeling of EQ-5D-5L Health States

A Tobit model for cTTO data (model 1), a mixed logit model for DCE data rescaled to the health utility scale (model 2), and a hybrid model that combined cTTO and DCE data (model 3) were used to model health state preferences. The dependent variables for the Tobit and mixed logit models were cTTO values and the DCE stated choice (ie, A or B for each given health state pair), respectively. If the respondent did not comprehend the choice task(s), then the respondent's preference data (cTTO and/or DCE) were excluded from the respective analysis. Respondent-flagged cTTO responses were excluded from the base-case analysis.

Model 1 (Tobit) estimated utility decrements for each parameter. Observed cTTO values were regressed onto 19 total parameters, that is, 4 dummy variables for each health dimension representing the disutility from level 1 (no problems), the referent category, except for usual activities levels 4 and 5, which were constrained to have the same utility decrement; the model also contained a random effect to account for dependency of repeated observations within respondents (Equation 1). Model 1 left-censored the cTTO data at −1, because respondents could hypothetically continue trading beyond the left bound at −1 for WTD values. Censoring at 1, that is, right-censoring, was not pursued because 1 is the theoretical upper bound in health utilities, and the EQ-5D-5L health states were valued against full health on the cTTO. Tobit models assume that a latent variable (cTTO*) underlies the observed cTTO values and uses a likelihood function to adjust the parameter estimates for the probability of the cTTO* value beyond the censored value (ie, <-1) (Equation 2).

$$\begin{aligned} \mathrm{cTTO}_{ij} &= \mu_j + u_i + \varepsilon_{ij} \\ \mu_j &= 1 + \beta_1 \mathrm{MO2}_j + \beta_2 \mathrm{MO3}_j + \beta_3 \mathrm{MO4}_j + \beta_4 \mathrm{MO5}_j + \\ &\quad + \beta_5 \mathrm{SC2}_j + \beta_6 \mathrm{SC3}_j + \beta_7 \mathrm{SC4}_j + \beta_8 \mathrm{SC5}_j + \\ &\quad + \beta_9 \mathrm{UA2}_j + \beta_{10} \mathrm{UA3}_j + \beta_{11} (\mathrm{UA4}_j + \mathrm{UA5}_j) + \\ &\quad + \beta_{12} \mathrm{PD2}_j + \beta_{13} \mathrm{PD3}_j + \beta_{14} \mathrm{PD4}_j + \beta_{15} \mathrm{PD5}_j + \\ &\quad + \beta_{16} \mathrm{AD2}_j + \beta_{17} \mathrm{AD3}_j + \beta_{18} \mathrm{AD4}_j + \beta_{19} \mathrm{AD5}_j \\ u_i &\sim iid\ N(0, \sigma_u^2) \text{ is a subject-level random intercept} \\ \varepsilon_{ij} &\sim iid\ N(0, \sigma_j^2) \text{ is a heteroscedastic error term} \end{aligned} \tag{1}$$

where MO is mobility, SC is self-care, UA is usual activities, PD is pain/discomfort, and AD is anxiety/depression; the number following the dimension indicates level of severity (eg, MO2 is mobility level 2); iid is independent and identically distributed; i is the respondent; and j accounts for the multiple tasks completed.

$$\mathrm{cTTO} = \begin{cases} \mathrm{cTTO}^* \text{if } \mathrm{cTTO}^* > -1 \\ -1 \quad \text{if } \mathrm{cTTO}^* \leq -1 \end{cases} \tag{2}$$

Furthermore, as the observed variance of the cTTO values increased with increasing severity of the health state, we investigated several methods of modeling the heteroscedasticity of the error term, such as log link with polynomials of μ_j up to the fourth degree and a log link with 21 dummy variables corresponding to each dimension level of the EQ-5D-5L and a constant. We chose the most appropriate model (fourth-degree μ_j polynomial) considering Akaike information criterion, Bayesian information criterion, and estimation stability (Equation 3).

$$\sigma_j = \exp\left(\gamma_0 + \gamma_1 \mu_j + \gamma_2 \mu_j^2 + \gamma_3 \mu_j^3 + \gamma_4 \mu_j^4\right) \tag{3}$$

Model 2 used a mixed logit, that is, random parameter logit that can account for repeated observations, to model the DCE data. The model included regression coefficients and random effects for each of the 20 dummy variables, reflecting utility decrements associated with levels 2, 3, 4, and 5 for each of the 5 domains (Equation 4).

$$\begin{aligned} U_{it} = &\beta_{1i}(\mathrm{MO2})_j + \beta_{2i}(\mathrm{MO3})_j + \beta_{3i}(\mathrm{MO4})_j + \beta_{4i}(\mathrm{MO5})_j + \beta_{5i}(\mathrm{SC2})_j \\ &+ \beta_{6i}(\mathrm{SC3})_j + \beta_{7i}(\mathrm{SC4})_j + \beta_{8i}(\mathrm{SC5})_j + \beta_{9i}(\mathrm{UA2})_j + \beta_{10i}(\mathrm{UA3})_j \\ &+ \beta_{11i}(\mathrm{UA4})_j + \beta_{12i}(\mathrm{UA5})_j + \beta_{13i}(\mathrm{PD2})_j + \beta_{14i}(\mathrm{PD3})_j \\ &+ \beta_{15i}(\mathrm{PD4})_j + \beta_{16i}(\mathrm{PD5})_j + \beta_{17i}(\mathrm{AD2})_j + \beta_{18i}(\mathrm{AD3})_j \\ &+ \beta_{19i}(\mathrm{AD4})_j + \beta_{20i}(\mathrm{AD5})_j + \varepsilon_{it} \end{aligned} \tag{4}$$

where i is the respondent, t is the choice alternative in choice sets, U_{it} represents latent utility, $\beta_{ki} \sim N(\beta_k, \sigma_{\beta j}{}^2)$ are random effects, and ε_{it} is the residual term with an extreme value distribution. The model estimates obtained using DCE are on a latent utility scale, and consequently cannot be used independently as the basis for generating a value set on a cardinal scale. Therefore, mixed logit (model 2; Table 2) coefficients were transformed to a 1 (11 111) to 0 (dead) scale using a rescaling parameter obtained from a line fit to the scatterplot of cTTO censored means and DCE mixed logit latent health values, assuming a linear relationship[39,40] (see Appendix F in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009).

