Volume 12 • Number 8 • 2009
VALUE IN HEALTH

South Korean Time Trade-Off Values for EQ-5D Health States: Modeling with Observed Values for 101 Health States

Yeon-Kyeng Lee, PhD,[1] Hae-Sung Nam, MD, PhD,[2] Ling-Hsiang Chuang, MSc,[3] Keon-Yeop Kim, MD, PhD,[4] Hae-Kyung Yang, MPH,[2] In-Sun Kwon, MS,[2] Paul Kind, MPhil,[3] Sun-Seog Kweon, MD, PhD,[5] Young-Tack Kim, MD, PhD[1]

[1]Division of Chronic Disease Surveillance, Korea Centers for Disease Control and Prevention, Seoul, Korea; [2]Department of Preventive Medicine, Research Institute for Medical Sciences, College of Medicine, Chungnam National University, Daejeon, Korea; [3]Centre for Health Economics, University of York, York, UK; [4]Department of Preventive Medicine, School of Medicine, KyungPook National University, Daegu, Korea; [5]Jeonnam Regional Cancer Center, Chonnam National University Hwasun Hospital, Jeonnam, Korea

ABSTRACT

Objectives: This study establishes the South Korean population-based preference weights for EQ-5D based on values elicited from a representative national sample using the time trade-off (TTO) method.
Methods: The data for this paper came from a South Korean EQ-5D valuation study where 1307 representative respondents were invited to participate and a total of 101 health states defined by the EQ-5D descriptive system were directly valued. Both aggregate and individual level modeling were conducted to generate values for all 243 health states defined by EQ-5D. Various regression techniques and model specifications were also examined in order to produce the best fit model. Final model selection was based on minimizing the difference between the observed and estimated value for each health state.
Results: The N3 model yielded the best fit for the observed TTO value at the aggregate level. It had a mean absolute error of 0.029 and only 15 predictions out of 101 had errors exceeding 0.05 in absolute magnitude.
Conclusions: The study successfully establishes South Korean population-based preference weights for the EQ-5D. The value set derived here is based on a representative population sample, limiting the interpolation space and possessing better model performance. Thus, this EQ-5D value set should be given preference for use with the South Korean population.
Keywords: EQ-5D, population values, preference-based measures, time trade-off.

Introduction

Economic evaluations of health-care interventions provide important evidence to decision-makers in charge of making efficient resource allocations within their jurisdictions. Quality-adjusted life year (QALY) is one of a number of measurement units in cost-utility analysis for economic evaluation. QALY stands for both quantity and quality of life. To calculate the value of a QALY, a set of value scores needs to be assigned to each of the various health states indicating weights for quality of life, also known as health-related quality of life (HrQoL). It is recommended that these values be calibrated using social preference weights elicited from the general population [1]. In addition, because the preferences for health states can differ across cultures [2], many countries have measured their own population-based preference weights for all possible health states. Several methods to quantify people's preferences for health status have been developed; these include visual analog scale (VAS), standard gamble, time trade-off (TTO), and person trade-off methods [3].

Together with EQ-5D [4], there are other preference-based health status measures that can be used to classify the health state of individuals and summarize the change of health outcome in a single index score. For example, there are the Health Utilities Index [5], SF-6D [6], and Quality of Well-Being Scale [7]. In Korea, as in many other countries, there is growing interest in EQ-5D due to the increasing need of measuring the change in HrQoL as an outcome of the health care program. The Korean version of EQ-5D has been under development for some time. Its reliability and validity has already been proven [8] and it was included in the Korea National Health and Nutrition Survey, designed to measure population health in 2005.

In order to develop a population-based preference weights for EQ-5D (also known as EQ-5D value set), a valuation study was conducted, in which a subset of health states defined by the EQ-5D descriptive system was directly valued. Based on these observed values, a regression modeling approach is adopted to exploit values for all 243 health states defined by EQ-5D. It must be noted here that there appears to be reported in the literature only one earlier study that attempted to develop the EQ-5D value set for the population in South Korea [9]. However, due to drawbacks in the design of its valuation study and modeling, the sample was not nationally representative and the average of absolute differences between observed and estimated scores was as great as 0.071. To the authors' knowledge, to this day the demand for a representative and reliable EQ-5D value set for South Korean population is still not met.

The current study establishes the South Korean population-based preference weights for EQ-5D based on the values elicited from a national representative sample using the TTO method. One of the main features of the survey where the preference data were collected is the number of health states involved in the study. Unlike previous valuation studies performed in Korea or in other countries, where either 43 health states defined by EQ-5D or less were directly valued, here the values for a total of 101 EQ-5D health states have been directly observed. Thus, with this unique dataset it is expected that the interpolation spaces in estimating a value set are minimized in comparison to other value sets.

