VALUE IN HEALTH ▮ (2018) ▮▮▮–▮▮▮

Available online at www.sciencedirect.com

ScienceDirect

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

Time Trade-Off Value Set for EQ-5D-3L Based on a Nationally Representative Chinese Population Survey

Lang Zhuo, PhD Candidate[1,2], Ling Xu, PhD[3], Jingtao Ye, MMeD[2], Sun Sun, PhD[4,5], Yaoguang Zhang, MS[3], Kristina Burstrom, PhD[4,6,7], Jiaying Chen, PhD[1,7,*]

[1]School of Health Policy and Management, Nanjing Medical University, Nanjing, Jiangsu, China; [2]School of Public Health, Xuzhou Medical University, Xuzhou, Jiangsu, China; [3]Center for Statistics and Information, National Health and Family Planning Commission, Beijing, China; [4]Health Outcomes and Economic Evaluation Research Group, Stockholm Centre for Healthcare Ethics, Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden; [5]Division of Epidemiology and Global Health, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden; [6]Equity and Health Policy Research Group, Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden; [7]Center for Health Policy Studies, Nanjing Medical University, Nanjing, Jiangsu, China

ABSTRACT

Objectives: To obtain a nationally representative Chinese three-level EuroQol five-dimensional questionnaire value set based on the time trade-off (TTO) method. **Methods:** A multistage, stratified, clustered random nationally representative Chinese sample was used. The study design followed an adapted UK Measurement and Valuation of Health protocol. Each respondent valued 11 random states plus state 33333 and "unconscious" using the TTO method in face-to-face interviews. Three types of models were explored: ordinary least squares, general least squares, and weighted least squares models. **Results:** In total, 5939 inhabitants aged 15 years and older were interviewed. Of these, 5503 satisfactorily interviewed participants were included in constructing models. An ordinary least squares model including 10 dummies without constant and N3 had a mean absolute error of 0.083 and a correlation coefficient of 0.899 between the predicted and mean values. Goodness-of-fit indices of two models based on split subsample were similar. **Conclusions:** TTO values were higher in our study compared with those in a study carried out in urban areas, which is mirrored by the higher values in rural areas. Several other aspects, in addition to the valuation procedure, might have influenced the results, such as factors beyond demographic factors such as view on life and death and believing in an afterlife, which need further investigation. Future studies using the three-level EuroQol five-dimensional questionnaire should consider using this value set based on a nationally representative sample of the Chinese population.

Keywords: China, EQ-5D-3L, nationally representative sample, time trade-off, value set.

Introduction

In China, the generic three-level EuroQol five-dimensional questionnaire (EQ-5D-3L) has been used to measure health status among the general population in the National Health Services Survey (NHSS) in years 2008 [1] and 2013, and also among patients with diabetes and hypertension [2,3]. The EQ-5D-3L descriptive system describes health-related quality of life (HRQOL) in terms of five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), and each dimension has three severity levels (no, moderate, and severe problems) defining a total of 243 health states. Quality-adjusted life-years combine life expectancy and health status into a single measure requiring values for health states as numeric scores (1 = full health and 0 = dead). Health values can be obtained either through direct methods, such as standard gamble (SG), time trade-off (TTO), and visual analogue scale (VAS), or through indirect methods, for example, via a value set that is attached to an existing multi-attribute instrument (also known as HRQOL measures) such as the EQ-5D-3L [4].

National value sets for the EQ-5D-3L exist for many countries [5], following the first value set established in the United Kingdom [6]. In Asian countries, value sets for the EQ-5D-3L have been established for Japan [7], Korea [8], and China [9,10].

Lang Zhuo and Ling Xu contributed equally to this work.

Conflicts of interest: The authors declare that they have no competing interests. S. Sun and K. Burstrom are members of the EuroQol Group. The views expressed by the authors in this publication do not necessarily reflect the views of the EuroQol Group.

* *Address correspondence to:* Jiaying Chen, Center for Health Policy Studies, Nanjing Medical University, Longmian Avenue 101, Jiangning District, Nanjing, Jiangsu 211166, China.

E-mail: jychen@njmu.edu.cn

https://doi.org/10.1016/j.jval.2018.04.1370

A Chinese TTO value set for the EQ-5D-3L has been published on the basis of interviews with 1222 convenient participants [9]. The authors of that study prove the feasibility of a valuation study; they, however, acknowledge that the main limitation of their study was the sampling method; that is, they had only convenient participants from big cities. In their sample, most were of Han nationality, and they point to the fact that 9% of the Chinese population is of minority nationality and might have different preferences for health. Accordingly, the authors suggest that "future health state valuations should target rural and minority populations in the country." Rural population accounts for half of the Chinese population [11], and there are large disparities in socioeconomic status, lifestyle, and health status between the urban and rural areas in China [12], which might lead to differences in health state valuation. A previous study based on experienced VAS values found such differences, and these differences remained even after controlling for socioeconomic status [1].

