VALUE IN HEALTH 20 (2017) 662–669

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Estimating an EQ-5D-5L Value Set for China

Nan Luo, PhD[1], *Gordon Liu, PhD*[2,3,*], *Minghui Li, MSc*[4], *Haijing Guan, MSc*[2], *Xuejing Jin, MSc*[2,5], *Kim Rand-Hendriksen, PhD*[6,7]

[1]*Saw Swee Hock School of Public Health, National University of Singapore, Singapore;* [2]*China Center for Health Economic Research, Peking University, Beijing, China;* [3]*National School of Development, Peking University, Beijing, China;* [4]*Department of Clinical Pharmacy and Outcomes Sciences, University of South Carolina, Columbia, SC, USA;* [5]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada;* [6]*Health Services Research Centre, Akershus University Hospital, Lørenskog, Norway;* [7]*Department of Health Management and Health Economics, University of Oslo, Oslo, Norway*

ABSTRACT

Objectives: To estimate a five-level EuroQol five-dimensional questionnaire (EQ-5D-5L) value set for China using the health preferences of residents living in the urban areas of the country. **Methods:** The values of a subset of the EQ-5D-5L–defined health states (n = 86) were elicited using the time trade-off (TTO) technique from a sample of urban residents (n = 1271) recruited from five Chinese cities. In computer-assisted personal interviews, participants each completed 10 TTO tasks. Two additive and two multiplicative regression models were evaluated for their performance in describing the relationship between TTO values and health state characteristics using a cross-validation approach. Final values were generated using the best-performed model and a rescaling method. **Results:** The 8- and 9-parameter multiplicative models unanimously outperformed the 20-parameter additive model using a random or fixed intercept in predicting values for out-of-sample health states in the cross-validation analysis and their coefficients were estimated with lower standard errors. The prediction accuracies of the two multiplicative models measured by the mean absolute error and the intraclass correlation coefficient were very similar, thus favoring the more parsimonious model. **Conclusions:** The 8-parameter multiplicative model performed the best in the study and therefore was used to generate the EQ-5D-5L value set for China. We recommend using rescaled values whereby 1 represents the value of instrument-defined full health in economic evaluation of health technologies in China whenever the EQ-5D-5L data are available.

Keywords: cross-validation, EQ-5D-5L, modeling, time trade-off.

Introduction

Preference-based instruments are widely used to estimate quality-adjusted life-years (QALYs) in cost-utility analysis of new health programs or technologies [1]. Consisting of a descriptive system and predetermined utility values, preference-based instruments provide an alternative approach to generating utility values through direct valuation.

The EuroQol Group's three-level EuroQol five-dimensional questionnaire (EQ-5D-3L) is the most frequently used preference-based instrument worldwide [2]. The descriptive system of the EQ-5D comprises five dimensions: mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD), and anxiety/depression (AD); each dimension is described at three levels (roughly corresponding to no problems, moderate problems, and extreme problems). A large number of national value sets, each consisting of the utility values of the 243 EQ-5D-3L health states to the general population, have been generated to provide health technology appraisers with the most relevant quality-of-life weights for calculation of QALYs [3].

Recently, the EQ-5D has been expanded, such that each dimension is described at five levels, corresponding roughly to no, slight, moderate, severe, and extreme problems. The new version is referred to as the EQ-5D-5L [4]. Several national value sets for the EQ-5D-5L have been published [5–10], and more are on the way.

The purpose of this study was to estimate the EQ-5D-5L value set on the basis of health preferences of urban residents of China. The Chinese version of the EQ-5D-5L description system for China has demonstrated some advantages over the EQ-5D-3L [11–13].

Methods

The study was part of a multinational research project coordinated by the EuroQol Research Foundation. Using a cross-sectional survey designed according to a research protocol and a computerized interview program that were developed by the EuroQol Group [14], we collected the preferences data needed for estimating an EQ-5D-5L value set from a sample of the general urban population in China. The detailed study design is described here.

Sampling and Recruitment

A sample of the general population living in urban China was recruited. Because of limited resources, a nonprobability sampling method was used to recruit community-dwelling residents from

* *Address correspondence to:* Gordon Liu, National School of Development, Peking University, 5 Yiheyuan Road, Beijing 100871, China. E-mail: gordonliu@nsd.pku.edu.cn.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2016.11.016

five metropolitan areas, namely, Beijing, Shenyang, Nanjing, Chengdu, and Guiyang. These areas were selected as representatives of cities of varying population size, geographical region, and economic development. In each area, quotas were set to recruit 250 participants whose characteristics in age, sex, and educational attainment resembled those of the target population, and participants were recruited through personal contact by a local data collection team from five or more different locations. Recruiting was conducted in publicly accessible places such as streets, shopping centers, parks, and university campuses, and in places with restricted access such as residential areas, schools, and factories. Inclusion criteria were that participants 1) were 18 years or older, 2) were literate and able to read text from a computer screen, 3) were able to understand survey questions, and 4) gave informed consent.