Address correspondence to: Hae-Sung Nam, Department of Preventive Medicine, College of Medicine, Chungnam National University, 6 Munhwa-dong, Jung-gu, Daejeon, 301-130, South Korea. E-mail: hsnam@cnu.ac.kr
10.1111/j.1524-4733.2009.00579.x

 1098-3015/09/1187 1187–1193

Methods

Study Subjects

The target population for the study is Korean adult citizens, aged 20 and older, estimated at 36.786 million based on the official residential registries on December 31, 2006 [10]. A multistage stratified random sampling was employed aiming at generating a sample representing the age and sex distribution in the target population. Due to limited resources, the target sample size was restricted to 1307. The sampling procedure is explained below.

In the first step, the entire sample was stratified using 15 regions (seven large cities and eight provinces) with the exception of Jeju province, which is an island with a population number equivalent to 1.2 % of the total population. Due to the relatively small number of residents in this province, its exclusion was expected to have only a limited impact on the sampling. The number of subjects was assigned to 15 regions in proportion to the population size of each region. The same process was subsequently repeated within each region using three categorized administrative units: "Dong," "Eup," and "Myun" ("Dong" is a town in a district of a city, "Eup" is a main town in a county, and "Myun" is a township in a county; every address can be categorized into one of these units). In the second step, the final fieldwork locations "Ban" and "Village" ("Ban" is a subdivision of "Dong" or "Eup," and "Village" is a subdivision of "Eup" or "Myun") were selected randomly within the strata defined in the first step. In the third step, 8 to 10 households were randomly selected for interview in each "Ban" or "Village." In those cases where a selected household had more than two persons aged 20 years or more, the interviewers invited the person whose birthday was closest within the next 12 months to the day of interview. Persons residing temporarily at a selected household, such as a lodger, family member in military service, and persons in long-period official trips or overseas duty were excluded.

EQ-5D

EQ-5D is one of the most widely used generic index measures of HrQoL [4]. It consists of two parts, the EQ-5D descriptive system and the EQ-5D VAS. The descriptive system contains five items that measure five dimensions of health including mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension is represented by a single item with three levels of responses: no problem, some problems, and extreme problems. A total of 243 health states are defined by this descriptive system. The EQ-5D VAS records the respondent's self-rate health status on a VAS, where the endpoints are labeled "best imaginable health state" and "worst imaginable health state." As mentioned earlier, the Korean version of EQ-5D has already been developed, and its validity and reliability has been proven [8].

Health State Selection

The survey included 100 EQ-5D health states together with states "33333" and "11111" for direct valuations. The 100 health states chosen comprise 25 mild, 50 moderate, and 25 severe states. The degree of severity was defined by a standard city-block distance metric in which any movement away from "11111" is simply counted for each dimension and aggregated. For instance, state "11121" and state "21113" are categorized into distance groups 1 and 3, respectively. Mild states are those within the distance groups 1 to 4, where there are no level 3 problems and up to three level 2 problems. Severe states are those within the distance groups 7 to 9, in which there are no level 1 problems and at least two level 3 problems. If a state is neither mild nor severe, then it is classified as a moderate state. For example, although state "21113" is in distance group 3, it would be categorized not as mild but as a moderate state due to having a problem of level 3. The 100 health states selected were distributed into 25 blocks (P. Kind, pers. comm.). To ensure that each block contained health states across different severities, each block had six health states composed of two randomly selected mild states, two severe states, and two moderate states. In the current study, each participant was assigned two blocks; one was picked following the numerical order assigned to 25 blocks (i.e., the 100th respondent evaluated the 25th block) and the other was randomly selected. Thus, each participant evaluated 12 health states from two blocks, in addition to the "11111" and "33333" states. The selection of health states for each of the 25 blocks can be found at: http://www.ispor.org/Publications/value/ViHsupplementary/ViH12i8_Nam.asp.

Data Collection

The survey instruments and protocol used were similar to those of the Measurement and Valuation of Health (MVH) study in the UK [11]. The details of the survey are as follows:

The survey was based on a face-to-face interview that can be divided into three stages. In the first stage, respondents described their own health at the time of the interview using the validated Korean version of EQ-5D, including answering the five-item descriptive system and self-rated VAS. In the second stage, the respondents were asked to rank the 12 health states from the two blocks assigned plus the states "11111" and "33333" by putting the "best" health state on top and the "worst" at the bottom. It was assumed that each state was experienced for 10 years followed by immediate death. Subsequently, respondents rated the above ranked 14 health states and the state of immediate death using VAS.