Most of the Chinese population (91%) is from the Han ethnic group, whereas the other 55 ethnic groups account for 9% of the population [9]. Some minority groups live among the Han and are very similar to the Han; some groups might live in certain areas, quite often in rural areas, and are different from the Han. It is difficult to set up a specific sampling frame on the basis of the ethnic groups. In the present study, the differences between different ethnic groups were assumed to be somewhat nested within the rural-urban and socioeconomic differences, which were already given consideration in the sampling method.

To carry out a representative valuation study for the EQ-5D-3L is of importance, because this version is in use, for example, in face-to-face interviews such as the NHSS with a large number of questions, and may also provide a possibility to carry out comparison over time when attached to previously collected EQ-5D-3L data in different disease groups in China.

The aim of the present study was to obtain a nationally representative Chinese EQ-5D-3L value set based on the TTO method.

Methods

Health State Description System

An EQ-5D-3L health state, for example, the state 12223, represents no problems with mobility (level 1); moderate problems with self-care, usual activities, and pain/discomfort (level 2); and severe problems with anxiety/depression (level 3) [6,13]. A total of 43 health states (out of the 243 possible) were selected to be valued. Each respondent was randomly assigned to value 11 of those health states (2 very mild states, 3 mild states, 3 moderate states, and 3 severe states) and the health state 33333 plus "unconscious" [6].

Sample Size

The Chinese population aged 15 years and older in 2014 was the target population for this study. This study attempted to draw a nationally representative sample of the Chinese population. According to the population structure in China, we decided that the proportion of urban and rural participants should be equal. Referring to the Dolan estimation, a sample size of 3235 was required [6]. Considering the size of the Chinese population, regional variation, and potential deviation, we increased the sample size to 6000 respondents.

Sampling Method

A multistage, stratified, clustered random sample was drawn from the target population from five different areas on the basis of geographic location and economic development involving the provinces of Jiangsu, Guangdong, Hebei, Chongqing, and Shaanxi. One county (rural area) and one city district (urban area) from each province were selected (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.04.1370 for details). These counties and city districts were sampled in NHSS 2013 [10]. Data were collected from July 10 to August 25, 2014. Ethical approval was obtained by the Nanjing Medical University (NMU) Ethics Committee (NMU_EA20140706004). Participants were informed that it was voluntary to participate and that they could terminate their participation at any time during the interview. Written informed consent was obtained before each interview.

Interviewer Training and Pilot Interviews

In total, 108 interviewers chosen from the local township/street health service center who had participated in NHSS 2013 carried out the survey. The research team at NMU trained 10 faculty members who had a medical background and who were familiar with this type of research and had experience in fieldwork using face-to-face interviews to become supervisors for the local interviewers (see Supplemental Materials for details).

The Valuation Task

Face-to-face interviews in the participants' homes, using paper and pencil, were carried out mainly on the basis of the Measurement and Valuation of Health protocol, however with several modifications of the TTO method [6]. The interview questionnaire included 17 demographic questions, the EQ-5D-3L descriptive system, the EQ VAS, the 11 preselected health states plus the state 33333 and "unconscious," and the TTO valuation exercise. The questionnaire also included the health state 11111 and "dead" for the VAS valuation exercise, which is not reported in this article. The mean time for each interview was 48.6 ± 16.7 minutes, including short periods of rest. The questionnaires were filled by the interviewers. The valuation task and data entry are further described in the Supplemental Materials.

Modeling

Table 1 presents the definitions of the independent dummy variables and the specifications of the models constructed in this study. The main effect within each of the five EQ-5D-3L dimensions is represented by a set of 10 dummy variables. The main effect of movement from no problems (level 1) to moderate problems (level 2) is represented by the dummy variable for level 2, and the movement from no problems (level 1) to severe problems (level 3) is represented by the dummy variable for level 3. If there was a level 2 or level 3 in any of the dimensions, it was represented by the constant and the variable N3. No other interaction terms were included for the avoidance of multicollinearity [9].

The TTO values (dependent variable) were calculated as $t/10$ (states better than dead), $-t/(10 - t)$ (states worse than dead), and 0 (states equal to dead). In the regression models, we used disutility (1 − TTO value) as the dependent variable (Y), that is, the same dependent variable as used in the value sets for the United Kingdom [6], Belgium [14], Denmark [15], and New Zealand [16].

Seven models were constructed on the basis of individual-level data with three types of regression method. First, ordinary least squares (OLS) models were constructed to comprehend the basic characteristics of the model. Then, general least squares (GLS) regression models with multilevel effect were also explored

Table 1 – Definition of the dummy variables and the models.