Data Collection Procedures

All consenting participants were invited to a face-to-face, one-on-one computer-assisted personal interview. A total of 20 graduate students and junior lecturers from five local universities, one university in each city, conducted the interviews. All the interviewers participated in a half-day training workshop on the study design, interview protocol, and computer software program designed for the study and on recruitment and interview skills. All interviews were conducted using a laptop computer for displaying questions and recording responses. During the data collection period, completed interviews were uploaded and analyzed on a daily basis. Whenever very short interviews or unusual response patterns were identified, the interviewers were contacted for clarification and retraining if necessary.

Interview

All interviews were conducted using the EuroQol Valuation Technology (version 1.0) program [14]. The interview had four sections. The first section familiarized participants with the EQ-5D-5L descriptive system by asking them to describe their own health using the EQ-5D-5L questionnaire. The second section contained 10 time trade-off (TTO) tasks, each for valuing a different EQ-5D-5L health state. The third section was designed to value selected EQ-5D-5L health states using discrete choice experiment (data not used in the present study). The last section assessed participants' socioeconomic characteristics.

The TTO tasks used a "composite" TTO technique whereby "better than dead" and "worse than dead" health states were valued by conventional TTO and lead-time TTO, respectively [14,15]. The composite TTO was described in detail elsewhere [14,16]. Briefly, for an impaired health state considered better than dead, the task is to elicit the x (0–10) value at which a respondent is indifferent between two alternatives: 1) living in full health for x years followed by death and 2) living in the impaired health state for 10 years followed by death. The utility value of the impaired health state is $x/10$. For health states considered worse than dead, the two alternatives are 1) living in full health for x years followed by death and 2) living in full health for 10 years and then in the impaired health state for another 10 years before death. The utility value is given by $(x-10)/10$. The values were bounded at -1 and 1, with 0 corresponding to death. The state of "in a wheelchair" was used as an example to make sure participants understood the concept of composite TTO before proceeding to the 10 formal valuation tasks.

Health States

By convention, the EQ-5D-5L health states are presented in a short form using five-digit numbers in which the digits represent the levels of functioning for the dimensions in order of presentation (MO, SC, UA, PD, and AD). Thus, state 11111 represents no problems on any dimension, whereas state 55555 represents extreme problems on all five dimensions.

In this study, 86 EQ-5D-5L health states were valued using the composite TTO technique, including the 5 mildest imperfect health states (i.e., 21111, 12111, 11211, 11121, and 11112), state 55555, and 80 other states of varying severity. The 86 health states were grouped into 10 health state blocks, all of which contained 1 mildest health state, state 55555, and 8 block-unique health states. Each participant was randomly assigned a health state block for TTO valuation.

Data Analysis

The aim was to determine a regression model on the basis of the observed values for the 86 health states, which would then be used to generate values for all the 3125 health states defined by the EQ-5D-5L system. We elected to predict utility values on the basis of only health state characteristics. Our data analysis involved four stages: model construction, model evaluation, model estimation, and value adjustment.

Model construction

Four core regression models were tested on the basis of their performance in a recent study of regression models for the EQ-5D-5L (K. Rand-Hendriksen, unpublished data). The standard, additive 20-parameter model, referred to as ADD20, has parameters representing levels 2, 3, 4, and 5 for each dimension. Let α represent the intercept, x_{dl}, the dummy variable indicating the presence of problems on dimension d at level l and β_{dl} the coefficient representing the estimated disutility of having problems on dimension d at level l (e.g., β_{MO3} representing the disutility of having moderate problems on mobility). The mathematical function of ADD20 is as follows:

$$y=\alpha+\Sigma_l\Sigma_d\beta_{dl}x_{dl}+e=\alpha+$$
$$\beta_{MO2}x_{MO2}+\beta_{SC2}x_{SC2}+\beta_{UA2}x_{UA2}+\beta_{PD2}x_{PD2}+\beta_{AD2}x_{AD2}+$$
$$\beta_{MO3}x_{MO3}+\beta_{SC3}x_{SC3}+\beta_{UA3}x_{UA3}+\beta_{PD3}x_{PD3}+\beta_{AD3}x_{AD3}+$$
$$\beta_{MO4}x_{MO4}+\beta_{SC4}x_{SC4}+\beta_{UA4}x_{UA4}+\beta_{PD4}x_{PD4}+\beta_{AD4}x_{AD4}+$$
$$\beta_{MO5}x_{MO5}+\beta_{SC5}x_{SC5}+\beta_{UA5}x_{UA5}+\beta_{PD5}x_{PD5}+\beta_{AD5}x_{AD5}+e$$