Finally, each respondent evaluated the same set of health states but without state "11111" and immediate death using TTO technique with the double-sided time board and a set of health state cards. The method is also known as TTO props method. A thorough description of the method can be found elsewhere [12] and is therefore not repeated in detail here. In short, the respondents were first asked to decide whether a state is better or worse than death. For states regarded as better than death, respondents decided a period of time t in the state "11111," which they consider as equivalent to 10 years in the target state. The shorter t is, the worse the target state. For the states worse than death, the choice was between dying immediately and spending a length of time $(10 - t)$ in the target state followed by t years in the state "11111." Consequently, the longer the time chosen to be in the state "11111" to compensate for a shorter time in the target state, the worse the target state is [12]. In TTO valuation scale, the states "11111" and immediate death were treated as anchors and assigned values of 1 and 0, respectively. Respondents were also surveyed on socioeconomic background questions after completing the TTO valuation.

The data were collected between February 6 and April 3, 2007. A total of 61 trained interviewers were recruited for this purpose. On completing the survey, each respondent was rewarded a gift certificate equivalent to about 10 US dollars.

Logical Consistency and Exclusion Criteria

The logical consistency approach was applied to examine the quality of data. Logical consistency is defined as follows: for a given pair of health states, if state A of a pair is better than the state B in at least one dimension and not worse in any other, then the valuation for the former state (TTO_A) must be at least as good as the valuation for the latter state (TTO_B) [13,14]. In a situation where this rule is breached, the logical inconsistency is said to occur. For instance, if state "11122" is valued higher than state "11121," this is logically inconsistent.

Following the MVH study, respondents were excluded if they valued less than three states, valued all states the same, valued all states worse than dead, or if there were four or more logical inconsistencies. These exclusion criteria take into account responses with incomplete or unreliable data. The choice of threshold level (three inconsistencies) is based on Ohinmaa and Sintonen's study of the Finnish population [15], where it was found that involving valuation data with more than three inconsistencies would significantly affect a derived value set.

Transforming the Data

The observed times t for each investigated health state in the TTO valuation task were converted into TTO value (h). For the states valued as better than death, the TTO value h is $t/10$. For states worse than death, h is $-t/(10-t)$. A linear transformation formula was applied to TTO values for states worse than death in order to bound negative values to a maximum of –1, where 0 is equivalent to death. This was done as follows [16]:

$$h' = h/39 \quad (1)$$

The lowest value (h) for a state worse than death is –39. This value occurs when a respondent preferred immediate death over the course of 3 months in the target state, followed by 9.75 years in state "11111." With the above formula, this minimum value of –39 was transformed to –1.

Statistical Analysis

Modeling method. The analyses were conducted at both aggregate and individual levels. In the aggregate level analysis, the mean is used to summarize the score of each health state and to estimate a value set based on the aggregated means. Both ordinary least square (OLS) and weighted least square (WLS) regressions were applied in the aggregate level analysis. The applied weights in WLS regression takes into account the number of respondents who rated a particular state. On the other hand, at the individual level each respondent's score for a given health state is introduced into the estimated model. OLS regression and either a random or fixed effects model, depending on the Hausman's test, were employed in the individual level analysis. The individual effects introduced by participants who might have systematically valued health states higher or lower can be eliminated by applying the fixed/random effects model.

Dependent variables. The dependent variable in the regression analysis was computed as 1 minus the transformed TTO value. It represents the measure of disutility by subtracting the value of a given health state from the value of full health. As a result, the predicted value for state "11111" is equal to 1.

Independent variables. In selecting a model, a range of models from earlier applications were reviewed. The simplest model is the main effects model, which comprises 10 dummy variables that indicate the presence of either a level 2 or 3 in a given dimension of the evaluated state. For instance, M2 for mobility level 2, M3 for mobility level 3, SC2 for self-care level 2, SC3 for self-care level 3, UA2 for usual activities level 2, UA3 for usual activities level 3, PD2 for pain or discomfort level 2, PD3 for pain or discomfort level 3, AD2 for anxiety or depression level 2, and AD3 for anxiety or depression level 3. Unlike the model focusing on only main effects, other models such as the N3 and D1 models also take into account the interaction effect. The N3 model includes the N3 term, indicating whether there is any dimension on level 3, along with 10 main effect variables. The D1 model consists of D1, I2, I22, I3, and I32 terms in addition to 10 main effect variables. The D1 term indicates the number of dimensions with problems beyond the first and replaces the constant term. The I2 term represents the number of dimensions at level 2 beyond the first. The I22 term is the square term of I2. The I3 term represents the number of dimensions at level 3 beyond the first. The I32 term is the square term of I3. The following interaction terms were also considered in the modeling process:

N2: whether there is any dimension on level 2,
C2: the number of dimensions on level 2,
C2sq: the square of the number of dimensions on level 2,
C3: the number of dimensions on level 3,
C3sq: the square of the number of dimensions on level 3,
X2: whether there are 2 or more dimensions on levels 2 or 3,
X3: whether there are 3 or more dimensions on levels 2 or 3,
X4: whether there are 4 or more dimensions on levels 2 or 3, and
X5: whether there are 5 dimensions on levels 2 or 3

Due to multicollinearity and the large number of possible combinations of interaction variables, it is not easy to explore all possible models. Thus, apart from the N3 and D1 models, each interaction term was examined individually and then new models were explored by adding the different combinations of the significant interaction terms ($P < 0.01$) to the main effects model, allowing a maximum of three additional interaction terms. For the D1 model, only the significant interaction terms ($P < 0.01$) were kept. The functional form was additive in all models.

Model selection. Given the purpose of this study, the best fitting model is the one that minimizes the difference between the observed and the estimated value in each health state. Hence the overall mean absolute error (MAE) was computed for each investigated model. The number of absolute errors greater than 0.05 and 0.10 in each model were also used as criteria. Note that due to the great variability of responses at the individual level, the goodness-of-fit such as adjusted or overall R^2 at this level is expected to be lower than that at aggregate level analysis.

Various tests were conducted to examine the assumptions made in the model. The normality of residual was investigated using Kolmogorov–Smirnov normality test and the Breusch–Pagan test for examining heteroskedasticity. When heteroskedasticity was found in OLS regression, the standard errors of coefficients in the model were corrected through the estimation of HC3 robust standard errors as described by Long and Ervin [17]. Hausman's test was used to decide between random effects and fixed effects model. The Ramsey regression equation specification error test (RESET) test for model misspecification was also examined.

To examine the robustness of the chosen model, respondents were randomly split into two half samples. The coefficients were first estimated from one half and then used to generate estimated scores and which were then compared with the observed scores in the other half sample. The final value set is then based on the whole valuation sample.

Comparison of final model with other studies. The mean observed TTO values for EQ-5D health states obtained in the current study were compared using *t*-tests to those published by

Table 1 Age and sex distribution of total sample and valuation sample compared to South Korean population

Characteristics	Total sample (n = 1307)	Valuation sample (n = 1264)	South Korean population*
Sex (%)			
Male	49.04	48.42	49.32
Female	50.96	51.58	50.68
Age, years (%)			
20–29	20.81	21.04	20.35
30–39	24.18	24.13	23.81
40–49	23.10	23.26	23.01
50–59	15.00	14.87	15.12
60 or more	16.91	16.69	17.72

*Source: National Statistical Office, Republic of Korea. The official residential registries on December 31, 2006 [10].

Jo et al. [9]. For comparability with Jo et al.'s study, the TTO values for states worse than death were based on monotonic transformation. Furthermore, the coefficient estimations from the main effects model in Jo et al.'s study [9] were compared with those from the current data set (also with monotonic transformation) using the same model.

The estimated value set in the current study was compared with the established value set in other countries such as Japan, UK, and USA, as well as with the estimation obtained in the previous Korean study. For this purpose, Spearman's rank correlation coefficients and mean absolute differences (MADs) between estimated values from the current study and those from others were calculated.

Results

Respondent Sample and Valuation Sample

Of the 1307 respondents, a total of 233 (17.8%) had logical inconsistencies, of which 39 had four or more inconsistencies and were excluded from the sample. In addition, four other respondents were also excluded: two subjects who gave the same values for all 13 states measured; and another two who valued all states as the state worse than dead. As a result, a total of 1264 respondents formed the valuation sample.

Despite the exclusion of 43 participants, as shown in Table 1 the age and sex distribution of the valuation sample was representative of the Korean population. Table 2 presents the sociodemographic and self-reported health characteristics in the total sample and in the valuation samples. There were no significant differences between the two samples for variables such as education, religion, marital status, experience of chronic condition, and self-reported health problems measured by EQ-5D.