Finally, model 3 used a hybrid approach. The hybrid approach added the log-likelihood functions of Tobit (cTTO; model 4) and

conditional logit (DCE; model 5) models[39,40] to generate a single log-likelihood function with heteroscedasticity of the variance (Equations 5 and 6).

$$\begin{aligned}\text{Log likelihood} = &-\frac{1}{2}\sum_{j\in C'}\left\{\text{Log}\left(2\pi\sigma_j^2\right)+\left(\frac{y_j-x\beta}{\sigma_j}\right)^2\right\}\\ &+\sum_{j\in L'}\text{Log}\left(\Phi\left(\frac{-1-x\beta}{\sigma_j}\right)\right)\\ &+\sum_{j\in D}-y_j\times\text{Log}\left(1+e^{-(x_A-x_B)\beta/\theta}\right)\\ &+(1-y_j)\times\text{Log}\left(\frac{e^{-(x_A-x_B)\beta/\theta}}{1+e^{-(x_A-x_B)\beta/\theta}}\right)\end{aligned} \tag{5}$$

$$\begin{aligned}\sigma_j = \exp(&\gamma_1\text{MO2}_j+\gamma_2\text{MO3}_j+\gamma_3\text{MO4}_j+\gamma_4\text{MO5}_j+\gamma_5\text{SC2}_j+\gamma_6\text{SC3}_j+\gamma_7\text{SC4}_j+\gamma_8\text{SC5}_j+\gamma_9\text{UA2}_j+\gamma_{10}\text{UA3}_j+\gamma_{11}\text{UA4}_j\\ &+\gamma_{12}\text{UA5}_j+\gamma_{13}\text{PD2}_j+\gamma_{14}\text{PD3}_j+\gamma_{15}\text{PD4}_j+\gamma_{16}\text{PD5}_j+\gamma_{17}\text{AD2}_j+\gamma_{18}\text{AD3}_j+\gamma_{19}\text{AD4}_j+\gamma_{20}\text{AD5}_j)\end{aligned} \tag{6}$$

where y_j is the dependent variable, C' represents cTTO responses greater than -1, L' represents TTO responses equal to -1, and x_A and x_B represent the attributes of alternatives A and B in the paired DCE comparisons.[39,40]

The component models of the hybrid model were Tobit for cTTO (model 4; heteroscedasticity allowance, censoring at -1) and conditional logit for DCE (model 5), neither of which could account for repeated measures within a respondent (see Appendix G in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009); several approaches were used to assess whether the cTTO and DCE as modeled by the component models were compatible and therefore combinable for the hybrid model. A Bland-Altman plot was developed to examine the difference and mean for the predicted cTTO and DCE values (see Appendix H in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009). Predicted health state values from the hybrid model and its component models were plotted, and Pearson correlation, Spearman correlation, and Lin concordance coefficients were also calculated to assess model compatibility (see Appendix I in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009).

Logical ordering of parameter estimates (ie, larger utility decrements with more severe problems), statistical significance of the parameters ($P<.050$), model parsimony, and theoretical considerations related to model specifications (eg, handling of censored values, accounting for panel data, and heteroscedasticity assumptions) were all considered in model selection. Kernel density distributions were plotted for the preferred US EQ-5D-5L value set, the US EQ-5D-3L value set,[14] and the crosswalk developed by van Hout et al[15] to show the distributional properties of each scoring algorithm (eg, skewness, modality, and range of scale). Finally, sensitivity analyses were performed to evaluate the robustness of the primary analysis: (1) re-inclusion of respondent-flagged cTTO responses, (2) re-inclusion of preference data from all respondents, and (3) examination of interviewer effects. Statistical analyses were performed using SAS version 9.4 (SAS Institute, Chapel Hill, NC) and STATA version 14.2 (StataCorp, College Station, TX). No missing preference data were noted.

Results

In total, 1134 interviews were conducted from May to September 2017. Mean interview time was 58.3 ± 71.1 minutes. Thirty-two respondents (2.8%) did not comprehend the cTTO nor DCE tasks per interviewer assessment, and all valuation data from these respondents were excluded from the primary analysis (Fig. 1). The full and analytic samples were generally representative of the US adult population with respect to age, sex, race, ethnicity, prevalence of chronic conditions, and general health status (Table 1).[41–45] Characteristics of respondents excluded from cTTO- and DCE-only models are provided in Appendix C in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009.

cTTO and Discrete Choice Data

For the EQ-5D-5L cTTO tasks, data from 72 respondents were excluded per interviewer assessment, resulting in 1062 respondents providing 10 620 cTTO responses (Fig. 1). Respondents flagged 1234 cTTO responses (11.6%) using the feedback module. Respondents took an average of 6.8 ± 4.7 iterative steps before they reached their point of indifference in cTTO tasks. The observed mean cTTO values ranged from −0.366 for health state 55 555 to 0.965 for health state 11 121 (see Appendix D in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009). A total of 324 (30.5%) respondents had at least 1 inconsistency; that is, health state A dominated health state B across all dimensions, but A was assigned a lower cTTO value; 123 respondents (11.6%) had an inconsistency involving 55 555. After the feedback module, 172 (16.2%) respondents provided an inconsistent response (33 respondents [3.11%] involving 55 555). Overall, 5.6% of all eligible cTTO responses were inconsistent, and inconsistency prevalence was reduced to 2.2% after the feedback module.