The second model was an 8-parameter multiplicative model, hereafter referred to as MULT8. This is a constrained variant of ADD20, in which five parameters representing the disutility of having problems at level 5 on each of the five dimensions (β_{MO}, β_{SC}, β_{UA}, β_{PD}, and β_{AD}) are multiplied by parameters for levels 2, 3, and 4 (L_2, L_3, and L_4). Thus, the disutility of having moderate problems on mobility is $\beta_{MO}\times L_3$. The mathematical function of MULT8 is as follows (note that x_{dl} still represents the dummy variable representing the presence of problems on dimension d at level l):

$$y=\alpha+\sum_l\left(\sum_d\beta_d x_{dl}\right)L_l+e=\alpha+$$
$$\left(\beta_{MO}x_{MO2}+\beta_{SC}x_{SC2}+\beta_{UA}x_{UA2}+\beta_{PD}x_{PD2}+\beta_{AD}x_{AD2}\right)L_2+$$
$$\left(\beta_{MO}x_{MO3}+\beta_{SC}x_{SC3}+\beta_{UA}x_{UA3}+\beta_{PD}x_{PD3}+\beta_{AD}x_{AD3}\right)L_3+$$
$$\left(\beta_{MO}x_{MO4}+\beta_{SC}x_{SC4}+\beta_{UA}x_{UA4}+\beta_{PD}x_{PD4}+\beta_{AD}x_{AD4}\right)L_4+$$
$$\beta_{MO}x_{MO5}+\beta_{SC}x_{SC5}+\beta_{UA}x_{UA5}+\beta_{PD}x_{PD5}+\beta_{AD}x_{AD5}+e$$

The third model, referred to as MULT9, extends MULT8, with an additional parameter L_5 to distinguish level 5 for PD and AD (described using the label "extreme") from level 5 for MO, SC, and UA (described using the label "unable to"). Thus, MULT9 assumes that the relative distance between the levels is shared across dimensions, with the exception of the distance between levels 4 and 5, which is shared across the first three and the last two dimensions only. The mathematical function of MULT9 is as follows:

$$y=\alpha+$$
$$\left(\beta_{MO}x_{MO2}+\beta_{SC}x_{SC2}+\beta_{UA}x_{UA2}+\beta_{PD}x_{PD2}+\beta_{AD}x_{AD2}\right)L_2+$$

$$(\beta_{MO}x_{MO3}+\beta_{SC}x_{SC3}+\beta_{UA}x_{UA3}+\beta_{PD}x_{PD3}+\beta_{AD}x_{AD3})L_3+$$
$$(\beta_{MO}x_{MO4}+\beta_{SC}x_{SC4}+\beta_{UA}x_{UA4}+\beta_{PD}x_{PD4}+\beta_{AD}x_{AD4})L_4+$$
$$(\beta_{MO}x_{MO5}+\beta_{SC}x_{SC5}+\beta_{UA}x_{UA5})+(\beta_{PD}x_{PD5}+\beta_{AD}x_{AD5})L_5+e$$

ADD20 was estimated using ordinary least squares with a fixed constant/intercept (ADD20f), and using a random intercept at the level of individual study participants (ADD20r). ADD20r corresponds to the method described as *random effects* or *generalized least squares* in several EQ-5D valuation studies [17]. On the basis of a previous study in which MULT8 or MULT9 performed poorly with a fixed intercept (K. Rand-Hendriksen et al., 2016; unpublished data), only the random intercept variants of these models were estimated. Hence, four core models were evaluated in the analysis, namely, ADD20f, ADD20r, MULT8r, and MULT9r.

Each core model function was also tested with each of the following three sets of interaction terms:

1. The N5 model, an extension of the EQ-5D-3L N3 model [17]: With N5, the core models were expanded with four additional dummy variables labeled N2, N3, N4, and N5, representing the presence of any problems at levels 2 through 5, respectively.
2. An extended D1 model based on the model used to fit the US EQ-5D-3L value set [18]: With D1, the four core models were expanded with terms representing the number of dimensions beyond the first at levels 2 through 5 (i2, i3, i4, and i5); the squares of these terms (i22, i32, i42, and i52) were also included. To avoid exacerbating the multicollinearity of the models, a constant term was maintained, rather than the D1 term, representing the number of dimensions beyond the first (not at level 1) [19].
3. The four core models were expanded using a single parameter taking on values representing the square root of the number of movements away from full health (i.e., a city block-metric).

Model evaluation

The core models and their derivatives were evaluated for identifying the best one by analyzing their predictions. First, model predictions were examined for logical consistency (monotonicity), which means that worse health states should have lower TTO values than objectively better health states. Models predicting logically inconsistent values would be discarded. Second, prediction accuracy was assessed by comparing predicted and observed mean values for health states valued in the study. The mean absolute error (MAE) and intraclass correlation coefficient (ICC) were calculated to assess overall prediction accuracy. Lower MAE and higher ICC values indicated better accuracy. Model parsimony was used as the selection criterion in case two or more models had similar prediction accuracy.

