View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Volume 13 • Number 2 • 2010
VALUE IN HEALTH

Valuation of EQ-5D Health States in Poland: First TTO-Based Social Value Set in Central and Eastern Europe

Dominik Golicki, MD, PhD,[1] Michał Jakubczyk, PhD,[1,2] Maciej Niewada, MD, PhD,[3] Witold Wrona, MD,[1] Jan J. V. Busschbach, PhD[4]

[1]Department of Pharmacoeconomics, Medical University of Warsaw, Poland; [2]Institute of Econometrics, Warsaw School of Economics, Poland; [3]Department of Experimental and Clinical Pharmacology, Medical University of Warsaw, Poland; [4]Department of Medical Psychology and Psychotherapy, Erasmus University Medical Center, Rotterdam, The Netherlands

ABSTRACT

Objective: Currently, there is no EQ-5D value set for Poland. The primary objective of this study was to elicit EQ-5D Polish values using the time trade-off (TTO) method.
Methods: Face-to-face interviews with visitors of inpatients in eight medical centers in Warsaw, Skierniewice, and Puławy were carried out by trained interviewers. Quota sampling was used to achieve a representative sample of the Polish population with regard to age and sex. Modified protocol from the Measurement and Value of Health study was used. Each respondent ranked 10 health states and valued 4 health states using the visual analog scale and 23 using the TTO. Mean and variance stability tests were performed to determine whether using a larger number of health states per respondent would yield credible results. Modeling included random effects and random parameters models.
Results: Between February and May 2008, 321 interviews were performed. Modeling based on 6777 valuations resulted in an additive model with all coefficients statistically significant, R^2 equal to 0.45, and value −0.523 for the worst possible health state. Means and variance did not differ significantly for states valued in the middle and at the end of the TTO exercise.
Conclusions: This is the first EQ-5D value set based on TTO in Central and Eastern Europe so far. Because the values differ considerably from those elicited in Western European countries, its use should be recommended for studies in Poland. Increasing the number of health states that each respondent is asked to value using TTO seems feasible and justifiable.
Keywords: EQ-5D, Poland, quality-adjusted life-years, social value set, utility.

Introduction

Economic evaluation of health technologies in the cost–utility analysis framework aims at providing maximal utility—as perceived by a given society—within a limited budget, and thus should be based on social preferences that are to be satisfied. The EQ-5D questionnaire is a widely known tool that can be used to elicit social preferences [1].

The EQ-5D was translated into Polish in 1997, after the EuroQol Group guidelines and in interaction with EuroQol translation review members [2]. This version is currently used in clinical research conducted in Poland. The main limitation against the wider application of the Polish EQ-5D in clinical and pharmacoeconomic studies in Poland is the lack of either population norms or a national EQ-5D value set. As a result, the Agency for Health Technology Assessment in Poland (AHTAPol) has been recommending the use of the EQ-5D European value set [3]. This value set was derived using the visual analog scale (VAS) methodology developed during the EuroQol BIOMED Research Programme funded by the European Union (1998–2001) [4]. This might not be an optimal choice, as in health economics, the preferred outcome measure is quality-adjusted life-years (QALYs) and the VAS scale is less associated with the QALY paradigm than choice-based valuation methods, like for instance time trade-off (TTO) [5]. Moreover, the EQ-5D European value set was based only on data collected in Western European countries (Finland, Germany, The Netherlands, Spain, Sweden, and the United Kingdom), lacking data from Poland or any other Central European country. The primary objective of this study was therefore to establish a Polish EQ-5D value set using TTO. The secondary objectives included comparison with the EQ-5D European VAS value set and other potentially useful value sets, as well as assessing the possible bias resulting from expanding the TTO experiment to 23 states per respondent.

Address correspondence to: Dominik Golicki, Department of Pharmacoeconomics, Medical University of Warsaw, Pawińskiego 3A, 02-106 Warsaw, Poland. E-mails: dominik.golicki@wum.edu.pl
10.1111/j.1524-4733.2009.00596.x

Methods

EQ-5D

EQ-5D essentially consists of two pages—the EQ-5D descriptive system (page 2) and the EQ visual analog scale (EQ VAS) (page 3) [1]. The EQ-5D descriptive system comprises the following five dimensions: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension has three levels: no problems, some problems, and severe problems. The respondent is asked to indicate his or her health state by ticking (or placing a cross) in the box against the most appropriate statement in each of the five dimensions. This decision results in a one-digit number expressing the level selected for that dimension. A health state is defined by combining one level from each of the five dimensions. A total of 243 possible health states are defined in this way. Each state is referred to in terms of a five-digit code. For example, state 11111 indicates no problems on any of the five dimensions, while state 11223 indicates no problems with mobility and self-care, some problems with performing usual activities, moderate pain or discomfort, and extreme anxiety or depression.

EQ-5D health states, defined by the EQ-5D descriptive system, may be converted into a single summary index by applying a formula that essentially attaches values (also called weights) to each of the levels in each dimension [6]. The index can be calculated by deducting the appropriate weights from 1, the

© 2009, International Society for Pharmacoeconomics and Outcomes Research (ISPOR) 1098-3015/10/289 289–297

value for full health (i.e., state 11111). Information in this format is useful, for example, in cost–utility analysis. Value sets have been derived for EQ-5D in several countries using the EQ-5D VAS or the TTO valuation techniques [20].

Sample

Between February and May 2008, 10 trained undergraduated medical students surveyed a representative sample of the Polish adult population. Two training workshops for the interviewers were conducted by two study investigators (JB, DG). Each interviewer had to conduct at least one simulated interview. Survey quotas with respect to age and sex were prepared based on demographic data from the Central Statistical Office in Poland [8]. Face-to-face interviews were conducted among visitors of inpatients in eight medical centers in Warsaw (Mazowieckie voivodship), Skierniewice (Łódzkie voivodship), and Puławy (Lubelskie voivodship). Respondents were not promised any compensation but were given an unexpected gift of limited value only after the interview had taken place. The study was approved by the Medical University of Warsaw ethics committee (KB/24/2008) and all respondents gave informed consent.