Dummy variables	Definition	
MO2	1 if mobility is level 2; 0 otherwise	
MO3	1 if mobility is level 3; 0 otherwise	
SC2	1 if self-care is level 2; 0 otherwise	
SC3	1 if self-care is level 3; 0 otherwise	
UA2	1 if usual activities is level 2; 0 otherwise	
UA3	1 if usual activities is level 3; 0 otherwise	
PD2	1 if pain/discomfort is level 2; 0 otherwise	
PD3	1 if pain/discomfort is level 3; 0 otherwise	
AD2	1 if anxiety/depression is level 2; 0 otherwise	
AD3	1 if anxiety/depression is level 3; 0 otherwise	
N3	1 if any dimension is level 3; 0 otherwise	
Model specification	**$f(x)$**	**Methods**
Model 1	f(MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3 N3)	OLS with constant and with N3
Model 2	f(MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3)	OLS with constant and without N3
Model 3	f(MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3)	OLS without constant and without N3
Model 4	f(MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3)	GLS multilevel effect with constant and without N3
Model 5	f(MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3)	GLS multilevel effect without constant and without N3
Model 6	f(MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3)	WLS with constant and without N3
Model 7	f(MO2 MO3 SC2 SC3 UA2 UA3 PD2 PD3 AD2 AD3)	WLS without constant and without N3
Model A	Model 3 + sex	OLS without constant and without N3, robust
Model B	Model 3 + sex + age group	OLS without constant and without N3, robust
Model C	Model 3 + sex + age group + region	OLS without constant and without N3, robust
Model D	Model 3 + sex + age group + region + marital status	OLS without constant and without N3, robust
Model E	Model 3 + sex + age group + region + educational level	OLS without constant and without N3, robust
Model F	Model 3 + sex + age group + region + employment status	OLS without constant and without N3, robust
Model G	Model 3 + sex + age group + region + economic status	OLS without constant and without N3, robust
Model H	Model 3 + sex + age group + region + marital status + education + employment + economic status	OLS without constant and without N3, robust

Dimensions: MO, mobility; SC, self-care; UA, usual activities; PD, pain/discomfort AD, anxiety/depression; GLS, general least squares; OLS, ordinary least squares; WLS, weighted least squares.

to investigate the potential intercorrelations considering that each respondent valued a subset of 13 health states. At last, weighted least squares (WLS) models were also explored to investigate the potential heteroscedasticity problem, where the inverse of the variance of residuals was used as the weight. All multiple linear regression models were built using STATA/SE 12.0 (StataCorp, College Station, TX) with an α of 0.05 [17].

To investigate the interviewer effect, we checked the variance partition coefficient with a multilevel zero model for six variables (participants, family, interviewer, district, urban-rural, and sex) and found that there was an interviewer effect as well as a family member effect, but those effects were far less than the participant effect [18]. Hence, we constructed our GLS models on the basis of the multilevel effect with a variable for participants' ID to investigate the potential intercorrelation considering that each respondent valued a subset of 13 health states.

Criteria for Choice of Model

The best performance model was chosen on the basis of the following criteria [19]: 1) logical consistency in each dimension; that is, level 3 (severe problems) should have a higher value (in absolute terms) than level 2 (moderate problems), and level 2 should have a higher value than level 1; 2) sign; that is, the sign of the main effect coefficients including N3 and constant should be positive when disutility is used as the dependent variable; 3) goodness of fit, how well the model explains the difference between observed and estimated health state value (e.g., adjusted R^2 and mean absolute error [MAE]); 4) parsimony, if goodness of fit makes little difference, the simplicity plays a more important role; and 5) transparency, easy for nonexperts to understand the modeling. Adjusted R^2, MAE, Akaike information criterion (AIC), and Bayesian information criterion (BIC) were used as goodness-of-fit indices [6,19,20].

Sensitivity and Subgroup Analyses and Reliability Tests

The database was randomly split into two subsamples to use one-half of the whole sample for the model development and the other half for estimation of goodness of fit. Demographic and socioeconomic variables were added to the final model to identify the effect on the coefficients.

To investigate the relationship between the values for the first two TTO tasks and the values given when the respondent revalued the same health states, the following reliability tests were performed: rates of consistency, Pearson correlation coefficient, kappa, intraclass correlation coefficient, and Cronbach α.

Results

Data Exclusion

A total of 6041 individuals were approached, 98 of whom were unable to finalize the interview because of various reasons (e.g., senile dementia and psychological problems), and 4 of them were younger than 15 years. There were no respondents in our data set that fulfilled any of the exclusion criteria, except that 10 respondents did not want to trade off any time for any of the health states, that is, valued all health states equally. These respondents are kept in the data set as the interviewer made sure the respondents did not misunderstand the task and hence the values represented their views. At this stage, 5939 respondents were kept (see Supplemental Materials for identification and exclusion of outliers). The final sample for analysis consisted of 5503 individuals, which represented 92.7% of the whole sample.

Sample Characteristics

This study includes not only urban and majority population of the country, but also rural and minority population from the eastern, middle, and western areas to achieve national representativeness. The demographic and socioeconomic characteristics of the sample are presented in Table 2. The proportion of participants from rural areas was 50.9%. Minorities, that is, other than the Han ethnicity, were represented by 10.9%. The proportion of graduates from at least senior high school was 42.7%. Census data in 2014 showed that the rural population was 45.2%. Han ethnicity was 91.0% and the proportion with least senior high school was 44.3% [1]. Our final sample had a little lower proportion in the age group of 15 to 44 years, and a relatively higher proportion in the other two groups. The composing of the final sample, regarding age and sex, compared with the Chinese population is presented in Appendix Table S1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.04.1370.

Health-related characteristics including problems reported on each EQ-5D-3L dimension are also presented in Table 2. Chronic disease that had been diagnosed by a doctor was reported by 25.0%, and disability that had been diagnosed by a doctor before was reported by 3.9%. Most problems were reported on the pain/discomfort dimension (14.6%), followed by anxiety/depression (7.2%).