Operationally, model evaluation was conducted in the manner of cross-validation. We repeatedly split the TTO data into two parts, fitted each model to one part of the data, and used the fitted models to predict values over the left-out part of the data. Two cross-validation methods were used. In the first method, all observations for a single health state were split out, all models were fitted to the remaining 85 observed health states, and the value of the left-out health state was predicted using each fitted model. This was repeated for each of the 86 observed health states. In the second method, health states were left out by block, as defined by the EuroQol Valuation Technology program. As in the first method, all models were fitted to the remaining data, and values for the health states in the left-out block were predicted. This was repeated for each of the 10 fixed health state blocks.

Model estimation

After model evaluation, coefficients of the most promising models were estimated using the entire set of data. Because of the nonlinear nature of some of the MULT8r and MULT9r, and the non-normal error distributions observed in TTO data, standard errors (SEs) and 95% confidence intervals (CIs) of model coefficients were estimated using bootstrapping. A total of 10,000 bootstrap samples were drawn at the level of individual study participants, and all models were fitted to each bootstrap sample.

Value adjustment

We applied linear adjustment to model-predicted values using the formula $\text{Value}_{\text{adjusted}} = \text{Value}_{\text{predicted}}/(1 - \text{intercept})$. This additional step was for removing the effect of the nonzero intercept in all models, which leads to a predicted value of less than 1 for full health (11111). Although most previous studies of this kind chose to adjust only the value for 11111 (to 1), we elected to adjust all the values to preserve the relative utility of all the health states.

All analyses were performed using R statistical package version 3.3.0 (R Development Core Team, Vienna, Austria) [20]. ADD20f and ADD20r were fitted using the built-in functions *lm* and *nls*. MULT8r and MULT9r were fitted using a nonlinear mixed-effects function from the *nlme* package [21].

Results

A total of 1332 individuals (response rate 68.6%) were recruited for this study. Among these, 1296 (97.3%) successfully completed the interview. After excluding those who were younger than 18 years at the time of interview (N = 25), 1271 individuals were included in the final analysis. The full sample characteristics are presented in Table 1.

Table 1 – Characteristics of participants (N = 1271).

Characteristics	N	%
Age group (y)		
18–29	313	24.63
30–39	244	19.20
40–49	272	21.40
50–59	220	17.31
60+	222	17.47
Sex		
Female	634	49.88
Male	637	50.12
Education		
Primary or lower	138	10.86
Junior high school	396	31.16
Senior high school	446	35.09
College or higher	291	22.90
Employment status		
Full-time employee	378	29.74
Temporary worker	301	23.68
Individual freelancer	148	11.64
Retired	240	18.88
Student	115	9.05
Unemployed	48	3.78
Other	41	3.22
Residence of origin		
City	749	58.93
County	82	6.45
Township or village	440	34.62
Health insurance		
Urban employee	551	43.35
New rural	289	22.74
Urban residence	339	26.67
Commercial	156	12.27
No	56	4.41
Other	171	13.46

Table 2 – Model fit results in cross-validation tests and for full data set.

Cross-validation method	ADD20f	ADD20r	MULT8r	MULT9r
Mean absolute error				
Method 1: leave-out by state	0.0447	0.0426	**0.0398**	**0.0398**
Method 2: leave-out by block	0.0437	0.0420	**0.0408**	0.0409
Full data set	0.0339	0.0349	0.0370	0.0365
Intraclass correlation				
Method 1: leave-out by state	0.9796	0.9820	0.9831	0.9838
Method 2: leave-out by block	0.9866	0.9880	0.9882	0.9886
Full data set	0.9882	0.9876	0.9854	0.9864
Number of nonmonotonic models*				
Method 1: leave-out by state	0	0	0	0
Method 2: leave-out by block	1	0	0	0
Full data set	0	0	0	0

Note. Boldfaced figures indicate the smallest observed mean absolute error.
ADD20f and ADD20r, additive 20-parameter model with a *fixed/random* constant/intercept; MULT8r and MULT9r, 8- and 9-parameter multiplicative model with a random intercept.
* The total number of models fitted was 86, 10, and 1 for method 1, method 2, and full data set, respectively.

All four main models gave logically consistent (monotonic) predictions in the cross-validation analysis and when fitted on the full set of data, except that ADD20f was not fully monotonic in cross-validation analysis by health state block. In terms of MAE, MULT8r and MULT9r performed the best when the first (leave-out by state) and second (leave-out by block) cross-validation methods were used, respectively (Table 2). ADD20r, MULT8r, and MULT9r displayed very small differences in terms of ICC, whereas ADD20f performed slightly worse.

The three groups of interaction terms were all rejected. The addition of the N5 terms to the four core models resulted in logically inconsistent models with substantially impaired predictive accuracy. The models with D1 terms added were less prone to logical inconsistency than the ones with the N5 terms, but those too substantially impaired predictive accuracy. The addition of the square-root city block-metric term resulted in logically inconsistent coefficients when added to ADD20r, and substantially changed the predicted values when added to ADD20f, MULT8r, and MULT9r, with a widened gap between the full health and the second-best state, and smaller differences between worse states. (Results are available on request.)