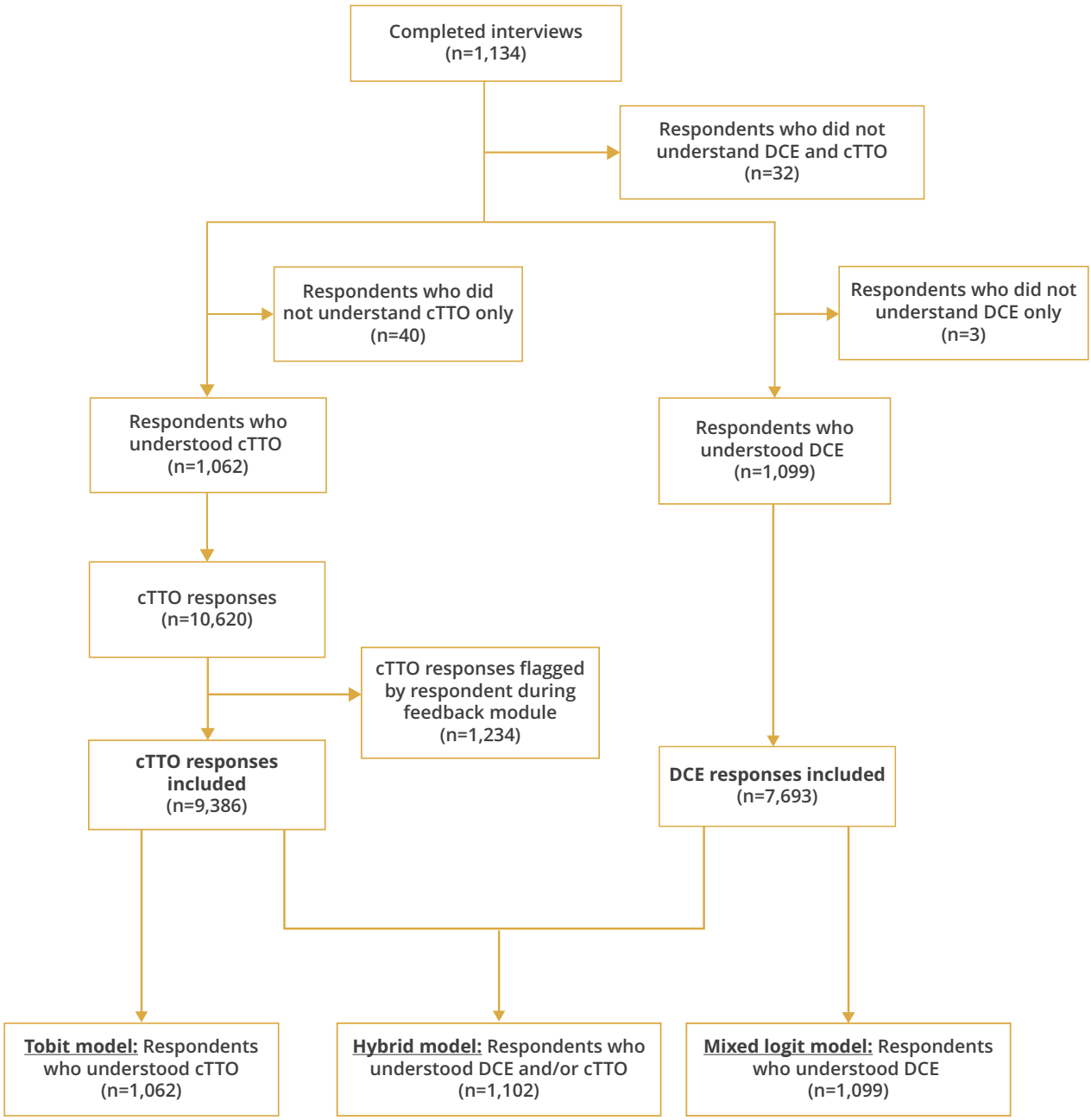
The main analysis included all unflagged cTTO valuations (10 620 − 1234 = 9386 responses) from respondents who comprehended the task per interviewer assessment. Of these, 2251 (24.0%) cTTO responses were considered WTD (Fig. 2A). The proportion of values clustered at 1, 0, and −1 was 20.5%, 5.1%, and 14.7%, respectively. Lower mean cTTO values and larger standard deviations were observed as the health state misery score increased (Fig. 2B).

For the DCE tasks, 1099 respondents were included in the DCE-only model; that is, 35 respondents did not understand the DCE. As the difference in overall severity between the 2 states increased, respondents were more likely to choose the state with the lower severity (see Appendix E in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009).

Modeling EQ-5D-5L Health States

Model 2 had 2 disordered parameters, whereas models 1 and 3 had no disordering (Table 2). Coefficients for model 2 were rescaled using the slope of the line fit to the censored cTTO disutility means and the mixed logit DCE latent values

Figure 1. Respondents included in the US EQ-5D-5L valuation study and responses retained.



cTTO indicates composite time trade-off; DCE, discrete choice experiment; EQ-5D-5L, 5-level EuroQol 5-dimensional questionnaire.

(scale = 0.034). All model parameter estimates were statistically significantly different from 0 ($P<.0001$). Dimension ranking in terms of relative importance differed slightly across the 3 models. For the mixed logit (model 2) and hybrid (model 3) models, the relative importance was as follows: pain/discomfort (most important), anxiety/depression, mobility, self-care, and usual activities (least important). For the Tobit TTO model (model 1), mobility was marginally more important than anxiety/depression (0.322 and 0.321, respectively). The hybrid model was in agreement with the cTTO and DCE as modeled by its component models (models 4 and 5; see Appendix G in Supplemental Materials).

Preferred Model and Value Set

The Tobit cTTO model was selected as the preferred model on the basis of its performance with respect to statistical significance of the estimates, ability to handle left-censored cTTO data, account for panel data, and heteroscedasticity (Table 2). The largest utility decrement for a dimension level was pain/discomfort level 5

Table 1. Respondent characteristics.