Study Design and Pilot Tests

This study was based on the most frequently cited EQ-5D valuation study, the Measurement and Value of Health (MVH), conducted in the United Kingdom in 1993 [9]. The MVH study was a large exercise, in which each of 3395 respondents valued 13 different health states. The MVH research group collected values for 43 (out of 245 potential EQ-5D health states [including "unconscious"]), frequently eliciting more than 500 observations per health state.

Because of budget limitations of the current study, it was necessary to modify the MVH protocol. Based on the findings of Lamers et al. [10] who studied the efficiency of the MVH design, the current study proposed to 1) carry out approximately 300 interviews; 2) collect approximately 150 valuations per health state; and 3) increase the number of health states valued per respondent to more than the 17 health states (as in the EQ-5D Dutch [10] and Japanese [11] studies).

Five pilot interviews were conducted by one of the study investigators (DG) and one of the student interviewers. Respondents ranked a set of 19 health states and valued the ranked health states using the EQ VAS and the TTO technique. These pilot interviews showed that the ranking and VAS valuation exercise may be more time-consuming than the TTO exercise alone. Given that TTO was the primary aim of this study, it was decided to 1) reduce the number of health states to be valued in the ranking exercise to 10; 2) reduce the number of ranked health states to be valued using the EQ VAS to 3 or 4 (the best, the worst, the closest to the average, "immediate death"); and 3) increase the number of health states ranked in the TTO exercise to 23. The above changes were the most substantial deviations from the MVH protocol. Possible problems resulting from the increase in the number of states to be valued using TTO, e.g., order effects, were addressed using statistical tests (see "MVH deviations verification" section). The data from pilot TTO exercises were included into the final study data set.

Interview Procedure

Each respondent was asked to perform the following tasks: 1) indicate his or her own health status using the EQ-5D descriptive system; 2) perform the ranking exercise; 3) value the ranked health states using the EQ-5D VAS; 4) rate his or her own health on the EQ VAS; 5) perform the TTO exercise; and 6) answer some socioeconomic background questions. Two sets of 25 cards describing health states according to the EQ-5D descriptive system were used alternatingly by the interviewers (Table 1). Dead and 11111 are not valued in the TTO exercise by its design. All health states used in the MVH study were used, except "unconscious." Two additional health states (23333 and 32333) were chosen from those proposed by Kind [12]. Cards describing health states were divided into two fixed sets in such a way that 1) there was equal representation of "very mild," "mild," "moderate," and "severe" health states in both sets; and 2) the largest number of logical comparisons was allowed (e.g., health states "11112" and "11113" were included in set 1, while health states "11121" and "11131" were included in set 2). This procedure was intended to facilitate ranking by the respondent, to shorten the survey time and to control data quality. Health states applied in the ranking exercise (Table 1, states marked with *) were also selected in the same way to facilitate the survey and shorten the duration of the preliminary part of the study. Equalization of the value of health states in both sets used in the ranking exercise was not a priority. Cards used in the ranking exercise were marked in the corner on the reverse side by a black dot to allow interviewers to make a quick selection among the cards in each set. Health state cards were shuffled before the TTO exercise and were presented in random order. Interviewers were asked not to present the first three cards describing health states worse than death.

The TTO exercise used the same visual probe as used in the United Kingdom and in the United States (where researchers also used a protocol similar to the UK MVH protocol) [9,13]. This probe, often called a "time board," allows for both positive and negative TTO values. The interview book used by Shaw et al. [14] was translated into Polish and used for training the inter-

Table 1 Sets of health state cards used during the interviews

Health state category	Set 1	Set 2
Very mild	11112*	11121*
	11211	12111
	21111	
Mild	11122*	12121
	22112	12211
		22121*
Moderate	11113*	11131*
	11133	11312
	12222	12223
	13332	13212
	21133	13311
	21232	21222*
	21312	21323
	22122*	22222*
	22222*	23313
	22331	32211
	23321	33212
	32313	33321*
	32331	
Severe	22323*	22233
	23333	23232
	32232	32223
	32333*	33232
	33333*	33323*
		33333*
Anchoring states	11111*	11111*
	DE*	DE*

States severity was defined following Kind [12]: very mild—one level 2 problem; mild state—with no level 3 problems and up to three level 2 problems; severe—no level 1 problems and at least two level 3 problems; moderate—neither mild nor severe.
*Cards used in ranking exercise.
DE, death.

viewers and in the pilot interviews. During these pilot tests, it emerged that the book was rather complex. For instance, the description of a single TTO exercise occupied three pages in English (US) and four pages in the Polish version. Because we had planned to elicit values for as many as 23 states using TTO, the TTO valuation task would have taken more than 90 pages. Moreover, it was considered likely that a 90-page-long protocol would obstruct the flow of the interview. Based on the interviewers' suggestions and subsequent pilot tests, the instruction and documentation of the TTO exercise in the protocol book was reduced to a graphic system, which allowed the registering of results of five TTO exercises on a single page (http://www.ispor.org/Publications/value/ViHsupplementary/ViH13i2_Golicki.asp). Every interviewer was trained and had continuous access to a separate instruction standardized on the TTO methodology as carried out in the US study [13].

Respondents were allowed to trade time in months and weeks instead of years when no valuation changes were noted for a period of 9 years on side 1 of the time board (positive values). This was a modification introduced during a TTO valuation study by German EuroQol Group members [15]. Results of the TTO exercise were read out from the scale in the protocol book with an accuracy of 0.25 of a year. States regarded as better than dead were anchored on a scale ranging from full health to dead: X/10. States regarded as worse than dead were calculated as X/10 – 1, so scores were bounded by 0 and –1 [9].