The four models were generally similar in predictions (Fig. 1A). Reflecting the larger number of observations per fitted parameter in the nonlinear models, the bootstrap-based SEs for MULT8r and MULT9r were smaller than those for ADD20f and ADD20r (Tables 3 and 4). The constant/intercept term was in the range of 0.121 (MULT8r) to 0.127 (ADD30f and ADD20r). The right axis of Figure 1B illustrates the effect of rescaling to the value of 11111 predicted by MULT8r.

MULT8r and MULT9r performed better than both ADD20f and ADD20r in terms of predictive accuracy. On the basis of the parsimony criterion, MULT8r was chosen as the basis for generating the final EQ-5D-5L value set for China. The parameters (SEs) of the four models estimated using the full sample data are presented in Table 3 and their counterparts for generating the rescaled values are presented in Table 4. A full set of predicted

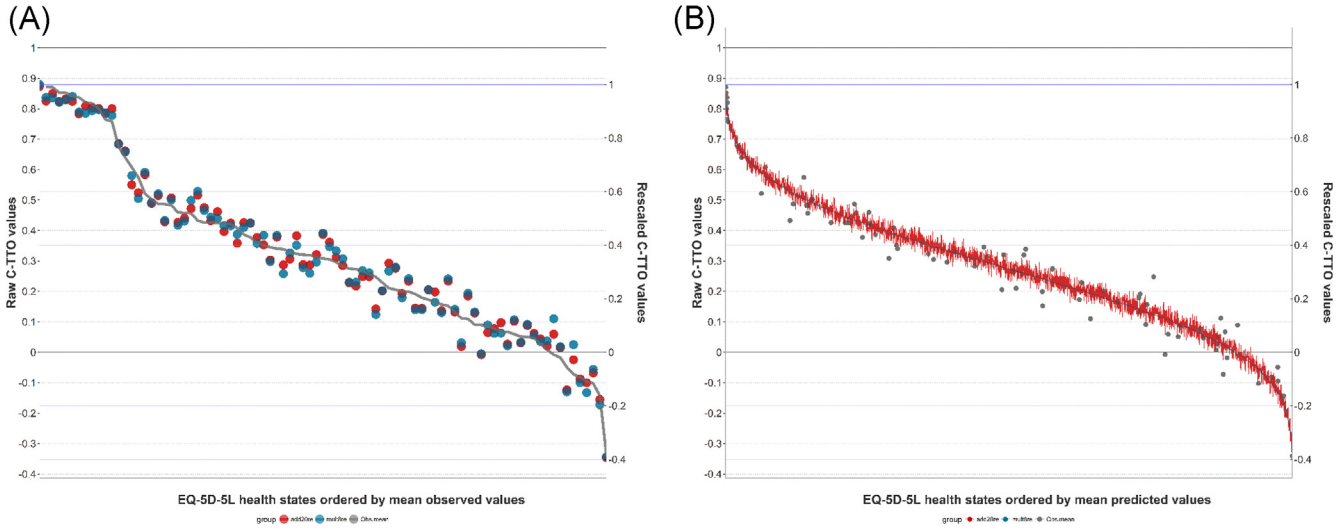

Fig. 1 – Observed mean and model predicted values for the EQ-5D-5L health states. (A) Values for 86 health states ordered from high to low by observed mean values. (B) Values (rescaled) for all 3125 health states ordered from high to low by values predicted by MULT8r. C-TTO, composite time trade-off; EQ-5D-5L, five-level EuroQol five-dimensional questionnaire; MULT8r, 8-parameter multiplicative model with a random intercept.

Table 3 – Estimated coefficients (SEs) of the fitted models on all data.

Model	ADD20f		ADD20r		MULT8r*		MULT9r*	
	Coefficient	SE†	Coefficient	SE†	Coefficient	SE†	Coefficient	SE†
				Nonlinear parameter				
Intercept	–		–		0.121	0.009	0.125	0.010
MO	–		–		0.303	0.010	0.312	0.011
SC	–		–		0.222	0.010	0.224	0.010
UA	–		–		0.205	0.009	0.205	0.009
PD	–		–		0.266	0.010	0.294	0.017
AD	–		–		0.227	0.010	0.255	0.017
L2	–		–		0.191	0.018	0.169	0.020
L3	–		–		0.458	0.019	0.429	0.021
L4	–		–		0.832	0.016	0.785	0.024
L5	–		–		–		0.875	0.047
				Linear parameter				
Intercept	0.127	0.015	0.127	0.010	0.121	0.009	0.125	0.010
MO2	0.056	0.012	0.050	0.010	0.058	0.006	0.053	0.006
MO3	0.121	0.014	0.118	0.011	0.139	0.007	0.134	0.007
MO4	0.230	0.014	0.227	0.012	0.253	0.010	0.245	0.010
MO5	0.316	0.013	0.307	0.012	0.303	0.010	0.312	0.011
SC2	0.041	0.014	0.048	0.010	0.043	0.005	0.038	0.005
SC3	0.127	0.014	0.124	0.011	0.102	0.006	0.096	0.006
SC4	0.186	0.018	0.186	0.013	0.185	0.009	0.176	0.010
SC5	0.233	0.011	0.235	0.011	0.222	0.010	0.224	0.010
UA2	0.036	0.013	0.049	0.010	0.039	0.005	0.035	0.005
UA3	0.099	0.018	0.095	0.012	0.094	0.006	0.088	0.006
UA4	0.165	0.014	0.171	0.012	0.171	0.008	0.161	0.009
UA5	0.201	0.012	0.206	0.010	0.205	0.009	0.205	0.009
PD2	0.045	0.011	0.041	0.009	0.051	0.005	0.050	0.005
PD3	0.106	0.014	0.114	0.012	0.122	0.007	0.126	0.008
PD4	0.235	0.015	0.231	0.012	0.221	0.010	0.231	0.010
PD5	0.250	0.014	0.251	0.013	0.266	0.010	0.258	0.011
AD2	0.024	0.015	0.023	0.010	0.043	0.005	0.043	0.005
AD3	0.109	0.018	0.104	0.013	0.104	0.007	0.109	0.007
AD4	0.195	0.016	0.187	0.012	0.189	0.009	0.200	0.010
AD5	0.217	0.013	0.221	0.012	0.227	0.010	0.223	0.010