Characteristic	General population* (%)	Full sample (n = 1134)	Analytic sample† (n = 1102)
Age (y), mean ± SD		46.9 ± 18.1	46.7 ± 18.1
Age group (y), n (%)			
18-34	30.5	358 (31.6)	354 (32.1)
35-54	34.5	394 (34.7)	381 (34.6)
>55	34.6	382 (33.7)	367 (33.3)
Range	–	18-99	18-99
Sex, n (%)			
Male	48.3	564 (49.7)	544 (49.3)
Female	51.4	565 (49.8)	553 (50.2)
Other[46]	0.3	5 (0.4)	5 (0.4)
Race, n (%)			
White	65.5	685 (60.4)	679 (61.6)
Black	11.9	152 (13.4)	144 (13.1)
Asian or Pacific Islander	5.3	79 (7.0)	74 (6.7)
American Indian/Alaskan Native	0.5	29 (2.6)	27 (2.5)
Hispanic ethnicity, n (%)	15.0	208 (18.3)	197 (17.9)
Education level greater than secondary, n (%)	58.9	732 (64.6)	718 (65.2)
Marital status, n (%)			
Not married	48.0	556 (49.0)	537 (48.7)
Married or common-law	–	322 (28.4)	317 (28.8)
Separated or divorced	–	180 (15.9)	174 (15.8)
Widowed	–	75 (6.6)	73 (6.6)
Child dependents, n (%)			
None	71.2	916 (80.8)	888 (80.6)
Child(ren), ≤5 y old	–	68 (6.0)	68 (6.2)
Child(ren), 6-17 y old	–	180 (15.9)	176 (16.0)
Primary health insurance,[47] n (%)			
None	9	98 (8.6)	93 (8.4)
Public	36	480 (42.3)	458 (41.6)
Private	56	555 (49.1)	550 (50.0)
Location, n (%)			
Midwest	21.3	358 (31.6)	346 (31.4)
Northeast	17.6	129 (11.4)	125 (11.3)
South	37.6	353 (31.1)	346 (31.4)
West	23.6	294 (25.9)	285 (25.9)
Country of birth, United States, n (%)		983 (86.7)	960 (87.1)
History of illness,[45] n (%)			
Hypertension	32.0	270 (23.8)	258 (23.4)
Arthritis	22.7	267 (23.5)	258 (23.4)
Diabetes	9.4	111 (9.8)	104 (9.4)
Heart failure	2.2	20 (1.8)	18 (1.6)
Stroke	1.8-2.4	23 (2.0)	22 (2.0)
Bronchitis	3.6	29 (2.6)	25 (2.3)
Asthma	7.5	132 (11.6)	130 (11.8)
Depression	25.7	295 (26.0)	285 (25.9)
Migraine	16.0	164 (14.5)	159 (14.4)
Cancer	5.9	65 (5.7)	64 (5.8)
None	–	372 (32.8)	364 (33.3)
Health status,[44] n (%)			
Excellent/very good/good	85.6	980 (86.4)	955 (86.7)
Fair/poor	14.4	154 (13.5)	146 (13.3)
Self-reported EQ-VAS			
Mean ± SD		80.4 ± 15.6	80.4 ± 15.6
Median (IQR)		85 (15)	85 (15)
Mobility, n (%)			
No problems		812 (71.6)	790 (71.7)
Slight problems		208 (18.3)	204 (18.5)
Some/moderate problems		79 (7.0)	77 (7.0)
Severe problems		31 (2.7)	28 (2.5)
Unable to walk about		4 (0.4)	3 (0.3)

continued on next page

Table 1. Continued

Characteristic	General population* (%)	Full sample (n = 1134)	Analytic sample† (n = 1102)
Self-care, n (%)			
No problems		1060 (93.5)	1032 (93.6)
Slight problems		42 (3.7)	41 (3.7)
Some/moderate problems		25 (2.2)	25 (2.3)
Severe problems		5 (0.44)	3 (0.3)
Unable to wash or dress		2 (0.18)	1 (0.1)
Usual activities, n (%)			
No problems		854 (75.3)	828 (75.1)
Slight problems		178 (15.7)	175 (15.9)
Some/moderate problems		80 (7.1)	79 (7.2)
Severe problems		16 (1.4)	16 (1.4)
Unable to do usual activities		6 (0.5)	4 (0.4)
Pain/discomfort, n (%)			
No pain or discomfort		556 (49.0)	540 (49.0)
Slight pain or discomfort		374 (33.0)	364 (33.0)
Moderate pain or discomfort		151 (13.3)	148 (13.4)
Severe pain or discomfort		39 (3.4)	38 (3.5)
Extreme pain or discomfort		14 (1.2)	12 (1.1)
Anxiety/depression, n (%)			
Not anxious or depressed		699 (61.6)	679 (61.6)
Slightly anxious or depressed		272 (24.0)	265 (24.1)
Moderately anxious or depressed		131 (11.6)	128 (11.6)
Severely anxious or depressed		24 (2.1)	23 (2.1)
Extremely anxious or depressed		8 (0.7)	7 (0.6)

cTTO indicates composite time trade-off; DCE, discrete choice experiment; IQR, interquartile range; SD, standard deviation; VAS, visual analogue scale.
*General population estimates were based on US Census estimates (2011-2015 American Community Survey 5-y estimates) unless otherwise referenced.
†Analytic sample included respondents who provided eligible responses for the cTTO or DCE.

(−0.414), and the smallest was usual activities level 2 (−0.068). Statistically significant differences were observed between levels for each of the 5 dimensions except mobility (levels 2-3), self-care (levels 2-3), and anxiety/depression (levels 4-5).

In applying the preferred model as a scoring algorithm for EQ-5D-5L health state utilities, an index-based summary score is obtained by subtracting parameter estimates for each dimension level of the health state from 1. For example, for the health state 21354, the utility would be 1 − (0.096 + 0 + 0.101 + 0.414 + 0.299) = 0.090 (see Appendix J in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.02.009).

Overall mean scores generated for the EQ-5D-3L using the Shaw et al value set, the crosswalk, and the EQ-5D-5L value set from the present study were 0.369 ± 0.225, 0.416 ± 0.194, and 0.236 ± 0.272, respectively.[14,15] The kernel density plot shows the distribution of EQ-5D-5L values to be unimodal and symmetric, with a wider range of scale (−0.573 to 1) than the EQ-5D-3L value set and crosswalk (−0.109 to 1) (Fig. 3). The EQ-5D-5L value set had 624 health states that were WTD (20.0%), whereas the EQ-5D-3L and crosswalk value sets had 10 (4.1% of 243) and 39 (1.2% of 3125) health states that were WTD, respectively.[14,15] Sensitivity analyses resulted in marginal differences in the model estimates. No significant interviewer effect was observed.