Exclusion Criteria

To ensure "rational" trade-offs, respondents who misunderstand the task were removed [16]. These respondents were identified according to the following exclusion criteria: fewer than three states valued, all states valued worse than dead, all states valued the same, and "serious logical inconsistencies." These respondents were distinguished from those who provided "irrational" values resulting from "normal cognitive imperfections." A logical inconsistency was defined as being an instance where one health state could be clearly seen to be better than another but the respondent ranked it as worse. A logical inconsistency was called "serious" if the difference in valuation was greater than 0.5. It was considered a clear sign that the respondent had misunderstood the task when he or she had 10 or more serious inconsistencies. In these cases, all responses relating to that particular respondent were excluded. Extreme values, defined as values more than 2 SD from the mean, were also excluded.

MVH Deviations Verification

One of the aims of the present study was to evaluate the possible bias resulting from the modifications made to the original MVH study design and by expanding the TTO experiment to 23 states per respondent. We anticipated that respondents might be too fatigued to credibly answer the last TTO questions. At least two issues would arise as a result: the mean valuation of a health state might be different or the variance of this valuation might increase. In the first case, it would pose a problem relating to the credibility of the valuation; in the second case, the overall error of estimation might increase.

To assess the possible bias involved, several tests were conducted. First, we tested whether the mean valuation for each health state differed when it was valued in the middle of the experiment (as 6th–17th state) or at the end of the experiment (as 18th–23rd state). The first five health states were omitted in this comparison, because they might have been perceived as a warm-up task in the TTO experiment. During valuation of first health states, respondents are just learning the rules of TTO exercise and the variance of the valuation may be significantly high. Beside that, the first three states differ from states valued later, because we asked interviewers not to show states worse than dead at the beginning of the TTO exercise. Because we intended to increase the power of the test and minimize the type II error (not finding a difference in means when there was one), a higher than usual significance level of $P = 0.1$ was used. At the same time, to control for the multiple hypotheses testing for the 44 individual states, we used the Hölm–Bonferroni correction for multiple hypotheses testing. A t test with separate variance estimation was used, but the common variance assumption did not change the results. Tests for equality of variances were performed analogously.

Modeling

The dependent variable was the loss of utility associated with a specific health state, i.e., $1 - u$, where u is the utility. The predictor variables included binary variables $d_{k,l}(j)$ equal to 1 or 0 depending on whenever level l of domain k of health state j was 1, 2, or 3. Furthermore, we used derived variables described in earlier EQ-5D valuation studies—N3, D1, I2, I2sq, I3, and I3sq (for definitions, see Table 2) [9,13].

The data had the panel structure with one level being the respondent index, and the other the health state being evaluated. Two approaches were used in modeling. In the first approach, a simple random effects model was built in. For example, it was assumed that the loss of utility assigned by i-th respondent to the j-th health state was described as

$$Y_{i,j} = \alpha_0 + \sum_{k=1}^{5}\sum_{l=2}^{3}\alpha_{k,l}dl(j) + \eta_{i,j} + \upsilon_i$$

where α denotes the model parameters, $d_{k,l}(j)$ was defined as above, $\eta_{i,j}$ denotes the error term associated with a single TTO experiment, and υ_i denotes the error term associated with i-th respondent (and fixed for this respondent across all TTO experiments). It was assumed that $\eta_{i,j}$ and υ_i were independent and normally distributed with zero means. Additionally, models with other variables describing the health state j than just $d_{k,l}(j)$ were used. Random effects modeling was performed with GRETL software [17].

Table 2 Definitions of the independent variables used in the analyses

Variable	Definition
MO2	1 if mobility is at level 2; 0 otherwise
MO3	1 if mobility is at level 3; 0 otherwise
SC2	1 if self-care is at level 2; 0 otherwise
SC3	1 if self-care is at level 3; 0 otherwise
UA2	1 if usual activities is at level 2; 0 otherwise
UA3	1 if usual activities is at level 3; 0 otherwise
PD2	1 if pain/discomfort is at level 2; 0 otherwise
PD3	1 if pain/discomfort is at level 3; 0 otherwise
AD2	1 if anxiety/depression is at level 2; 0 otherwise
AD3	1 if anxiety/depression is at level 3; 0 otherwise
N3	1 if any dimension is at level 3; 0 otherwise
D1	Number of movements away from full health beyond the first (ranging from 0 to 4); D1 = max (0; $d_{1,2} + d_{2,2} + d_{3,2} + d_{4,2} + d_{5,2} + d_{1,3} + d_{2,3} + d_{3,3} + d_{4,3} + d_{5,3} - 1$)
D1sq	The square of the D1 variable
I2	Number of dimensions at level 2 beyond the first I2 = max (0; $d_{1,2} + d_{2,2} + d_{3,2} + d_{4,2} + d_{5,2} - 1$)
I2sq	The square of the I2 variable
I3	Number of dimensions at level 3 beyond the first I3 = max (0; $d_{1,3} + d_{2,3} + d_{3,3} + d_{4,3} + d_{5,3} - 1$)
I3sq	The square of the I3 variable