ADD20f and ADD20r, additive 20-parameter model with a *fixed/random* constant/intercept; Dimensions: MO, mobility; SC, self-care; UA, usual activities; PD, pain/discomfort; and AD, anxiety/depression; MULT8r and MULT9r, 8- and 9-parameter multiplicative model with a random intercept; SE, standard error.

* Nonlinear models in 20-parameter form for comparison purposes.

† All estimates of SEs were derived by fitting the models using bootstrapping with 10,000 resamples. Bootstrap samples were performed at the level of individual study participants.

values for all 3125 health states, including SEs and 95% CIs, is available in CSV format (see Digital Content File 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.11.016). Using the recommended MULT8r in Table 4, the values for 11211 (the second-best state) and 55555 (the worst state) are 0.955 and −0.391, respectively.

Discussion

With the advance of evidence-based decision making in health policy, economic evaluation has been considered as an important tool for drug pricing and reimbursement in China. Recently published pharmacoeconomic guidelines recommend cost-utility analysis and instruments that can generate utility values reflecting the health preferences of the general Chinese population for calculation of QALYs [22]. Therefore, the availability of a national EQ-5D-5L value set will propel the development of pharmacoeconomics and its use in the official decision-making process of the country.

With the exception of Canada [7], published EQ-5D-5L valuation studies in the United Kingdom [5], Japan [6], Uruguay [8], the Netherlands [9], and Korea [10] have used a 20-parameter model for generating their final value sets. It is not surprising that the 20-parameter model performed well in those studies, given that the health states and the number of observations per health state presently used in the computer-assisted EQ-5D-5L valuation tool were determined with the 20-parameter model in mind. In this study, however, we decided to inform model selection using cross-validation as a proxy for assessing predictive accuracy outside the scope of the observed health states, rather than model fit on the observed data set. The distinction is important; in terms of model fit on the observed data set, the 20-parameter model will always and without exception perform equal to or better than the 8- and 9-parameter models tested here because the 8- and 9-parameter models are constrained variants of the 20-parameter model (see Tables 3 and 4 in which MULT8r and MULT9r have been presented in the 20-parameter format). When model selection is informed by fit on the observed data, models of increased complexity will be preferred. Nevertheless, increasing

Table 4 – Rescaled coefficients (SEs) of the fitted models for predicting values anchored at 0 (dead) and 1 (11111).

Model	ADD20r		MULT8r*		MULT9r*	
	Coefficient	SE†	Coefficient	SE†	Coefficient	SE†
			Nonlinear parameters			
MO	–		0.345	0.012	0.356	0.013
SC	–		0.253	0.011	0.256	0.011
UA	–		0.233	0.010	0.234	0.010
PD	–		0.302	0.012	0.336	0.020
AD	–		0.258	0.011	0.291	0.019
L2	–		0.191	0.018	0.169	0.020
L3	–		0.458	0.019	0.429	0.021
L4	–		0.832	0.016	0.785	0.024
L5	–		–		0.875	0.047
			Linear parameters			
MO2	0.057	0.011	0.066	0.006	0.060	0.007
MO3	0.136	0.013	0.158	0.008	0.153	0.008
MO4	0.260	0.013	0.287	0.012	0.279	0.011
MO5	0.351	0.015	0.345	0.012	0.356	0.013
SC2	0.055	0.011	0.048	0.005	0.043	0.005
SC3	0.142	0.013	0.116	0.007	0.110	0.007
SC4	0.213	0.015	0.210	0.010	0.201	0.011
SC5	0.269	0.012	0.253	0.011	0.256	0.011
UA2	0.056	0.011	0.045	0.005	0.040	0.005
UA3	0.109	0.014	0.107	0.006	0.100	0.007
UA4	0.196	0.013	0.194	0.009	0.184	0.010
UA5	0.236	0.012	0.233	0.010	0.234	0.010
PD2	0.047	0.010	0.058	0.006	0.057	0.006
PD3	0.131	0.013	0.138	0.008	0.144	0.009
PD4	0.264	0.014	0.252	0.011	0.264	0.013
PD5	0.287	0.014	0.302	0.012	0.294	0.012
AD2	0.027	0.012	0.049	0.005	0.049	0.005
AD3	0.119	0.015	0.118	0.007	0.125	0.008
AD4	0.215	0.014	0.215	0.010	0.228	0.012
AD5	0.253	0.013	0.258	0.011	0.255	0.011