Discussion

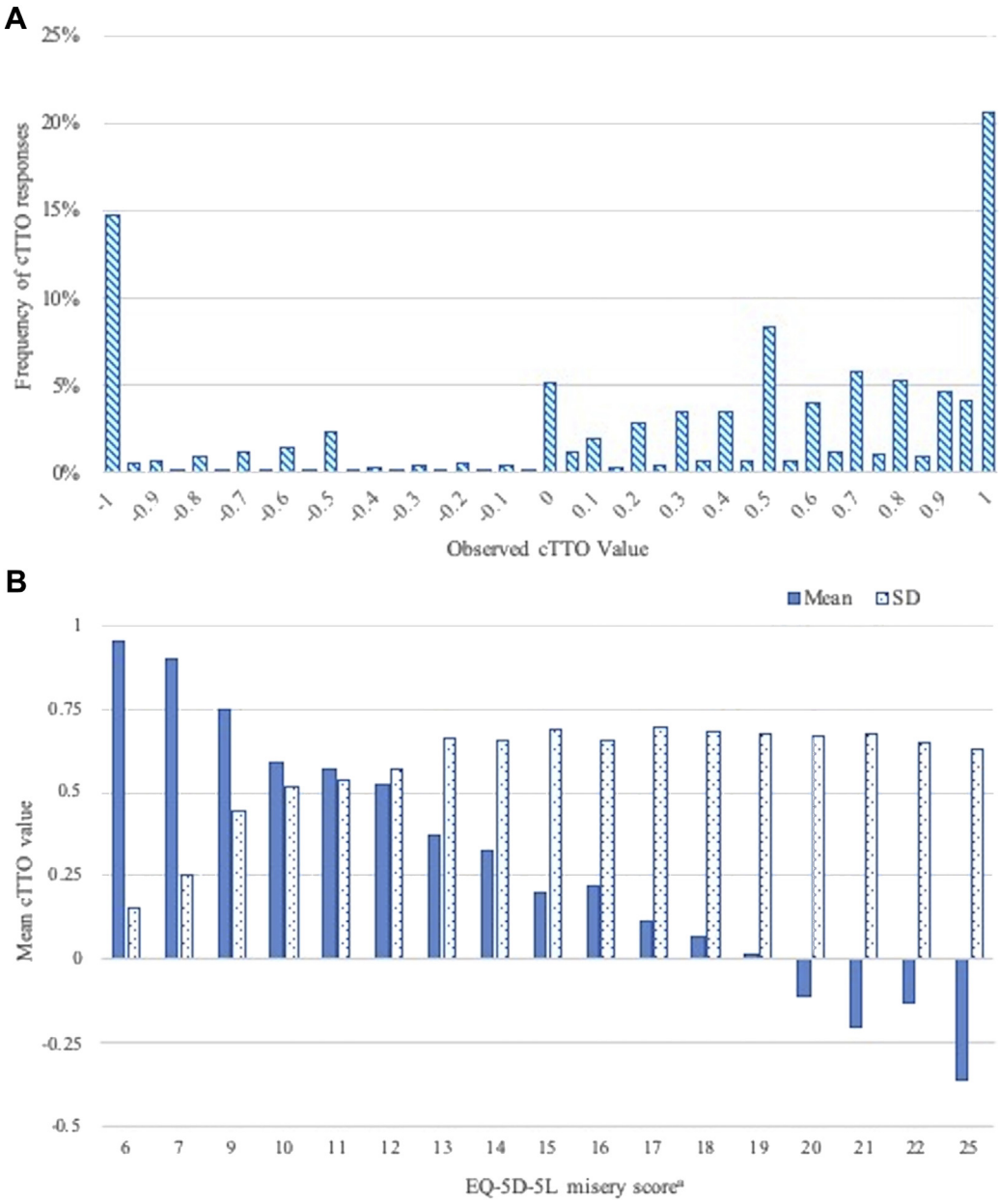
A US-based value set for the EQ-5D-5L was developed on the basis of a standardized international protocol that can help inform economic evaluations and decision making in US health systems as well as facilitate cross-country comparisons of health preferences and cost effectiveness. The value set fulfills criteria outlined by the Second Panel of Cost-Effectiveness in Health and Medicine for generating QALYs for the purposes of cost-utility analyses, providing quality weights that are preference-based, interval-scaled, and sourced from a "community-based" (societal) sample.[48]

The Tobit model (model 1) based on cTTO data was selected as the preferred model for the value set. Although there is no criterion standard for modeling preference data, combining data from different elicitation techniques such as the cTTO and DCE required assumptions of statistical and conceptual appropriateness that were unnecessary in this study; the Tobit model using cTTO-only data provided a value set that fulfilled our criteria. In addition, we were able to account for within-subject dependency of the observations by using a respondent-level random effect. The hybrid model is, however, a pragmatic compromise to combine preferences from methods with different shortcomings, for example, scale compatibility and loss aversion[49] for the cTTO, attribute nonattendance and lexicographic preferences for the DCE,[50–52] and would fully use the collected preference data.

The predicted value for 55 555 using the Tobit model (model 1) was −0.573 and the mean observed cTTO value was −0.366; this difference was due to both the extrapolation of predicted values (ie, censoring at −1) and accounting for heteroscedasticity of the variance. During interviews, respondents often expressed the desire to trade beyond the cTTO value of −1; furthermore, homoscedasticity of the error was rejected, indicating the presence of heteroscedasticity. Thus, these modeling aspects should be included in the final value set. A similar pattern was seen in other countries that have also used the Tobit model with censoring at −1 and/or heteroscedasticity, which include value sets in Germany[53] and Indonesia.[54]

Although an evaluation of the implications of transitioning from the EQ-5D-3L to the EQ-5D-5L is beyond the scope of this study, a few differences between the US EQ-5D-3L and the US EQ-

Figure 2. Distribution of cTTO observations by (A) value and (B) health state severity. Misery score is calculated by summing the severity levels across all 5 dimensions; for example, the misery score for health state 21354 would be 15 (2+ 1 + 3 + 5 + 4).



cTTO indicates composite time trade-off.

5D-5L valuation studies are known.[14] Although both value sets were based on TTO tasks, the WTD elicitation tasks differed between studies (EQ-5D-3L: conventional TTO, scaled to −1 to 1 QALY scale; EQ-5D-5L: lead-time TTO).[32] Furthermore, the statistical models differed in specification, that is, presence of interaction terms, heteroscedasticity of variance, and censoring. The TTO technique and modeling variations may both contribute to a greater range of scale for the EQ-5D-5L: −0.573 to 1 (EQ-5D-5L) and −0.109 to 1 (EQ-5D-3L).