With the whole sample, the mean TTO value for 43 evaluated states was 0.713 ± 0.446.

Analysis of Models

In Table 3, regression analyses on TTO values—coefficient estimates and statistics of OLS, GLS, and WLS models—are presented. All models were significant ($P < 0.001$). We chose the best-fit model according to the following steps. Among the three OLS models, model 1 (OLS with constant and N3) and model 2 (OLS with constant and without N3) were discarded because of the positive sign of the constant and the N3 term, which resulted in logical errors. Consequently, model 3 (OLS without constant and without N3) was preferred at this stage. In model 4 (GLS with constant and without N3) and model 5 (GLS without constant and without N3), which were based on GLS methods, the significance of all coefficients was inspected ($P < 0.001$), and this meant potential inner correlation did not lead to an invalid significance of the coefficients in OLS methods. Model 4 still had a positive constant, resulting in a logical error. Appendix Figure S2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.04.1370 shows the predicted value of model 5 compared with model 3. The lines are almost overlapping with each other, which means they have almost the same ability of prediction. Comparatively, model 5 has an inferiority of goodness of fit. So model 3 still had superiority at the second stage.

Models 6 and 7 were constructed on the basis of the WLS method to deal with heteroscedasticity [21]. We adopted the inverse of the variance of residuals as the weight, which means the more is the variance of residuals in the specific states, the less important the states would be. Model 6 (WLS with constant and without N3) still had a positive constant. Although model 7 (WLS without constant and N3) had no logical error, the coefficient of PD2 was not significant. In addition, different weights might lead to different models, which brought another arbitrary

Table 2 – Demographic, socioeconomic, and health-related characteristics of the sample (n = 5503).

Characteristic	Category	%	n
Sex	Male	48.2	2653
	Female	51.8	2850
Age (y)	15–24	11.0	604
	25–34	14.9	817
	35–44	17.1	942
	45–54	20.4	1123
	55–64	20.2	1109
	65–74	11.2	617
	75–97	5.3	291
Region	Urban	49.1	2700
	Rural	50.9	2803
District	North (Hebei)	19.8	1090
	East (Jiangsu)	21.4	1177
	South (Guangdong)	17.4	957
	Middle (Chongqing)	19.5	1074
	West (Shaanxi)	21.9	1205
Marital status	Single	13.7	754
	Married	80.1	4410
	Divorced	1.6	84
	Widowed	4.6	252
	Other	0.1	3
Educational level	Below primary school	7.1	389
	Primary school	18.8	1036
	Junior middle school	31.3	1723
	Senior middle school	17.2	948
	Technology school	6.6	364
	Junior college	10.6	585
	University and above	8.3	458
Economic status*	Good	16.5	909
	Common	73.3	4033
	Poor	10.2	561
Occupational status	Employed	60.4	3323
	Retired	18.2	1001
	Students	5.5	305
	Unemployed	1.7	93
	No occupation	14.2	781
Ethnicity	Han	89.1	4905
	Other	10.9	598
Chronic disease†	Yes	25.0	1377
Disability†	Yes	3.9	213
Self-rated health	Very good	25.0	1375
	Good	46.6	2562
	Fair	22.8	1255
	Poor	5.0	275
	Very poor	0.7	36
Dimension			
Mobility	No problem	94.5	5201
	Moderate problem	5.2	285
	Severe problem	0.3	17
Self-care	No problem	98.2	5402
	Moderate problem	1.5	83
	Severe problem	0.3	18
Usual activities	No problem	95.6	5260
	Moderate problem	3.6	199
	Severe problem	0.8	44
Pain/discomfort	No problem	85.4	4701
	Moderate problem	14.2	783
	Severe problem	0.4	19
Anxiety/depression	No problem	92.8	5107
	Moderate problem	6.9	378

continued on next page

Table 2 – *continued*

Characteristic	Category	%	n
	Severe problem	0.3	18
11111	Yes	81.1	4461
At least one at level 3	Yes	1.2	68

[*] Subjective perception compared with neighborhood.
[†] Reported by participants who had been diagnosed before by a doctor in a county hospital or higher.

bias. Nevertheless, the WLS model did not pass the test of heteroscedasticity and so we abandoned the WLS models at last.

From model 1 to model 5, the differences between coefficients of the same variable were all less than 0.05 in absolute terms. Half of them were less than 0.015. Hence, the models were relatively robust, and model 3 appeared preferable because of better consistency, parsimony, and transparency as well as better goodness of fit. So we recommended model 3 as the best-fit model.

Tests and Comparisons

Figure 1 shows the observed mean values compared with the predicted values based on model 3 for evaluated states. The mean error between observed and predicted values was small. Figure 2 shows the predicted value of model 3 compared with the UK and Japanese value sets as well as the Chinese value set by Liu et al. [5,9]. Model 3 gave a remarkable difference with the others, which indicated the necessity of the study.