Note. Rescaling was done by dividing the dimension parameters (MO, SC, UA, PD, and AD) in the nonlinear models, and all parameters in the linear models, by (1−intercept). SE estimates recalculated using bootstrapping, because the uncertainty around the intercept will influence uncertainty for all parameters.

ADD20r, additive 20-parameter model with a random intercept; Dimensions: MO, mobility; SC, self-care; UA, usual activities; PD, pain/discomfort; and AD, anxiety/depression; MULT8r and MULT9r, 8- and 9-parameter multiplicative model with a random intercept; SE, standard error.

* Nonlinear models in 20-parameter form for comparison purposes.

† All estimates of SEs were derived by fitting the models using bootstrapping with 10,000 resamples. Bootstrap samples were performed at the level of individual study participants.

complexity comes at the risk of overfitting to random variance, effectively reducing predictive accuracy and validity beyond the observed data [23]. Cross-validation methods, although imperfect, can reveal situations in which simplified models improve predictions. Indeed, the 8- and 9-parameter models we tested here outperformed the 20-parameter model in the Singaporean and Spanish EQ-5D-5L valuation data sets (K. Rand-Hendriksen et al., 2016; unpublished data), suggesting that China is not the only country where the more parsimonious 8- and 9-parameter models could be applied. The concept of the constrained models may be applied to other health descriptive systems that use a common set of descriptors in all or multiple dimensions. For example, the same set of frequency descriptors (none/a little/some/most/all of the time) is used in three of the six dimensions of the six-dimensional health state short form [24].

Nonlinear regression models are not without drawbacks. From a practical perspective, methods for fitting nonlinear regression models are not taught as extensively as linear regression, and methods for fitting such models are less accessible than linear regression methods (STATA [StataCorp LP, College Station, TX] provides linear mixed-effects models and nonlinear regression functions, but we have not found any functions in STATA able to fit nonlinear mixed-effects regression, required for MULT8r and MULT9r). Furthermore, although the nonlinear regression functions that are available (at least in R and STATA) readily provide estimates of SEs for the fitted coefficients, these cannot be used directly to calculate SEs for predicted values, because the product of two Gaussian random variables is not itself a Gaussian random variable. For this reason, in addition to the non-normality of TTO distributions, we used bootstrapping to estimate SEs and CIs for all models in this study. Details and codes for fitting the various regression models in R and STATA can be found in a separate study (K. Rand-Hendriksen et al., 2016; unpublished data). To facilitate future research, anonymized raw TTO data for the 1271 included study participants are also available for download in CSV format (see Digital Content File 2 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2016.11.016).

Although the nonzero intercept in our models clearly described the data, we chose to remove it by rescaling so that 1 represents the estimated value of state 11111, rather than the maximum possible value in the TTO task. The rescaling was done on the basis of two separate lines of reasoning. First, there is reason to believe that values for mild states are biased downward. Because the scale is capped at 1, errors or random variance for mild states will be capped upward, but not downward. Second, from the point of ethics, we have to consider that the intended use of the EQ-5D value sets is to inform priority setting in health care. Allowing a large constant term would favor, and therefore could result in overinvestment in, treatments for very mild health problems. For instance, a value of 0.9 for slight problems walking about would imply that correcting a limp for 10 years is as valuable as prolonging a life in full health for 1 year. Our rescaling method maintains the relative impact of all transitions between the 3125 health states, while avoiding controversies that would otherwise occur when using the value set to appraise technologies targeting very mild health problems.

In this study, we did not exclude any observations or participants from our analysis on the basis of logical inconsistency or idiosyncratic behaviors of participants or interviewers. As in other studies [10,25], some of the participants in this study gave logically inconsistent values, that is, valuing logically better states as more undesirable than worse states (e.g., 11112 is logically better than 11113). We also found that logical inconsistency was associated with certain interviewers and decreased toward the end of the survey, suggesting existence of interviewer effects (Z. Yang et al., 2016; unpublished data). In addition, a small number of participants valued all 10 health states as worse than dead, as good as full health, or of the same desirability. Although excluding these minorities improved model fit, the representative of the study sample was impaired, and we had no strong evidence for the invalidity of that part of data.