This study has several strengths, particularly the implementation of a standardized protocol on the basis of evidence generated by researchers in and outside the EuroQol Group.[16,21] The standardized protocol ensured similar data collection methods, interviewer training, and data quality control. To tailor the protocol and EQ-VT platform to US-specific factors, scientific and stakeholder advisors were consulted to guide study design and recruitment strategies and inform user uptake. Another study recently sought to estimate preferences for EQ-5D-5L health states for the United States[55] on the basis of DCE with duration data from a EuroQol-sponsored experimental modeling exercise.[56] It was not based on the standard protocol and data quality control processes developed by the EuroQol Group, and used exclusively online panel respondents, which may have data quality and generalizability issues.[57] Future

Table 2. Parameter estimates for main-effects models.

Dimension/level	Model 1: cTTO (Tobit with heteroscedasticity, censored at −1, RE) (preferred model)			Model 2: DCE (mixed logit, rescaled to censored cTTO mean values)			Model 3: hybrid (Tobit with heteroscedasticity, censored at −1, conditional logit)		
	Estimate	SE	*P* value	Estimate	SE	*P* value	Estimate	SE	*P* value
MO2	−0.096	0.015	<.0001	−0.092	0.011	<.0001	−0.077	0.009	<.0001
MO3	−0.122	0.016	<.0001	−0.090	0.015	<.0001	−0.102	0.012	<.0001
MO4	−0.237	0.018	<.0001	−0.232	0.016	<.0001	−0.247	0.012	<.0001
MO5	−0.322	0.016	<.0001	−0.331	0.021	<.0001	−0.364	0.012	<.0001
SC2	−0.089	0.014	<.0001	−0.079	0.011	<.0001	−0.068	0.009	<.0001
SC3	−0.107	0.017	<.0001	−0.071	0.013	<.0001	−0.08	0.012	<.0001
SC4	−0.220	0.018	<.0001	−0.251	0.019	<.0001	−0.225	0.012	<.0001
SC5	−0.261	0.016	<.0001	−0.299	0.023	<.0001	−0.288	0.011	<.0001
UA2	−0.068	0.015	<.0001	−0.044	0.011	<.0001	−0.051	0.009	<.0001
UA3	−0.101	0.016	<.0001	−0.055	0.013	<.0001	−0.068	0.011	<.0001
UA4	−0.255	0.013	<.0001	−0.166	0.016	<.0001	−0.205	0.012	<.0001
UA5	−0.255	0.013	<.0001	−0.207	0.016	<.0001	−0.236	0.011	<.0001
PD2	−0.060	0.013	<.0001	−0.094	0.013	<.0001	−0.065	0.008	<.0001
PD3	−0.098	0.017	<.0001	−0.151	0.017	<.0001	−0.108	0.012	<.0001
PD4	−0.318	0.015	<.0001	−0.393	0.027	<.0001	−0.367	0.013	<.0001
PD5	−0.414	0.017	<.0001	−0.399	0.026	<.0001	−0.441	0.013	<.0001
AD2	−0.057	0.014	<.0001	−0.076	0.015	<.0001	−0.057	0.008	<.0001
AD3	−0.123	0.018	<.0001	−0.150	0.018	<.0001	−0.133	0.012	<.0001
AD4	−0.299	0.016	<.0001	−0.310	0.025	<.0001	−0.329	0.012	<.0001
AD5	−0.321	0.015	<.0001	−0.369	0.027	<.0001	−0.371	0.012	<.0001

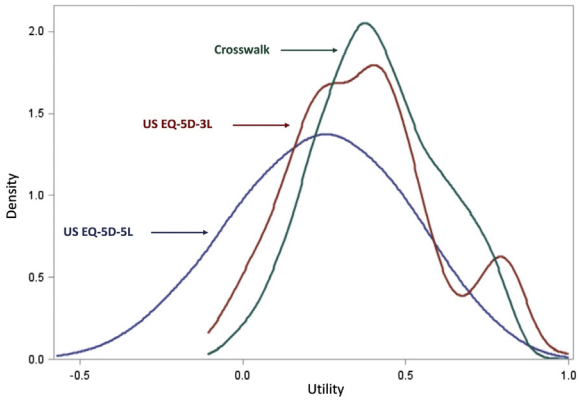
Dimension ranking	PD-MO-AD-SC-UA	PD-AD-MO-SC-UA	PD-AD-MO-SC-UA
Estimated utility values by health state			
21 111	0.904	0.908	0.923
12 111	0.911	0.921	0.932
11 211	0.932	0.956	0.949
11 121	0.940	0.906	0.935
11 112	0.943	0.924	0.943
55 555	−0.573	−0.605	−0.699
No. of health states WTD, n (%)	624 (20.0)	669 (21.4)	733 (23.5)

AD indicates anxiety/depression; cTTO, composite time trade-off; DCE, discrete choice experiment; MO, mobility; PD, pain/discomfort; RE, random effects; SC, self-care; SE, standard error; UA, usual activities; WTD, worse than dead.

studies comparing the scoring approaches may be inevitable, but initial indicators such as the range of the values suggest that the EQ-5D-5L value set in the present study has a wider range of scale and thus may generate larger QALY gains. Furthermore, the EuroQol Group has issued guidance on country-specific value sets used by users for health technology assessment–related decision making and promotes the use of value set–based data collected using the standardized international protocol, such as the present value set.[21]

This study also has several potential limitations. Alternative model specifications were explored to only a limited extent, because the experimental design of the EQ-VT protocol was optimized for a main-effects model. Law et al[58] explored the use of N45 terms on the cTTO data and found minimal model fit improvements with the inclusion of interaction terms, suggesting that a main-effects model would be best for parsimony and interpretation. Also, generalizability of study respondents to the US general population may be limited in that data collection was conducted in metropolitan areas in the summer. Non–quota-sampled characteristics of respondents do closely match reported estimates (Table 1), lending confidence to representativeness of the respondents. A seasonal effect has been detected in self-reported health[59]; nevertheless, it is unclear whether stated preferences for hypothetical health states, as in this study, are impacted by season. Quota sampling was based on the US Census, which does not inquire about

Figure 3. Kernel density plot of US EQ-5D value sets.