Model 3 was chosen as the final scoring algorithm. For example, the predicted value for the health state 22322 using this algorithm is:

$$1 - 0.0766 - 0.0441 - 0.0538 - 0.0274 - 0.0359 = 0.7622.$$

All the 243 predicted values based on model 3 are presented in Appendix Table S2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.04.1370.

Influencing Factors

The sociodemographic influence on health state valuation is presented in Appendix Table S3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.04.1370 based on model 3. The coefficient for rural area was significantly positive (0.052) when controlling for sex and age (model C) and significantly positive (0.041) in the full model (model H), controlling also for marital status, educational level, employment status, and self-assessed economic situation. In the full model, the coefficient for self-assessed bad economic status was significantly positive (0.044). The coefficient for sex was not significant, but the coefficients for age groups were significantly positive for nearly all age groups (youngest as reference group). For educational level, all coefficients, except secondary technical school, were significantly negative (below primary school as reference group). The coefficient for unemployment was significantly negative. Table 4 presents the goodness-of-fit indices for the whole sample and the split sample. The indices were similar, which indicated the robustness of the final model.

Table 3 – Regression analyses on TTO values—coefficient estimates and statistics of OLS, GLS, and WLS models.

Parameters	Model 1 (OLS)	Model 2 (OLS)	Model 3 (OLS)	Model 4 (GLS)	Model 5 (GLS)	Model 6 (WLS)	Model 7 (WLS)
11111	1	1	1	1	1	1	1
Constant	0.0318	0.0405	–	0.0391	–	0.0080	–
MO2	−0.0842	−0.0826	−0.0766	−0.0852	−0.0811	−0.0439	−0.0769
MO3	−0.2708	−0.2657	−0.2668	−0.2724	−0.2733	−0.2191	−0.2657
SC2	−0.0556	−0.0589	−0.0441	−0.0575	−0.0475	−0.0514	−0.0746
SC3	−0.3046	−0.2999	−0.2912	−0.3007	−0.2951	−0.1986	−0.2076
UA2	−0.0687	−0.0473	−0.0370	−0.0446	−0.0377	−0.0423	−0.0591
UA3	−0.1005	−0.0671	−0.0538	−0.0647	−0.0557	−0.0801	−0.0844
PD2	−0.0336	−0.0397	−0.0274	−0.0370	−0.0286	−0.0409	−0.0063[*]
PD3	−0.0696	−0.0515	−0.0409	−0.0468	−0.0396	−0.0930	−0.0603
AD2	−0.0552	−0.0520	−0.0359	−0.0513	−0.0404	−0.0408	−0.0449
AD3	−0.2106	−0.1886	−0.1771	−0.1901	−0.1824	−0.0823	−0.1717
N3	0.0651	–	–	–	–	–	–
Adjusted R^2	0.3559	0.3543	–	0.3866	–	0.2832	–
AIC	56070.54	56249.10	56444.80	52051.00	52196.69	107127.40	41832.41
BIC	56180.67	56350.06	56536.58	52170.32	52306.83	107026.40	41924.19
MAE	0.0788	0.0817	0.0838	0.0826	0.0846	0.0808	0.0855
MAE > 0.05[†]	21	25	24	26	26	18	21
MAE > 0.1[‡]	13	12	12	13	14	12	14
r	0.9037	0.9030	0.8989	0.9026	0.8998	0.8834	0.8903
Logical error[§]	33	1	0	1	0	0	0

Note. All models were significant and P = 0.0000; $P < 0.001$ for all regression coefficients unless otherwise stated.
AIC, Akaike information criterion; BIC, Bayesian information criterion; Dimensions: MO, mobility; SC, self-care; UA, usual activities; PD, pain/discomfort AD, anxiety/depression; GLS, general least squares; MAE, mean absolute error between actual mean and predicted value; OLS, ordinary least squares; r, correlation coefficient of mean value and predicted value; TTO, time trade-off; WLS, weighted least squares.
[*] Significance for this coefficient was at P = 0.104.
[†] Number of MAE >0.05 out of 42 states.
[‡] Number of MAE >0.1 out of 42 states.
[§] Number of inconsistencies occurred in 7533 comparable states.

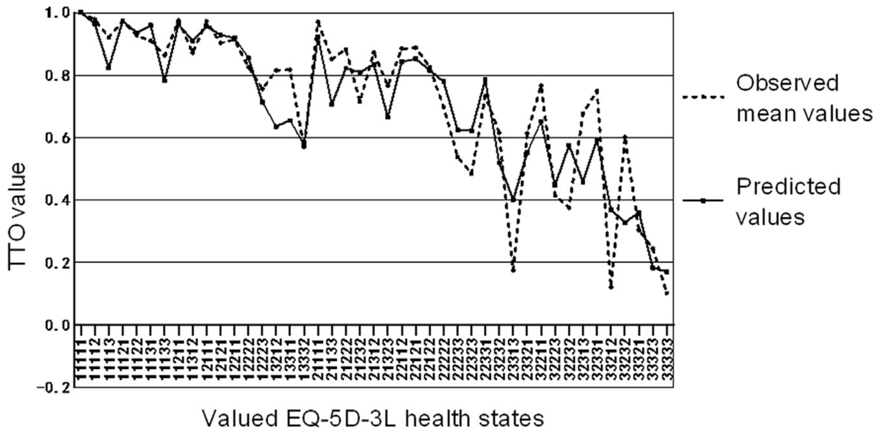

Fig. 1 – Observed mean values compared with the predicted values based on model 3 for the valued EQ-5D-3L health states. EQ-5D-3L, three-level EuroQol five-dimensional questionnaire; TTO, time trade-off.