As a second approach, a more complex random parameters model was estimated using Bayesian statistics. In this more complex model, it was assumed that the respondents could differ not only in the error terms, but also in the model parameters. This approach allowed for a full incorporation of demographic differences. The specification of a model was as follows:

$$Y_{i,j} = (\alpha_0 + \varepsilon_{i,0}) + \sum_{k=1}^{5}\sum_{l=2}^{3}(\alpha_{k,l} + \varepsilon_{i,k,l})d_{k,l}(j) + \eta_{i,j}$$

where α, $d_{k,l}(j)$, and $\eta_{i,j}$ denote as above, and $\varepsilon_{i,0}$ and $\varepsilon_{i,k,l}$ denote the random variability of model parameters on the individual level (notice that υ_i is incorporated in $\varepsilon_{i,0}$ term). It was assumed that $\varepsilon_{i,0}$ and $\varepsilon_{i,k,l}$ were independent normally distributed random variables with fixed variance across all respondents. The random parameters model was estimated using the Bayesian approach and Markov Chain Monte Carlo (MCMC) method in WinBUGS software (MRC Biostatistics Unit, Cambridge, UK) [18]. Noninformative priors were used as follows:

$$\alpha_0, \alpha_{k,l} \sim N(0, 1000)$$

$$\mathrm{Var}(\varepsilon_{i,0}), \mathrm{Var}(\varepsilon_{i,k,l}) \sim \Gamma(\mathrm{mean} = 1, \mathrm{variance} = 1000)$$

During the MCMC simulation, 10,000 initial simulations and 10,000 sample simulations were used. Ninety-five percent confidence intervals were calculated using the percentile method.

The quality of the models was assessed on two levels—on the individual TTO valuation level and on the health state level. On the individual level, standard R^2 coefficient and mean absolute difference between theoretical and empirical value (mean absolute error [MAE]) were calculated. For all 44 health states used in the experiment, the mean absolute difference between value predicted by the given model and average valuation (for the data set used in modeling) was calculated as well as number of states, for which this difference was larger than 0.05 or 0.1.

Comparison with Other Countries' Value Sets

EQ-5D model coefficients and health state values estimated in the present Polish valuation study were compared with those estimated in other countries. Two TTO and two VAS value sets were chosen: 1) the United Kingdom TTO value set (MVH A1 [9]) as it is the "original" standard; 2) the German TTO value set [15] as it emanates from the country closest geographically to Poland; 3) the European VAS value set [5] as recommended by the AHTAPol; and 4) the Slovenian VAS [19] as an example of a country that can be defined as Central European with a similar political history (although not necessarily sharing cultural similarities with Poland). We performed plain comparison of model coefficients and calculated the mean absolute difference between health states values, the number of health states (out of 243) with values more than 0.05 (or 0.1) different from Polish value, and the correlation coefficient between value sets.

Results

In total, 321 respondents completed the interview, of which 53% were females. Age and sex distributions of the respondents after exclusions are shown in Table 3. The interviewed sample group was representative of the Polish general population in terms of age and sex, but contained a large proportion of individuals with higher education, employed people, and students as well as a low percentage of widowed individuals. Approximately 62% of survey participants came from Warsaw, 19% from other towns, and 15% from rural areas. Overall, 86% of individuals in the sample group were inhabitants of Mazowieckie voivodship. The majority of problems reported in the EQ-5D descriptive system were pain/discomfort (40.1%) or anxiety/depression (37.8%). The mean health state recorded on the EQ VAS was 81.4 (SD 14.5), and the mean interview time was 42 minutes (SD 13).

Table 3 Study sample characteristics compared with the Polish general population data

	Study sample, after exclusions (n = 305)	Polish general population (%)
Male (years)	46.9%	49.9
18–24	7.5%	7.5
25–34	9.8%	11.1
35–44	8.9%	8.7
45–54	9.5%	9.8
55–64	7.5%	7.8
65–74	3.6%	4.1
Female (years)	53.1%	51.1
18–24	7.5%	7.2
25–34	10.5%	10.8
35–44	9.5%	8.5
45–54	10.5%	10.1
55–64	9.5%	8.8
65–74	5.6%	5.7
Mean age (SD)	42.8 (15.7)	Not available
Educational level		
Low	5.2%	23.7
Middle	52.1%	58.0
High	42.6%	18.3
Marital status		
Single	18.4%	20.3
Married/living together	72.1%	64.2
Widowed	4.6%	10.4
Divorced	4.9%	5.1
Work		
Employed	62.3%	53.7
Unemployed	3.0%	4.4
Pensioner	3.0%	4.8
Retired	14.8%	14.8
Student	11.1%	6.3
Housewife/househusband	3.9%	Not available
Belief in life after death	63.8%	Not available
EQ-5D		
Those reporting problems on		
Mobility	16.8%	Not available
Self-care	3.3%	
Usual activities	13.8%	
Pain/discomfort	40.1%	
Anxiety/depression	37.8%	
EQ VAS own health		
Mean (SD)	81.4 (14.5)	Not available

EQ VAS, EQ visual analog scale.

There was no response with fewer than three states valued, or with all states valued worse than dead, or with all states valued the same. We identified 532 serious logical inconsistencies in 120 (37%) interviews. Sixteen respondents with 10 or more serious logical inconsistencies were excluded from the final analysis. These respondents did not differ in demographic characteristics from the whole sample group. Eleven of the 16 excluded respondents (69%) were interviewed by the same interviewer, suggesting that the surveyor himself might have been the cause of the logical inconsistencies. Additionally, 206 extreme values, deviating from the mean score by more than 2 SD, were considered invalid and were excluded from the analysis. As a result, the number of useable valuations was reduced from 7351 to 6983 by excluding the 16 respondents with a high number of serious logical inconsistencies and, subsequently, to 6777 after excluding the extreme values (Table 4).