EQ-5D-3L/5L indicates 3-/5-level EuroQol 5-dimensional questionnaire.

citizenship status; thus, the quota sampling approach pertains to US residents, not just citizens.

Conclusions

A US value set for the EQ-5D-5L is an important resource to support economic evaluations of healthcare interventions. The wider range of scale and improved measurement properties associated with the 5-level descriptive system, such as greater sensitivity to changes in health status, are likely to translate into differences in QALY calculations compared with the EQ-5D-3L[14] and the crosswalk.[15] In addition, use of a standard protocol in the present study will facilitate comparisons with international value sets for the EQ-5D-5L.[21] Studies that compare the EQ-5D-5L value set with the EQ-5D-3L value set such as Law et al[58] and other utility measures will assist in understanding the implications of this value set for estimating QALYs and the adoption of health technology. The full value set can be retrieved by contacting the corresponding author or the EuroQol Group.

Acknowledgments

We are grateful to the following individuals for their guidance and assistance in the study preparation, data collection, analysis, and/or interpretation of results: Jennifer Graff (National Pharmaceutical Council), Juan Manuel Ramos-Goñi (Axentiva Solutions), Kristina Ludwig (EuroQol), Elly Stolk (EuroQol), Fredrick Purba (Erasmus-Padjadjaran), Alan Schwartz (University of Illinois at Chicago), Kim Rand (University of Oslo), and Ben van Hout (Sheffield University); University of Illinois at Chicago interviewers: Ashley Cha, Michelle Cho, Rachel Harrington, Thomas Hopkins, Jonathan Nazari, Christopher Saffore, Dharmi Shah, and Connie Yan; the International Society of Pharmacoeconomics and Outcomes Research student chapters for their assistance with on-site data collection: Universities of Washington, Thomas Jefferson, Colorado, and Illinois at Chicago. Further, we want to thank the study participants for providing their time and views on health-related quality of life and the staff at interview locations for their assistance; without them, the study would not have been possible.

Source of Financial Support

This study was sponsored in part by research grants from the EuroQol Research Foundation and Bristol-Myers Squibb. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and writing and publishing the report. R Jiang was supported by the University of Illinois at Chicago/Takeda Fellowship in Health Economics and Outcomes Research; JW Shaw is an employee and shareholder of Bristol-Myers Squibb; KS Boye is an employee and minor shareholder of Eli Lilly and Company; AS Pickard, F Xie, M Oppe, JW Shaw, KS Boye, E Pullenayegum, and JJV Busschbach are members of the EuroQol Group.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2019.02.009.

REFERENCES

1. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med*. 2001;33(5):337–343.
2. 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.
3. Janssen MF, Pickard AS, Golicki D, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res*. 2013;22(7):1717–1727.
4. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics*. 2018;36(6):675–697.
5. Devlin N, Brazier J, Pickard AS, et al. 3L, 5L, What the L? A NICE conundrum. *Pharmacoeconomics*. 2018;36(6):637–640.
6. Schnipper LE, Davidson NE, Wollins DS, et al. American Society of Clinical Oncology statement: a conceptual framework to assess the value of cancer treatment options. *J Clin Oncol*. 2015;33(23):2563–2577.
7. Pizzi LT. The Institute for Clinical and Economic Review and its growing influence on the US healthcare. *Am Health Drug Benefits*. 2016;9(1):9–10.
8. Anderson JL, Heidenreich PA, Barnett PG, et al. ACC/AHA statement on cost/value methodology in clinical practice guidelines and performance measures. *Circulation*. 2014;129(22):2329–2345.
9. Westrich K. *Current Landscape: Value Assessment Frameworks*. Washington, DC: National Pharmaceutical Council; 2016.
10. Neumann PJ, Weinstein MC. Legislating against use of cost-effectiveness information. *N Engl J Med*. 2010;363(16):1495–1497.
11. AMCP Format Executive Committee. *The AMCP format for formulary submissions: a format for submission of clinical and economic evidence of pharmaceuticals in support of formulary consideration. Version 4.0. April 2016*. http://www.amcp.org/FormatV4/. Accessed May 17, 2019.
12. Neumann PJ, Cohen JT. Measuring the value of prescription drugs. *N Engl J Med*. 2015;373(27):2595–2597.
13. *ICER. Overview of the ICER value framework and proposals for an update for 2017-2018*. 2017. https://icer-review.org/wp-content/uploads/2016/02/ICER-VAF-Update-Proposals-020117.pdf. Accessed May 19, 2019.
14. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43(3):203–220.
15. van Hout B, Janssen M, Feng Y-S, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15(5):708–715.
16. Krabbe PF, Devlin NJ, Stolk EA, et al. Multinational evidence of the applicability and robustness of discrete choice modeling for deriving EQ-5D-5L health-state values. *Med Care*. 2014;52(11):935–943.
17. Krabbe PF, Stolk EA, Devlin NJ, et al. Head-to-head comparison of health-state values derived by a probabilistic choice model and scores on a visual analogue scale. *Eur J Health Econ*. 2017;18(8):967–977.
18. Ramos-Goñi JM, Oppe M, Slaap B, et al. Quality control process for EQ-5D-5L valuation studies. *Value Health*. 2017;20(3):466–473.
19. Stolk E, Ludwig K, Rand K, et al. 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.
20. Oppe M, van Hout B. *The "power" of eliciting EQ-5D-5L values: the experimental design of the EQ-VT*. Rotterdam, The Netherlands: EuroQol Working Paper Series, EuroQol Research Foundation; 2017.
21. EuroQol. EQ-5D 5L | Valuation: Standard value sets. 2018. https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/valuation-standard-value-sets/. Accessed January 14, 2019.
22. Oppe M, Devlin NJ, van Hout B, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health*. 2014;17(4):445–453.
23. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: past, present and future. *Appl Health Econ Health Policy*. 2017;15(2):127–137.
24. 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.
25. Devlin N, Shah K, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ*. 2018;27(1):7–22.
26. Ikeda T, Shiraiwa K, Igarashi M, et al. *The development of the scoring method in the Japanese version of EQ-5D-5L*. 2016. https://www.niph.go.jp/journal/data/64-1/201564010008.pdf. Accessed May 30, 2018.