Reliability Tests of Values for the First Two TTO Tasks

All reliability tests carried out to investigate the relationship between the values for the first two TTO tasks and the values given when the respondent revalued the same TTO tasks at the end of the interview showed that the tests passed the threshold for high reliability [22,23]. The results for the first and second TTO tasks, respectively, are as follows: rate of consistency 90.3% and 89.5%; Pearson correlation coefficient 0.95 and 0.93; kappa 0.88 and 0.88; intraclass correlation coefficient 0.97 and 0.96; Cronbach α 0.97 and 0.96.

Discussion

This study reported on an estimation of a Chinese TTO value set for EQ-5D-3L health states based on a nationally representative sample. The choice of the final model was an OLS model with only the main effects.

Although GLS models have been used in many value sets from different countries, there were a host of authors still preferring the OLS models [7,17,24]. In our study, we constructed OLS models as comparable foundation, and then we used the GLS method to investigate the correctness of significance and the WLS method tentatively to rectify the bias from heteroscedasticity. The decision was a choice based on full comparison according to our criteria.

The N3 is an interactive dummy variable and if it could not improve the goodness of fit markedly, it would be omitted for the sake of parsimony. Almost half of the published value sets had no N3 term [25]. Models without a constant seemed relatively rare to date. D1 models of the United States had no constant [26]. The constant, often labeled as intercept, is the expected mean value of Y when all X equals 0. When all dummies equaled 0, it meant the respondent should be in full health (11111). Then Y, as disutility, equaled 0, which was reasonable and acceptable theoretically. As a result, MAE increased by 0.0021, from 0.0817 to 0.0838, but the number of MAE greater than 0.05 decreased from 25 to 24 and the logical errors were eliminated. Pearson correlation coefficients were almost the same. So, forcing constant equaled 0 almost did little harm to the model practically.

For severe problems, self-care (−0.2912) had the greatest impact on HRQOL in our study, followed by mobility (−0.2668), anxiety/depression (−0.177), usual activities (−0.054), and pain/discomfort (−0.041). Moderate problems in all dimensions except mobility (−0.0766) were less than 0.05 in absolute terms. Severe pain/discomfort had less effect on HRQOL, which was also found in some other eastern countries [7,24]. Nevertheless, in other countries, for example, the United Kingdom, the United States, Zimbabwe, Germany, and the Netherlands [5], the most important effect on HRQOL was exerted by severe problems of pain/discomfort (>0.3 in absolute terms).

Comparing the impact on health dimensions between the present study (model 3) and the study by Liu et al. (model 8) showed that the coefficients for severe problems were the greatest in absolute terms for the mobility dimension followed by the pain/discomfort, self-care, and anxiety/depression dimensions in

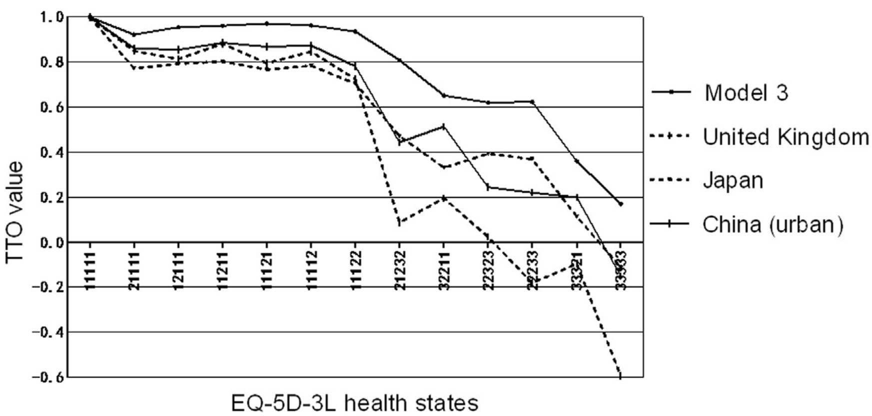

Fig. 2 – Predicted values of model 3 compared with value sets of the United Kingdom [6], Japan [6], and China (urban) [9]. EQ-5D-3L, three-level EuroQol five-dimensional questionnaire; TTO, time trade-off.

Table 4 – Goodness-of-fit indices for the whole and the split sample models.

Paramters	Whole sample model	Split sample model
MAE	0.0838	0.0835
MAE > 0.05*	24	22
MAE > 0.1†	12	12
r	0.8989	0.8968
Logical error‡	0	0

MAE, mean absolute error between observed mean value and predicted value; *r*, correlation coefficient of observed mean value and predicted value.
* Number of MAE >0.05 out of 42 states.
† Number of MAE >0.1 out of 42 states.
‡ Number of inconsistencies occurred in 7533 comparable states.

the study by Liu et al. [9]. The same pattern between the two studies was seen for the coefficients for moderate problems.