The EQ-5D-5L values estimated in this study are concordant with the EQ-5D-3L value set for China in some ways, indicating face validity [26]. First, the rank orders of disutility associated with problems at the worst level of the different health dimensions according to the two value sets are very similar—mobility, pain/discomfort, and usual activities were ranked at the first, second, and fifth place, respectively, in both systems. Second, the values for the EQ-5D-5L mildest states (0.934–0.955) are higher than those for the EQ-5D-3L mildest states (0.856–0.887). This is logical because the former involves only slight problems, whereas the latter involves moderate problems in one dimension. It should be noted, however, that the values of the worst health states defined by the EQ-5D-5L and the EQ-5D-3L are −0.149 and −0.391, respectively, although the two worst states were almost identical. This difference could be due to the different TTO methods used to value worse than dead health states in the two studies: lead-time TTO in the present study and conventional TTO in the Chinese EQ-5D-3L valuation study. The difference could also be due to the different ways of handling negative values in the two studies. Regardless of the reasons, the two value sets are not equivalent and should not be used interchangeably.

Conclusions

We found that the simplest model tested, the nonlinear 8-parameter model, performed the best in predicting values for observed health states in cross-validation. We recommend using rescaled values predicted by the 8-parameter model as presented in Table 4 (column 3) to assess health programs and technologies in China whenever the EQ-5D-5L data are available.

Acknowledgments

Source of financial support: This study was cofunded by the EuroQol Group and the Peking University China Center for Health Economic Research.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2016.11.016 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Drummond MF, O'Brien BJ, Stoddart GL, et al. Methods for the Economic. Evaluation of Health Care Programmes. Oxford, UK: Oxford University Press, 1997.
[2] Wisloff T, Hagen G, Hamidi V, et al. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. Pharmacoeconomics 2014;32:367–75.
[3] Szende A, Oppe M, Devlin N. EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Dordrecht, The Netherlands: Springer, 2007.
[4] Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011;20:1727–36.
[5] Devlin N, Shah K, Feng Y, et al. Valuing health-related quality of life: an EQ-5D-5L value set for England. 2016. Available from: https://www.ohe.org/publications/valuing-health-related-quality-life-eq-5d-5l-value-&set-england. [Accessed April 26, 2016].
[6] Ikeda S, Takeru S, Igarashi A, et al. Developing a Japanese version of the EQ-5D-5L value set. J Natl Inst Public Health 2015;64:47–55.
[7] 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;54:98–105.
[8] 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:A810–1.
[9] Versteegh MM, Vermeulen KM, Evers SMAA, et al. Dutch tariff for the five-level version of EQ-5D. Value Health 2016;19:343–52.
[10] Kim SH, Ahn J, Ock M, et al. The EQ-5D-5L valuation study in Korea. Qual Life Res 2016;25:1845–52.
[11] Luo N, Li M, Chevalier J, et al. A comparison of the scaling properties of the English, Spanish, French, and Chinese EQ-5D descriptive systems. Qual Life Res 2013;22:2237–43.
[12] Jia YX, Cui FQ, Li L, et al. Comparison between the EQ-5D-5L and the EQ-5D-3L in patients with hepatitis B. Qual Life Res 2014;23:2355–63.
[13] Pan CW, Sun HP, Wang X, et al. The EQ-5D-5L index score is more discriminative than the EQ-5D-3L index score in diabetes patients. Qual Life Res 2015;24:1767–74.
[14] Oppe M, Devlin NJ, van Hout B, et al. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health 2014;17:445–53.
[15] Devlin N, Buckingham K, Shah K, et al. A comparison of alternative variants of the lead and lag time TTO. Health Econ 2013;22:517–532.
[16] Oppe M, Rand-Hendriksen K, Shah K, et al. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics 2016;34:993–1004.
[17] Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
[18] Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care 2005;43:203–20.
[19] Rand-Hendriksen K, Augestad LA, Dahl FA. A critical re-evaluation of the regression model specification in the US D1 EQ-5D value function. Popul Health Metr 2012;10:2.
[20] R Development Core Team. R: a language and environment for statistical computing. Available from: http://www.R-project.org/. [Accessed April 26, 2016].

[21] Pinheiro J, Bates D, DebRoy S, et al. nlme: linear and nonlinear mixed effects models. 2016. Available from: http://CRAN.R-project.org/package=nlme. [Accessed April 26, 2016].
[22] Liu GG, Dong Z, Wu J. China Guidelines for Pharmacoeconomic Evaluations and Guide. Beijing, China: Science Press, 2015.
[23] Picard RR, Cook RD. Cross-validation of regression models. J Am Stat Assoc 1984;79:575–83.
[24] Brazier JE, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92.
[25] Andrade MV, Noronha K, Kind P, et al. Logical inconsistencies in 3 preference elicitation methods for EQ-5D health states: a study in the Brazilian population. Med Decis Making 2016;36:242–52.
[26] Liu GG, Wu H, Li M, et al. Chinese time trade-off values for EQ-5D health states. Value Health 2014;17:597–604.