The two final models resulted from the random effects modeling are presented in Table 5. The first model encompasses all

Table 4 Descriptive statistics on health state level—data before and after quality check

	Observed value (data before quality check)				Observed value (data after quality check)			
Health state	No. of observations	Mean	SD	% of negative	No. of observations	Mean	SD	% of negative
11112	171	0.896	0.212	1	157	0.925	0.116	0
11113	171	0.656	0.425	9	150	0.753	0.250	2
11121	149	0.880	0.206	1	137	0.912	0.132	0
11122	173	0.826	0.287	2	160	0.848	0.249	1
11131	149	0.286	0.619	28	140	0.333	0.595	26
11133	170	0.195	0.648	34	163	0.232	0.630	31
11211	170	0.900	0.168	0	154	0.935	0.091	0
11312	147	0.685	0.362	5	135	0.743	0.246	2
12111	148	0.901	0.168	0	133	0.934	0.100	0
12121	150	0.853	0.203	0	135	0.891	0.140	0
12211	149	0.849	0.178	0	131	0.886	0.129	0
12222	170	0.727	0.356	4	157	0.781	0.237	0
12223	149	0.527	0.462	11	131	0.635	0.304	4
13212	150	0.615	0.403	7	134	0.712	0.234	1
13311	150	0.490	0.513	16	132	0.605	0.373	8
13332	170	–0.071	0.655	49	162	–0.040	0.653	48
21111	170	0.915	0.140	0	156	0.934	0.094	0
21133	170	0.202	0.635	29	162	0.232	0.623	28
21222	149	0.760	0.259	1	137	0.799	0.195	0
21232	170	0.287	0.631	26	160	0.324	0.603	24
21312	170	0.549	0.479	11	151	0.673	0.287	3
21323	149	0.417	0.554	20	133	0.530	0.430	13
22112	170	0.783	0.306	3	156	0.826	0.196	0
22121	149	0.803	0.262	1	140	0.825	0.195	0
22122	170	0.754	0.311	3	155	0.797	0.212	0
22222	319	0.663	0.405	7	290	0.747	0.238	0
22233	150	0.058	0.620	40	142	0.081	0.619	39
22323	172	0.296	0.595	25	158	0.366	0.534	21
22331	171	0.071	0.657	39	163	0.099	0.652	37
23232	149	0.046	0.627	41	141	0.061	0.626	40
23313	149	0.129	0.616	38	141	0.165	0.601	36
23321	173	0.293	0.598	25	160	0.356	0.551	21
23333	169	–0.204	0.626	60	161	–0.213	0.627	62
32211	149	0.464	0.559	19	132	0.573	0.445	13
32223	149	0.187	0.587	34	141	0.216	0.580	32
32232	171	–0.050	0.650	50	163	–0.027	0.645	49
32313	171	0.024	0.653	45	163	0.058	0.647	43
32331	169	–0.110	0.627	53	161	–0.092	0.623	52
32333	172	–0.295	0.597	69	152	–0.384	0.508	74
33212	149	0.278	0.600	29	139	0.319	0.574	27
33232	150	–0.183	0.600	60	142	–0.167	0.607	60
33321	148	0.033	0.648	48	140	0.068	0.640	46
33323	150	–0.150	0.606	55	142	–0.143	0.611	56
33333	318	–0.362	0.542	70	285	–0.461	0.458	78
Total or mean	7351	0.383	0.474	24	6777	0.424	0.411	22

Table 5 Comparison of regression models

	Basic		I3sq		Bayesian	
	Coefficient (SD)	*P*-value	Coefficient (SD)	*P*-value	Coefficient (SD)	95% CI
Constant	0.049 (0.018)	0.007	0.035 (0.018)	0.053	0.054 (0.013)	0.028–0.080
MO2	0.052 (0.011)		0.048 (0.011)		0.051 (0.009)	0.035–0.069
MO3	0.331 (0.014)		0.363 (0.016)		0.325 (0.018)	0.289–0.361
SC2	0.054 (0.012)		0.057 (0.012)		0.047 (0.010)	0.028–0.067
SC3	0.235 (0.015)		0.269 (0.016)		0.224 (0.015)	0.196–0.253
UA2	0.046 (0.014)		0.032 (0.014)	0.023	0.048 (0.011)	0.026–0.069
UA3	0.212 (0.014)		0.224 (0.014)		0.212 (0.014)	0.183–0.239
PD2	0.057 (0.011)		0.063 (0.012)		0.058 (0.009)	0.042–0.075
PD3	0.489 (0.012)		0.513 (0.013)		0.485 (0.021)	0.443–0.526
AD2	0.026 (0.013)	0.036	0.030 (0.013)	0.018	0.027 (0.010)	0.007–0.046
AD3	0.207 (0.012)		0.235 (0.013)		0.204 (0.013)	0.179–0.229
I3sq			–0.012 (0.002)			
R^2 overall	0.452		0.452			
MAE	0.039		0.033		0.041	
No. (of 44) > 0.05	10		12		12	
No. (of 44) > 0.10	3		3		3	

All coefficients were significant at $P < 0.001$ unless otherwise stated.
CI, confidence interval; MAE, mean absolute error.

statistically significant variables, including I3sq (tested for significance of individual variables with *t* test and the whole set by *F* test). The R^2 value of the model amounted to 0.4524. Because this model contains the nonintuitive I3sq variable, it might be considered as less creditable. Therefore, a second model using only $d_{k,l}$ variables was estimated. On the other hand, exclusion of statistically significant variables and the correlation with the other independent variables introduces bias. Nonetheless, the parsimonious model was much more intuitive and had an R^2 value of 0.4517, which is only marginally different from the saturated model.

Ninety-five percent confidence intervals for the Bayesian model excluded zero; thus, all domains on all levels significantly influenced the utility values (Table 5). In a Bayesian estimation, R^2 was not meaningful and is not reported here. It is also worth noting that the random parameters modeling results were very similar to the random effects modeling, indicating that the heterogeneity of the surveyed population had generally no impact on the results. Because Bayesian modeling resulted in no extra predictive value (MAE of 0.039 for the parsimonious model vs. 0.041 for the random parameters model), we therefore decided to base the Polish value set on the classical random effects model, with only $d_{k,l}$ variables and without any interaction variables. The final full Polish EQ-5D value set is presented in Table 6.