For severe problems, the coefficients in our study were larger than those in the study by Liu et al. [9] in the mobility (0.267 vs. 0.246) and self-care (0.291 vs. 0.208) dimensions, but smaller in the usual activities (0.054 vs. 0.193), pain/discomfort (0.041 vs. 0.236), and anxiety/depression (0.177 vs. 0.205) dimensions. For moderate problems, the coefficients for all the dimensions in our study were smaller than those in the study by Liu et al., and the largest differences were observed in the self-care (0.044 vs. 0.105), pain/discomfort (0.027 vs. 0.092), and anxiety/depression (0.036 vs. 0.086) dimensions. Furthermore, because the N3 term was included in the model of Liu et al. but not in our model, values based on their model will be lower for all health states having a dimension at the severe level.

China had a population of 1.37 billion by the end of 2014 and 50% lived in rural areas [11]. A few studies on HRQOL using the EQ-5D about the Chinese population have been carried out in recent years [1–3,10,21,27,28]. In our study, we found that the values from rural populations were higher than the values from the urban populations. These results remained after controlling for age, sex, and socioeconomic status. This might explain why the values in our study are higher than the values in the study by Liu et al., because that study was conducted only in big cities in urban areas [9]. Another possible explanation might be that we used an open-ended TTO question; that is, we did not use an iteration-based procedure. Because of the cultural reasons, as "death" is a taboo to be mentioned in China especially in rural areas and among the older, we did not introduce immediate death after the hypothetical life scenarios. By doing so, we might have introduced a possibility for the respondent to make different assumptions about the length and health status of the continued lives, which might have led to variations in the elicited TTO values. If respondents assumed that life continues after the 10 years in the health state to be valued, they could be unlikely to consider this alternative as worse than dead even if the health state is 33333. Avoiding the mention of "death" in the TTO valuation might have led to an overestimation of the TTO values in our study. Further investigation of the TTO question in the Chinese context would be important.

Socioeconomic factors might have an impact on health state values. Furthermore, people from different cultures might value health states differently. Studies have suggested that respondents with lower socioeconomic status might have lower expectations of health, and therefore, given the same health condition, they might rate their own health status higher than those respondents in higher socioeconomic groups [1]. To some extent, rural and lower socioeconomic status in our study and the convenient sampling in the study by Liu et al. [9] might explain why the values in our study are in general higher than the values in their study. Nevertheless, other factors beyond demographic factors might also be of importance, as found in a study by Jin et al. [29] who investigated the impact on health preferences among the Chinese in relation to their attitudes toward whether bad living is better than good death and their belief in an afterlife. Country-specific value sets based on a nationally representative sample of the population are essential [9,10], and it is important to also consider factors beyond demographic factors in reaching a representative sample in valuation studies in China as stated by Jin et al. [29].

The main strength of our study is the nationally representative sample and a large data set of 5503 participants in face-to-face interviews. Sensitivity analysis showed that the large sample size diluted some potential bias and led to a robust model. The procedure of implementation of the study is an advantage as authorized by the National Health and Family Planning Commission of China, designed and supervised by NMU researchers, supported by local health administrative institutions, and implemented by health service employees. An advantage of the study design was that each questionnaire was checked by the supervisor on the same day of the interview, which resulted in no respondents fulfilling the exclusion criteria. To familiarize the respondent with the TTO task, we initiated for the first two health states the trade-off procedure from 1 day, and then 1 week, 1 month, and 1 year. The reliability test between the initial valuations and the revaluations of those first two health states showed a high reliability. That indicates that the respondents were helped to understand the concept of TTO under the initial trade-off procedure.

A weakness of this study was that we used 108 interviewers. Even though the training was extensive, the interviewers had different backgrounds and used different skills when conducting the interviews. We found some interviewer effect, which was a little higher than the effect of districts, but far less than the participant effect. There are several issues that need to be addressed regarding the use of the TTO method. The willingness to trade off years might be influenced by several aspects, for example, taboo on talking about death; filial piety being one of the classical traditional Chinese cultures that requires that no one is allowed to impair their body or life, which would then lead to huge sorrow for their parents; and general belief that living under any health condition is far better than glorious death [29,30]. The complexity of the TTO method might cause confusion and decrease the adherence even though we introduced the trade-off procedure in the first two TTO tasks and also used the visual aid (Appendix Figure S1. in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.04.1370) The possible variation in the TTO values because of respondents' potentially different assumptions about the length and health status of the continued lives is an issue for further research.

Future studies should investigate factors beyond demographic factors that possibly influence willingness to trade off the length of life to improve the quality of life. Qualitative studies would be valuable to improve knowledge on this issue. To our knowledge, this is the first attempt to explore a value set based on a nationally representative sample in China using the TTO method.

Conclusions

A nationally representative Chinese TTO value set for EQ-5D-3L health states was obtained on the basis of a main effect OLS regression model without constant. TTO values were higher in our study compared with those in the study carried out in urban areas, which is mirrored by the higher values in rural areas.