Modeling Analysis

Table 3 presents the coefficient estimates and fit statistics results for the aggregate level models using OLS and WLS regression. Only the main effects, N3 and D1 models, are reported here because other models with different interaction terms did not perform better than the above three models. The main effects model based on OLS had an MAE of 0.031 and the number of absolute errors greater than 0.05 and 0.10 were 20 and 1, respectively. All the coefficient estimations were theoretically consistent, having the expected sign and magnitude. The D1 model included significant interaction terms, D1 and I3, only. Despite

Table 2 Demographic characteristics of total sample and valuation sample

Characteristics	Total sample (n = 1307)	Valuation sample (n = 1264)	*P*-value*
Education, years (%)			
6 or less	10.25	10.05	0.116
7–12	52.11	52.61	
13 or more	37.64	37.34	
Religion (%)			
Buddhist	25.79	25.71	0.958
Christian or Catholic	33.82	33.86	
Others	0.38	0.40	
Unbeliever	39.17	39.16	
No answer	0.84	0.87	
Marital status (%)			
Married	72.61	72.55	0.941
Single	23.87	23.89	
Widowed	2.75	2.77	
Divorced/separated	0.77	0.79	
Experience of chronic condition (%)			
Yes	10.41	10.36	0.990
No	89.59	89.64	
In EQ-5D, those reporting problems on (%)			
Mobility	5.89	5.85	1.000
Self-care	0.77	0.79	1.000
Usual activities	4.05	4.11	0.848
Pain/discomfort	21.27	21.04	0.372
Anxiety/depression	17.44	17.09	0.102

*Comparison between valuation sample and excluded respondents by Chi-square test.

Table 3 Parameter estimates and fit statistics of aggregate level models using OLS and WLS regression

	OLS						WLS					
	Main effects		N3		D1		Main effects		N3		D1	
Variable	Coefficient	SE	Coefficient	SE‡	Coefficient	SE	Coefficient	SE	Coefficient	SE	Coefficient	SE
Constant	0.062	0.013	0.050	0.012			0.061	0.013	0.060	0.013		
M2	0.097	0.010	0.096	0.009	0.145	0.012	0.087	0.011	0.084	0.011	0.140	0.012
M3	0.429	0.012	0.418	0.012	0.518	0.016	0.422	0.012	0.415	0.012	0.523	0.018
SC2	0.043	0.010	0.046	0.009	0.096	0.013	0.044	0.010	0.043	0.010	0.098	0.012
SC3	0.148	0.011	0.136	0.013	0.235	0.017	0.164	0.011	0.158	0.012	0.264	0.019
UA2	0.053	0.011	0.051	0.011	0.101	0.012	0.050	0.011	0.047	0.011	0.099	0.012
UA3	0.223	0.011	0.208	0.014	0.308	0.014	0.221	0.012	0.213	0.012	0.318	0.016
PD2	0.039	0.010	0.037	0.009	0.086	0.013	0.039	0.010	0.036	0.010	0.087	0.013
PD3	0.166	0.011	0.151	0.013	0.250	0.016	0.175	0.011	0.168	0.012	0.268	0.018
AD2	0.046	0.010	0.043	0.010	0.093	0.013	0.046	0.010	0.042	0.010	0.093	0.012
AD3	0.175	0.011	0.158	0.012	0.258	0.015	0.182	0.011	0.173	0.012	0.275	0.016
N3			0.050	0.016†					0.027	0.014*		
I3					–0.050	0.014					–0.102	0.023
I32											0.012	0.003
D1					–0.050	0.013					–0.048	0.013
Adjusted R^2	0.980		0.984		0.996		0.987		0.988		0.997	
MAE	0.031		0.029		0.029		0.031		0.030		0.030	
No. (of 101) >0.05	20		15		15		18		22		20	
No. (of 101) >0.1	1		0		0		1		1		0	

*$P > 0.05$; $0.001 <$ †$P < 0.01$; otherwise $P \leq 0.001$.
‡HC3 robust standard error.
M2, mobility level 2; M3, mobility level 3; SC2, self-care level 2; SC3, self-care level 3; UA2, usual activities level 2; UA3, usual activities level 3; PD2, pain or discomfort level 2; PD3, pain or discomfort level 3; AD2, anxiety or depression level 2; AD3, anxiety or depression level 3; N3, any dimension on level 3; I3, the number of dimensions at level 3 beyond the first; I32, the square term of I3; D1, the number of dimensions with problems beyond the first; MAE, mean absolute error; OLS, ordinary least square; WLS, weighted least square; SE, standard error.

the different interaction terms involved, the N3 and D1 models produced identical results with an MAE of 0.029 and the number of states with absolute errors greater than 0.05 and 0.10 as 15 and 0, respectively. All coefficient estimations in the N3 model had positive signs. In contrast, the two interaction terms in the D1 model, the D1 and I3 terms, had negative signs. The negative sign implies a higher value for a health state with more severe problems. The results of the main effects, N3 and D1 models using WLS regression, were generally worse than those based on OLS regression. Particularly, the number of states with absolute error greater than 0.05 in the main effects, N3 and D1 models using WLS regression, were 18, 22 and 20, respectively. A possible cause could be the outweighed number of values for state "33333" ($n = 1264$, SD = 0.44) and, as a result, a greater weight assigned to this particular state.