Table 6 Polish EQ-5D value set

State	Utility	State	Utility	State	Utility	State	Utility
11111	1.000	13132	0.201	22223	0.535	31321	0.351
11112	0.925	13133	0.020	22231	0.310	31322	0.325
11113	0.744	13211	0.670	22232	0.284	31323	0.144
11121	0.894	13212	0.644	22233	0.103	31331	–0.081
11122	0.868	13213	0.463	22311	0.633	31332	–0.107
11123	0.687	13221	0.613	22312	0.607	31333	–0.288
11131	0.462	13222	0.587	22313	0.426	32111	0.566
11132	0.436	13223	0.406	22321	0.576	32112	0.540
11133	0.255	13231	0.181	22322	0.550	32113	0.359
11211	0.905	13232	0.155	22323	0.369	32121	0.509
11212	0.879	13233	–0.026	22331	0.144	32122	0.483
11213	0.698	13311	0.504	22332	0.118	32123	0.302
11221	0.848	13312	0.478	22333	–0.063	32131	0.077
11222	0.822	13313	0.297	23111	0.664	32132	0.051
11223	0.641	13321	0.447	23112	0.638	32133	–0.130
11231	0.416	13322	0.421	23113	0.457	32211	0.520
11232	0.390	13323	0.240	23121	0.607	32212	0.494
11233	0.209	13331	0.015	23122	0.581	32213	0.312
11311	0.739	13332	–0.011	23123	0.400	32221	0.463
11312	0.713	13333	–0.192	23131	0.175	32222	0.437
11313	0.532	21111	0.899	23132	0.149	32223	0.256
11321	0.682	21112	0.873	23133	–0.032	32231	0.031
11322	0.656	21113	0.692	23211	0.618	32232	0.005
11323	0.475	21121	0.842	23212	0.592	32233	–0.176
11331	0.250	21122	0.816	23213	0.411	32311	0.354
11332	0.224	21123	0.635	23221	0.561	32312	0.328
11333	0.043	21131	0.410	23222	0.535	32313	0.147
12111	0.897	21132	0.384	23223	0.354	32321	0.297
12112	0.871	21133	0.203	23231	0.129	32322	0.270
12113	0.690	21211	0.853	23232	0.103	32323	0.090
12121	0.840	21212	0.827	23233	–0.078	32331	–0.135
12122	0.814	21213	0.646	23311	0.452	32332	–0.161
12123	0.633	21221	0.796	23312	0.426	32333	–0.342
12131	0.408	21222	0.770	23313	0.245	33111	0.385
12132	0.382	21223	0.589	23321	0.395	33112	0.359
12133	0.201	21231	0.364	23322	0.369	33113	0.178
12211	0.851	21232	0.338	23323	0.188	33121	0.328
12212	0.825	21233	0.157	23331	–0.037	33122	0.302
12213	0.644	21311	0.687	23332	–0.063	33123	0.121
12221	0.794	21312	0.661	23333	–0.244	33131	–0.104
12222	0.768	21313	0.480	31111	0.620	33132	–0.130
12223	0.587	21321	0.630	31112	0.594	33133	–0.311
12231	0.362	21322	0.604	31113	0.413	33211	0.339
12232	0.336	21323	0.423	31121	0.563	33212	0.313
12233	0.155	21331	0.198	31122	0.537	33213	0.132
12311	0.685	21332	0.172	31123	0.356	33221	0.282
12312	0.659	21333	–0.009	31131	0.131	33222	0.256
12313	0.478	22111	0.845	31132	0.105	33223	0.075
12321	0.628	22112	0.819	31133	–0.076	33231	–0.150
12322	0.602	22113	0.638	31211	0.574	33232	–0.176
12323	0.421	22121	0.788	31212	0.548	33233	–0.357
12331	0.196	22122	0.762	31213	0.367	33311	0.173
12332	0.170	22123	0.581	31221	0.517	33312	0.147
12333	–0.011	22131	0.356	31222	0.491	33313	–0.034
13111	0.716	22132	0.330	31223	0.310	33321	0.116
13112	0.690	22133	0.149	31231	0.085	33322	0.090
13113	0.509	22211	0.799	31232	0.059	33323	–0.091
13121	0.659	22212	0.773	31233	–0.122	33331	–0.316
13122	0.633	22213	0.592	31311	0.408	33332	–0.342
13123	0.452	22221	0.742	31312	0.382	33333	–0.523
13131	0.227	22222	0.716	31313	0.201	Dead	0.000

Test of Additional TTO Valuations

The comparison of health state values when assigned during the middle of the experiment (position 6 to 17) or at the end (position 18 to 23) showed no statistically significant differences neither in mean nor in variance using the Hölm–Bonferroni correction (the smallest *P*-values for means and variances comparison between groups were equal to 0.0161 and 0.006, respectively, with Hölm–Bonferroni threshold of 0.002273). The results are shown in Table 7 (the values have been ordered according to *P*-values in variance testing). We therefore inferred that additional states were credibly valued (with identical means) and increased the precision of the final estimation (i.e., did not inflate the total variance). Hence, the extension of the number of states was justifiable.

Comparison with Other Countries' Value Sets

In a comparison of estimated values, for all health states, it can be seen that the Polish TTO values are higher than the UK values from the MVH survey (Fig. 1a) and were similar to the German TTO values, although estimation of individual states differed (Fig. 1b). Comparison with the European and Slovenian value sets showed the classical pattern of differences between TTO and VAS values, with Polish TTO values higher for the better health states and lower for the worse health states (Fig. 1c,d).