27. Kim SH, Ahn J, Ock M, et al. The EQ-5D-5L valuation study in Korea. *Qual Life Res*. 2016;25(7):1845–1852.
28. 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.
29. Yang Z, van Busschbach J, Timman R, et al. Logical inconsistencies in time trade-off valuation of EQ-5D-5L health states: whose fault is it? *PLoS One*. 2017;12(9):e0184883.
30. Xie F, Pickard AS, Krabbe PF, et al. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). *Pharmacoeconomics*. 2015;33(8):867–877.
31. Attema AE, Edelaar-Peeters Y, Versteegh MM, et al. Time trade-off: one methodology, different methods. *Eur J Health Econ*. 2013;14(suppl 1):S53–S64.
32. Janssen BM, Oppe M, Versteegh MM, et al. Introducing the composite time trade-off: a test of feasibility and face validity. *Eur J Health Econ*. 2013;14(1):5–13.
33. Tilling C, Devlin N, Tsuchiya A, et al. Protocols for time tradeoff valuations of health states worse than dead: a literature review. *Med Decis Making*. 2010;30(5):610–619.
34. Bernardo R. 2016's metro areas that most and least resemble the U.S. WalletHub. 2016. https://wallethub.com/edu/metro-areas-that-most-and-least-resemble-the-us/6109/. Accessed January 11, 2017.
35. Kolko J. *"Normal America" is not a small town of white people. FiveThirtyEight*. 2016. https://fivethirtyeight.com/features/normal-america-is-not-a-small-town-of-white-people/. Accessed January 11, 2017.
36. Iceland J, Weinberg DH. *Racial and Ethnic Residential Segregation in the United States 1980-2000*. Washington, DC: US Census Bureau; 2002. https://www.census.gov/prod/2002pubs/censr-3.pdf. Accessed January 17, 2019.
37. Purba FD, Hunfeld JA, Iskandarsyah A, et al. Employing quality control and feedback to the EQ-5D-5L valuation protocol to improve the quality of data collection. *Qual Life Res*. 2017;26(5):1197–1208.
38. van Reenen M, Janssen B. *EQ-5D-5L User Guide—Basic Information on How to Use the EQ-5D-5 L Instrument*. Rotterdam, The Netherlands: EuroQol Group; 2013. https://euroqol.org/wp-content/uploads/2016/09/EQ-5D-5L_UserGuide_2015.pdf. Accessed January 17, 2019.
39. Ramos-Goni JM, Pinto-Prades JL, Oppe M, et al. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Med Care*. 2017;55(7):e51–e58.
40. Ramos-Goni JM, Craig BM, Oppe M, et al. *Combining continuous and dichotomous responses in a hybrid model. EuroQol Working Paper Series*. Rotterdam, The Netherlands: EuroQol Research Foundation; 2016. https://euroqol.org/wp-content/uploads/working_paper_series/EuroQol_Working_Paper_Series_Manuscript_16002_-_Juan_Ramos-Goni.pdf. Accessed January 30, 2019.
41. US Census Bureau. *2015 American Community Survey 1-Year Estimates*. Washington, DC: US Census Bureau; 2015. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk. Accessed January 30, 2018.
42. Humes KR, Jones NA, Ramirez RR. *Overview of Race and Hispanic Origin: 2010*. Washington, DC: US Department of Commerce, Economics and Statistics Administration, US Census Bureau; 2011. https://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf. Accessed January 30, 2019.
43. Hobbs F, Stoops N. *Demographic Trends in the 20th Century. Census 2000 Special Reports, Series CENSR-4*. Washington, DC: US Census Bureau; 2002. https://www.census.gov/prod/2002pubs/censr-4.pdf. Accessed January 30, 2019.
44. Li C, Balluz LS, Okoro CA, et al. Surveillance of certain health behaviors and conditions among states and selected local areas—Behavioral Risk Factor Surveillance System, United States, 2009. *MMWR Surveill Summ*. 2011;60(9):1–249.
45. CDC. Data and Statistics. 2018. https://www.cdc.gov/datastatistics/index.html. Accessed January 30, 2018.
46. Gates WJ. *How Many People Are Lesbian, Gay, Bisexual and Transgender?*. 2011. https://williamsinstitute.law.ucla.edu/research/census-lgbt-demographics-studies/how-many-people-are-lesbian-gay-bisexual-and-transgender/. Accessed January 17, 2019.
47. Garfield R, Damico A, Stephens J, et al. *The Coverage Gap: Uninsured Poor Adults in States That Do Not Expand Medicaid—An Update*. Menlo Park, CA: Kaiser Family Foundation; 2016.
48. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second Panel on Cost-Effectiveness in Health And Medicine. *JAMA*. 2016;316(10):1093–1103.
49. Bleichrodt H. A new explanation for the difference between time trade-off utilities and standard gamble utilities. *Health Econ*. 2002;11(5):447–456.
50. Ryan M, Watson V, Entwistle V. Rationalising the "irrational": a think aloud study of discrete choice experiment responses. *Health Econ*. 2009;18(3):321–336.
51. Lagarde M. Investigating attribute non-attendance and its consequences in choice experiments with latent class models. *Health Econ*. 2013;22(5):554–567.
52. Lancsar E, Louviere J. Deleting "irrational" responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Econ*. 2006;15(8):797–811.
53. Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36(6):663–674.
54. Purba FD, Hunfeld JA, Iskandarsyah A, et al. The Indonesian EQ-5D-5L value set. *Pharmacoeconomics*. 2017;35(11):1153–1165.
55. Craig BM, Rand K. Choice defines QALYs: a US valuation of the EQ-5D-5L. *Med Care*. 2018;56(6):529–536.
56. 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.
57. Craig BM, Hays RD, Pickard AS, et al. Comparison of US panel vendors for online surveys. *J Med Internet Res*. 2013;15(11):e260.
58. Law EH, Pickard AS, Xie F, et al. Parallel valuation: a direct comparison of EQ-5D-3L and EQ-5D-5L societal value sets. *Med Decis Making*. 2018;38(8):968–982.
59. Jia H, Lubetkin EI. Time trends and seasonal patterns of health-related quality of life among U.S. adults. *Public Health Rep*. 2009;124(5):692–701.