Considering the consistency of the coefficient estimations and minimizing the difference between observed and estimated values, the N3 model based on OLS regression was selected as the best performing model at the aggregate level. It passed the Kolmogorov–Smirnov test for normality of the residuals (D = 0.083, $P = 0.497$). There was no model or functional form misspecification as suggested by the Ramsey RESET test (F = 0.54, $P = 0.466$). However, it failed the Breusch–Pagan test for heteroskedasticity (F = 4.92, $P = 0.026$). Theoretically heteroskedasticity in the OLS regression can be addressed by applying WLS regression. However, the number of absolute errors greater than 0.05 was significantly increased in the results of applying WLS regression as compared to those resulting from the OLS. Therefore, the N3 model based on OLS regression is still preferred. The HC3 procedure was used to correct the biased standard error of coefficient in the OLS.

The results of modeling at the individual level are shown in Table 4. Based on OLS regression, four models were identified, including the main effects, N3, D1, and X5 models. The D1 model here includes I3, I32, and D1 terms. The MAE was 0.031 for the main effects model and 0.030 for the rest. The number of absolute errors greater than 0.05 and 0.10 were 18, 22, 20, and 19, and 1, 1, 0, and 2, respectively. All models generated theoretically consistent coefficient estimations, apart from the D1 model where the I3 and D1 terms had negative signs. The Hausman test rejected the random effects model and in favor of the fixed effects model (F = 116.89, $P < 0.001$), therefore only results of the fixed effects model are reported. There are two models, main effects and N3, which are based on fixed effects regression and are presented here. The main effects model had an MAE of 0.032 and the number of absolute errors greater than 0.05 and 0.10 were 20 and 2, respectively. The N3 model produced similar results: the MAE was 0.031 and the number of absolute errors greater than 0.05 and 0.10 were 20 and 3, respectively.

The differences between the compared models at the individual level were marginal and therefore it was difficult to select one as the best performing. Also most of the models at the individual level showed significant heteroskedasticity, non-normality of error distribution and model misspecification. For example, the N3 model failed in each of the following tests: Kolmogorov–Smirnov normality, Breusch–Pagan and Ramsey RESET (D = 0.086, $P < 0.001$; F = 1945.99, $P < 0.001$; F = 17.90, $P < 0.001$, respectively).

According to the findings, the N3 model based on OLS regression with aggregate data is the best fitting model, minimizing the difference between the observed and the estimated value in each health state, and is thus chosen as the final model to estimate the value set. When the model robustness was examined by comparing the estimated values from one half sample and the observed values from the other half, both values were highly correlated ($r = 0.983$) with an MAE of 0.040.

Predicted values are calculated using the final model. For example, we calculated the predicted values of state "32322" as follows:

Predicted values = full health – disutility
Full health = 1.000
Disutility for 32322 state = 0.050 + 0.418 (M3) + 0.046 (SC2) + 0.208 (UA3) + 0.037 (PD2) + 0.043 (AD2) + 0.050 (N3) = 0.852.
Predicted values = 1 – 0.852 = 0.148

Table 4 Parameter estimates and fit statistics of individual level models using OLS and fixed effect regression

	OLS								FE			
	Main effects		N3		D1		X5		Main effects		N3	
Variable	Coefficient	SE†	Coefficient	SE†	Coefficient	SE	Coefficient	SE†	Coefficient	SE	Coefficient	SE
Constant	0.061	0.007	0.060	0.007			0.082	0.008	0.057	0.007	0.057	0.007
M2	0.087	0.006	0.084	0.006	0.140	0.008	0.079	0.006	0.088	0.005	0.086	0.005
M3	0.422	0.008	0.415	0.008	0.523	0.013	0.412	0.009	0.428	0.006	0.424	0.006
SC2	0.044	0.006	0.043	0.006	0.098	0.009	0.034	0.006	0.039	0.005	0.039	0.005
SC3	0.164	0.007	0.158	0.008	0.264	0.013	0.155	0.008	0.158	0.006	0.155	0.006
UA2	0.050	0.006	0.047	0.006	0.099	0.008	0.042	0.006	0.049	0.005	0.047	0.005
UA3	0.221	0.007	0.213	0.008	0.318	0.012	0.212	0.008	0.220	0.006	0.215	0.006
PD2	0.039	0.006	0.036	0.006	0.087	0.009	0.030	0.006	0.040	0.005	0.039	0.005
PD3	0.175	0.008	0.168	0.008	0.268	0.013	0.162	0.008	0.183	0.006	0.179	0.006
AD2	0.046	0.006	0.042	0.006	0.093	0.009	0.037	0.006	0.051	0.005	0.049	0.005
AD3	0.182	0.007	0.173	0.008	0.275	0.012	0.172	0.008	0.182	0.006	0.176	0.006
N3			0.027	0.009*							0.017	0.007*
I3					–0.102	0.016						
I32					0.012	0.002						
D1					–0.048	0.009						
X5							0.031	0.010*				
Adjusted R^2	0.533		0.533		0.801		0.533		0.533		0.533	
MAE	0.031		0.030		0.030		0.030		0.032		0.031	
No. (of 101) >0.05	18		22		20		19		20		20	
No. (of 101) >0.1	1		1		0		2		2		3	