Table 8 presents a statistical summary of cross-country comparisons. Polish health state values correlate significantly with UK values from the MVH A1 value set ($R^2 = 0.90$), but at the same time, mean absolute difference between Polish and UK values is the largest (0.245, compared with 0.117 between Polish and German values).

Discussion

In this study, we performed 321 face-to-face TTO interviews, directly measuring population values for 44 EQ-5D health states and estimated the Polish EQ-5D value set using random effects modeling.

The Polish EQ-5D valuation study differed in several aspects from the original UK MVH protocol. First, we reduced the ranking exercise to 10 health states and nearly eliminated the VAS valuation of ranked health states (four states) while increasing the number of health states to 23 valued in the TTO exercise. To date, this is the highest number of health states per respondent valued in a national valuation study. Second, similar to valuation studies performed in The Netherlands [10] and in Japan [11], predetermined health state sets were used; yet, in contrast to the before-mentioned studies, there were two sets, not one. Third, we developed a detailed description of the TTO exercise in a separate instruction booklet, leaving the graphic form of recording

Table 7 Equality of mean and variance tests for evaluations in the middle vs. at the end of the time trade-off experiment

	In position 6th–17th		In position 18th–23rd			
State	N	Mean (SD)	N	Mean (SD)	*P*-value for equality of variance test	*P*-value for equality of mean test
12223	97	0.482 (0.511)	36	0.627 (0.235)	**0.0060**	0.0278
22112	80	0.856 (0.196)	33	0.704 (0.323)	0.0062	**0.0161**
22323	102	0.279 (0.56)	41	0.174 (0.694)	0.0282	0.3911
11312	63	0.659 (0.439)	50	0.742 (0.246)	0.0411	0.2057
13212	69	0.546 (0.501)	47	0.68 (0.289)	0.0597	0.0702
23321	102	0.309 (0.579)	52	0.202 (0.655)	0.1040	0.3206
21133	98	0.154 (0.65)	50	0.383 (0.591)	0.2232	0.0337
32313	106	0.044 (0.625)	40	−0.042 (0.683)	0.2405	0.4937
23232	81	−0.001 (0.676)	40	0.006 (0.59)	0.2496	0.9547
12211	66	0.863 (0.165)	29	0.839 (0.205)	0.2651	0.5883
11122	68	0.834 (0.192)	47	0.785 (0.397)	0.3050	0.4419
33212	94	0.262 (0.604)	40	0.347 (0.554)	0.3200	0.4287
21111	80	0.932 (0.109)	30	0.905 (0.169)	0.3264	0.4206
33323	75	−0.106 (0.595)	47	−0.19 (0.647)	0.3708	0.4718
22331	109	0.081 (0.68)	43	0.048 (0.626)	0.3979	0.7761
22122	66	0.741 (0.369)	49	0.769 (0.249)	0.4051	0.6307
32223	80	0.141 (0.59)	48	0.26 (0.575)	0.4301	0.2632
21232	89	0.311 (0.631)	53	0.144 (0.658)	0.4500	0.1400
22222	151	0.671 (0.378)	91	0.613 (0.437)	0.4626	0.2922
32333	90	−0.374 (0.574)	49	−0.277 (0.578)	0.4784	0.3435
32232	87	−0.087 (0.642)	62	−0.047 (0.69)	0.4806	0.7222
11133	102	0.208 (0.638)	45	0.169 (0.684)	0.5035	0.7473
13332	106	−0.1 (0.643)	42	0.029 (0.7)	0.5391	0.3080
11113	91	0.666 (0.403)	47	0.658 (0.449)	0.5941	0.9236
33232	88	−0.142 (0.603)	45	−0.266 (0.581)	0.6072	0.2538
11211	70	0.919 (0.169)	38	0.93 (0.087)	0.6145	0.6455
13311	85	0.51 (0.534)	37	0.497 (0.484)	0.6186	0.8958
21222	74	0.716 (0.263)	35	0.776 (0.303)	0.6453	0.3140
21312	86	0.518 (0.503)	51	0.548 (0.475)	0.6605	0.7319
11121	47	0.866 (0.254)	35	0.851 (0.224)	0.6729	0.7844
11112	61	0.932 (0.102)	37	0.942 (0.122)	0.6751	0.6629
12111	58	0.887 (0.194)	19	0.908 (0.198)	0.6751	0.6946
11131	81	0.328 (0.606)	36	0.188 (0.656)	0.7216	0.2815
33321	97	0.034 (0.623)	36	−0.073 (0.657)	0.7608	0.3984
33333	171	−0.357 (0.568)	104	−0.358 (0.515)	0.7611	0.9873
21323	89	0.437 (0.545)	43	0.471 (0.55)	0.7744	0.7393
22233	91	0.004 (0.599)	47	0.131 (0.652)	0.8198	0.2675
22121	70	0.803 (0.282)	23	0.793 (0.215)	0.8251	0.8522
32211	91	0.471 (0.531)	32	0.522 (0.558)	0.8261	0.6504
23333	99	−0.224 (0.638)	36	−0.25 (0.663)	0.8477	0.8411
12121	62	0.862 (0.199)	30	0.83 (0.218)	0.9053	0.4989
12222	86	0.74 (0.364)	37	0.712 (0.37)	0.9309	0.6982
23313	87	0.149 (0.61)	43	0.195 (0.599)	0.9459	0.6814
32331	106	−0.093 (0.622)	37	−0.243 (0.643)	0.9901	0.2228
All	3851	0.326 (0.649)	1912	0.33 (0.651)	0.9487	0.8028

The smallest value in each *P*-value column has been put in bold. The minimal Hölm–Bonferroni threshold amounts to 0.0023.

the TTO results in the protocol book. Additionally, we confirmed the results of random effects modeling in random parameters modeling (using the Bayesian approach), proving that heterogeneity of the surveyed population had no influence on the results.