Several other aspects, in addition to the valuation procedure, might have influenced the results, such as factors beyond demographic factors such as view on life and death and believing in an afterlife, which need further investigation. Future studies using the EQ-5D-3L should consider using this value set based on a nationally representative sample of the Chinese population.

Acknowledgments

We thank the National Health and Family Planning Commission and local health administrative institutes in five provinces for the implementation of the survey. We also appreciate all the dedicated supervisors and interviewers. We are thankful for valuable comments from the Health Outcomes and Economic Evaluation Research Group during earlier stages in the modeling procedure. Valuable comments and suggestions from three reviewers are gratefully acknowledged.

Source of financial support: This study was funded by the National Natural Science Foundation of China (grant no. 71373183).

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at https://doi.org/10.1016/j.jval.2018.04.1370 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Sun S, Chen J, Johannesson M, et al. Population health status in China: EQ-5D results, by age, sex and socio-economic status, from the National Health Services Survey 2008. Qual Life Res 2011;20:309–20.
[2] Pan CW, Sun HP, Zhou HJ, et al. Valuing health-related quality of life in type 2 diabetes patients in China. Med Decis Making 2016;36:234–41.
[3] Tan Z, Liang Y, Liu S, et al. Health-related quality of life as measured with EQ-5D among populations with and without specific chronic conditions: a population-based survey in Shaanxi Province, China. PLoS One 2013;8:e65958.
[4] Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33:337–43.
[5] Agota S, Mark O, Nancy D. EQ-5D Value Sets: Inventory, Comparative Review and User Guide (1st ed.). Dordrecht, the Netherlands: Springer, 2007.
[6] Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
[7] Tsuchiya A, Ikeda S, Ikegami N, et al. Estimating an EQ-5D population value set: the case of Japan. Health Econ 2002;11:341–53.
[8] Jo MW, Yun SC, Lee SI. Estimating quality weights for EQ-5D health states with the time trade-off method in South Korea. Value Health 2008;11:1186–9.
[9] Liu GG, Wu H, Li M, et al. Chinese time trade-off values for EQ-5D health states. Value Health 2014;17:597–604.
[10] Sun S, Chen J, Kind P, et al. Experience-based VAS values for EQ-5D-3L health states in a national general population health survey in China. Qual Life Res 2015;24:693–703.
[11] National Bureau of Statistics of China. One percent of Chinese population of census data from National Bureau of Statistics of China. Available from: http://www.stats.gov.cn/tjsj/ndsj/2015/indexch.htm. [Accessed January 3, 2018].
[12] Sun S, Chen J, Johannesson M, et al. Regional differences in health status in China: population health-related quality of life results from the National Health Services Survey 2008. Health Place 2011;17:671–680.
[13] Oppe M, Rand-Hendriksen K, Shah K, et al. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics 2016;15:1–12.
[14] Cleemput I. A social preference valuations set for EQ-5D health states in Flanders, Belgium. Eur J Health Econ 2010;11:205–13.
[15] Wittrup-Jensen KU, Lauridsen J, Gudex C, Pedersen KM. Generation of a Danish TTO value set for EQ-5D health states. Scand J Public Health 2009;37:459–66.
[16] Devlin N, Hansen P, Herbison P. Variations in self-reported health status: results from a New Zealand survey. N Z Med J 2000;113:517–20.
[17] Augustovski FA, Irazola VE, Velazquez AP, et al. Argentine valuation of the EQ-5D health states. Value Health 2009;12:587–96.
[18] 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:e0184883.
[19] Burstrom K, Sun S, Gerdtham UG, et al. Swedish experience-based value sets for EQ-5D health states. Qual Life Res 2014;23:431–42.
[20] Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). Psychol Methods 2012;17:228–43.
[21] Yang Z. Inconsistency in the valuations of EuroQol EQ-5D-5L health states in China was more related to interviewer and to interview process than to respondents' characteristics. Value Health 2015;18:A737–8.
[22] Leske DA, Hatt SR, Holmes JM. Test-retest reliability of health-related quality-of-life questionnaires in adults with strabismus. Am J Ophthalmol 2010;149:672–6.
[23] Hussain ZI, Lim M, Stojkovic S. The test-retest reliability of fecal incontinence severity and quality-of-life assessment tools. Dis Colon Rectum 2014;57:638–44.
[24] Lee YK, Nam HS, Chuang LH, et al. South Korean time trade-off values for EQ-5D health states: modeling with observed values for 101 health states. Value Health 2009;12:1187–93.
[25] Xie F, Gaebel K, Perampaladas K, et al. Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. Med Decis Making 2014;34:8–20.
[26] 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.
[27] Wang HM, Patrick DL, Edwards TC, et al. Validation of the EQ-5D in a general population sample in urban China. Qual Life Res 2012;21:155–60.
[28] Wang X, Jia X, Zhu M, Chen J. Linking health states to subjective well-being: an empirical study of 5854 rural residents in China. Public Health 2015;129:655–66.
[29] Jin X, Liu GG, Luo N, et al. Is bad living better than good death? Impact of demographic and cultural factors on health state preference. Qual Life Res 2016;25:979–86.
[30] Yin Z, Li J, Ma K, et al. Development of palliative care in China: a tale of three cities. Oncologist 2017;22:1362–7.