0.001 < *P < 0.01; otherwise P < 0.001.
†HC3 robust standard error.
M2, mobility level 2; M3, mobility level 3; SC2, self-care level 2; SC3, self-care level 3; UA2, usual activities level 2; UA3, usual activities level 3; PD2, pain or discomfort level 2; PD3, pain or discomfort level 3; AD2, anxiety or depression level 2; AD3, anxiety or depression level 3; N3, any dimension on level 3; I3, the number of dimensions at level 3 beyond the first; I32, the square term of I3; D1, the number of dimensions with problems beyond the first; X5, whether there are 5 dimensions on levels 2 or 3; MAE, mean absolute error; OLS, ordinary least square; FE, fixed effect regression; SE, standard error.

The observed and predicted means and the difference between the two values can be found at: http://www.ispor.org/Publications/value/ViHsupplementary/ViH12i8_Nam.asp.

Comparison with Previous Korean Study and Other Studies

Among 23 health states in common, nine health states had significantly lower observed means in the current study than in Jo et al.'s [9], including five severe and two moderate states. Such differences are translated into the coefficient estimations in the model. With the same transformation method for a state worse than death and model specification, all coefficient estimations for level 3 are higher for all dimensions apart from self-care in the current study. The biggest magnitude is observed in the mobility dimension with almost a twofold increase (from 0.310 in Jo et al.'s to 0.606 here).

In comparison with other studies, the value set obtained from our final model is highly correlated with the official value set in Japan ($\rho = 0.969$, $P < 0.001$), USA ($\rho = 0.908$, $P < 0.001$), and UK ($\rho = 0.855$, $P < 0.001$), respectively. The MAD between our Korean study and Japan is 0.056, with USA it is 0.105 and with the UK it is 0.322.

Discussion

This study collected TTO values for 101 EQ-5D health states from a South Korean representative sample. Based on these values the population-based preference weights for EQ-5D are developed using the N3 model. This model yields the best fit for the observed TTO value at aggregate level, with an MAE of 0.029 and only 15 (out of a total 101) prediction errors exceeding 0.05 in absolute magnitude.

At the aggregate level, despite the D1 model producing identical results to the N3 model, the negative sign of coefficient estimations for both interaction terms in the D1 model make it less transparent in calculation. It also becomes conceptually difficult to understand why, for instance, health states with more level 3 problems result in an increased value. Thus, the N3 model is preferred.

The empirical comparisons between modeling at aggregate and individual levels in the current study support the use of aggregate level analysis. However, the choice of either aggregate or individual level based analysis is an ongoing debate. Advantages associated with the aggregate level approach include simple modeling, easy interpretation, and being intuitive. On the other hand, the advantages associated with the individual level approach include utilizing the maximum amount of information and treating each respondent's value on an equal basis. Theoretically, individual level analyses might be expected to produce better results with their capacity to adjust for individual effects. However, in practice it is commonly found that there is too much noise in individual level data that hinders the performance of the estimates. In contrast, aggregate level analysis can alleviate such a problem by regressing at aggregate measures to minimize the unwanted variations.

The choice of central tendency measures, such as the mean or median in the aggregate level analysis, is a debatable issue and the exploration of the impact of the choice of central tendency is beyond the scope of this study.

Compared with other valuation studies, the major contribution of the current study is the number of health states that were directly valued. Unlike other studies, either following the 43 EQ-5D health states in the MVH project or decreasing the number of health states investigated to fewer than 43, this study increases the number of health states. A total of 101 health states were valued, of which 23 overlap with the health states in the MVH set. Therefore, there are at least 2.4 times more health states investigated than in other studies, covering almost 42% of the total health states (101/243) defined by the EQ-5D descriptive system. As a result, this study provides more information regarding how values (observed) are distributed in the valuation space defined by EQ-5D, and consequently it limited the interpolation space in the estimations.

There are three possible ways to transfer the TTO value for states worse than death: monotonic, linear, and truncated transformations [16]. The choice of transformation method in our study was purely based on empirical evidence showing that the linear transformation results in the smallest MAE amongst the three methods. There is no theoretical ground for the choice of one method over another. However, there should be awareness of the effect of applying different transformations in the resulting EQ-5D value set and consequently on the cost-effectiveness analysis. For instance, the value set based on linear transformation produces a smaller range of values and therefore the QALY estimation, and possibly QALY gain, will be smaller than those estimated from a value set based on monotonic transformations.

A possible contribution to the discrepancies in observed TTO values and, consequently, coefficient estimation between the current study and Jo et al.'s [9] is the sampling difference. The sampling in our study is from 15 regions representing the whole country (except the Jeju region), while in the latter study it was confined to two adjacent regions only (Seoul and Gyeonggi-do). Our data suggests that the values obtained from the other 13 regions are different from values elicited from Seoul and Gyeonggi-do regions (data not shown). Therefore, the values elicited from these two regions alone cannot be used as a representative preference for the population in South Korea as a whole.

Another possible explanation for differences in coefficient estimations between the two studies could be the number of health states involved. In this study there are 101 health states with directly observed values, whereas Jo et al. [9] use only 42 states. In other words, in our study there is more information available regarding the valuation space defined by EQ-5D, and therefore it minimizes the interpolation spaces in the estimation. Particularly, this study values 26 severe health states directly. In contrast, only seven severe health states were investigated in the latter study. Thus, the coefficient estimation for level 3 problems is likely to be more robust in the current study.

We are confident that the EQ-5D preference weights developed in this study are better than the ones published previously and should be used preferentially for South Korean population. There are three main reasons for this: First, the data was collected from a national representative sample. Second, it is based on the values of 101 health states, which is twice the number of health states used in the earlier study and, on average each health state (apart from state "33333") has about 150 observations to provide a reliable estimate. Finally, the performance of the chosen model in our study is superior to the final model in the previous study in terms of the size of MAE and of the proportion of health states with an absolute estimation error greater than 0.05 and 0.10.

When considering the correlation coefficients and MADs between the estimated value set in the current study and the official value set in other studies, the estimates here are closer to values in the Japanese study than those in the USA and UK. This observation could represent the cultural similarity between Korea and Japan which was also observed in the previous Korean study [7].

In conclusion, the study successfully establishes a set of South Korean population-based preference weights for the EQ-5D. The value set derived here is based on a population representative sample, limiting the interpolating space and possessing better model performance. Thus, this EQ-5D value set should be used preferentially for the South Korean population.

Source of financial support: This study was prepared by financial support of Korea Centers for Disease Control and Prevention (KCDC) in the Ministry of Public Health and Social Welfare.

References

1 Busschbach J, Weihnen T, Nieuwenhuizen M, et al. A comparison of EQ-5D time trade-off values obtained in Germany, The United Kingdom and Spain. In: Brooks R, Rabin R, Charro F, eds. The Measurement and Valuation of Health Status Using EQ-5D: A European perspective. Dordrecht: Kluwer Academic Publisher, 2003.
2 Saxena S, Carlson D, Billington R, et al. The WHO quality of life assessment instrument (WHOQOL-Bref): the importance of its items for cross-cultural research. Qual Life Res 2001;10:711–21.
3 Jelsma J, Hansen K, De Weerdt W, et al. How do Zimbabweans value health states? Popul Health Metr 2003;1:1–10.
4 Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33:337–43.
5 Horsman J, Furlong W, Feeny D, et al. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes 2003;1:1–13.
6 Brazier J, Usherwood T, Haper R, et al. Deriving a preference-based single index from the UK SF-36 Health Survey. J Clin Epidemiol 1998;51:1115–28.
7 Lemon SM, Mahmoud AA. The threat of pandemic influenza: are we ready? Biosecur Bioterror 2005;3:70–3.
8 Kim MH, Cho YS, Uhm WS, et al. Cross-cultural adaptation and validation of the Korean version of the EQ-5D in patients with rheumatic diseases. Qual Life Res 2005;14:1401–6.
9 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.
10 Korea National Statistical Office. Resident of Registration Population (As of December 31, 2006). Daejeon: Korea National Statistical Office, 2007.
11 Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
12 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.
13 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.
14 Dolan P, Kind P. Inconsistency and health state valuations. Soc Sci Med 1996;42:609–15.
15 Ohinmaa A, Sintonen H. Inconsistencies and modelling of the Finnish EuroQol (EQ-5D) preference values. In: Greiner W, Graf von dem Schulenburg JM, Piercy J, eds. EuroQol Plenary Meeting, Hannover, 1st-2nd October. Hannover: Uvi-Verlag Witte, 1999.
16 Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. Med Care 2007;45:238–44.
17 Long JS, Ervin LH. Using heteroscedasticity consistent standard errors in the linear regression model. Am Stat 2000;54:217–24.