The limited number of respondents can be regarded as a potential weakness of our study, although the size of the sample group is similar to the sample size in the German [15] and Dutch [10] valuation studies. The sample size proved to be sufficient in obtaining statistically significant model coefficients. Another important weakness of our study may be lack of formal random allocation of responders to our study population. The representativeness of the sample was controlled with respect to two characteristics—age and sex—through quotas. Of course the sample is not representative with respect to other features: education, geography, or having a relative/friend inpatient. It is difficult to state a priori whether this characteristics impact the preferences among health states. On one hand, it can be argued that it can impact the preferences depending on the relative's illness; on the other, it may increase the awareness of the respondent and improve the quality of the results. Although patient relatives may not be an ideal population for valuation exercise, they were interviewed in earlier valuation studies in Spain and Argentina [20,21]. Our highly educated study sample, although not representative for general population, may better comprehend and perform TTO exercise, making the results more consistent and credible, and actually improve goodness of fit. Nevertheless, it may be the case that the highly educated differ from worse educated in the preferences among health states either directly or indirectly via another characteristics (e.g., increased wealth). In this case, our results would be biased. Because there were no previous health surveys using EQ-5D in Poland, it is unlikely to assess the representativeness and compare our study sample with the Polish general population considering health status data. This line of research should be pursued in the following studies.

Finally, by using fewer interviewers, we would probably have improved the quality of data, but for practical reasons, we were unable to reduce their number below 10. Despite these limitations, goodness-of-fit analysis proved that direct utilities from surveys fit the model-derived utilities relatively well with R^2 of 0.45, which is close to 0.46 obtained in the MVH study [9].

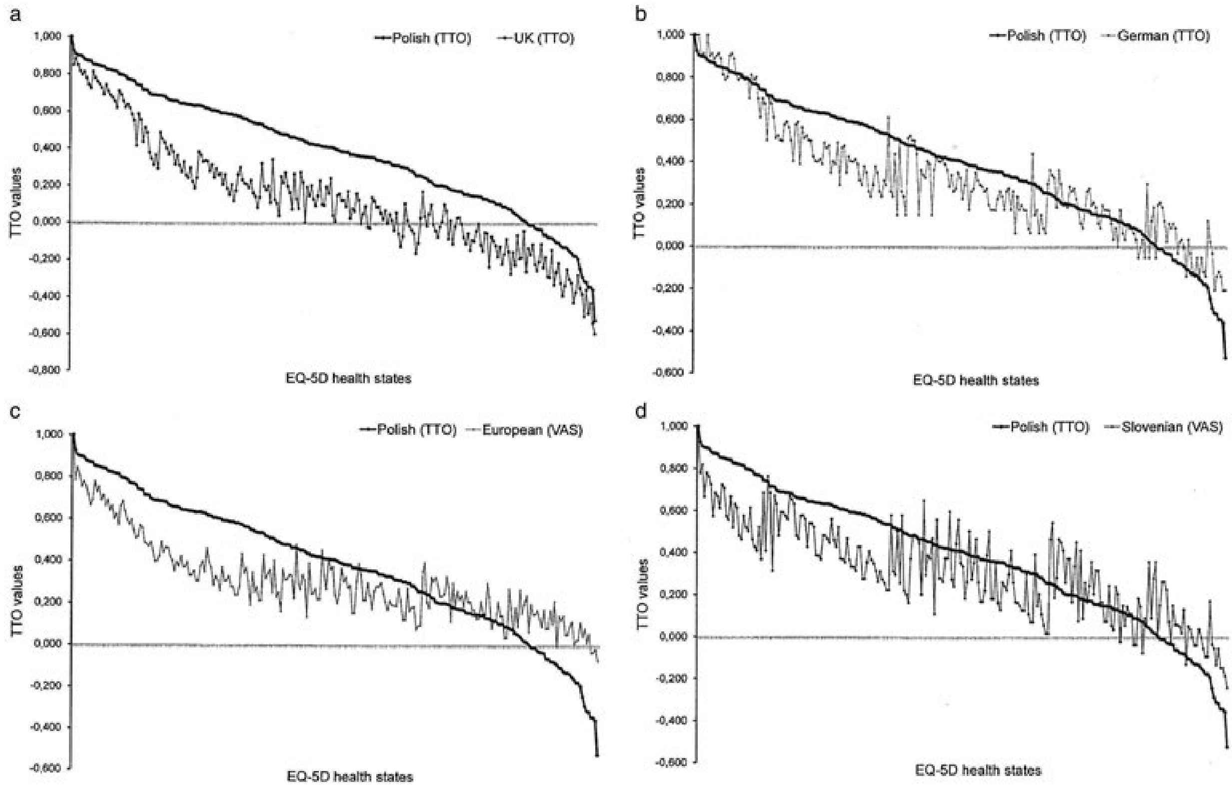

Figure 1 Graphical comparison of Polish EQ-5D TTO value set versus: (a) UK TTO, (b) German TTO, (c) European VAS, and (d) Slovenian VAS value set. TTO, time trade-off; VAS, visual analog scale.

Moreover, we believe that the present study supports the use of more health states in TTO experiments than previously thought. We found no statistical differences neither in mean nor in variance between valuations when a given health state was valued as 6th–17th in a row or 18th–23rd. Therefore, there is no risk of a bias or efficiency decrease in the estimation. This finding provides evidence for improving the efficiency of valuation protocols and supports the estimation of national value sets in other countries, including Central and Eastern Europe.

In the Polish version of EQ-5D, as in the Dutch and Italian versions, the wording chosen for the third level of "mobility," "confined to bed," implies being bedridden. Some of the authors were concerned that this could have caused the Polish values to be lower for health states that included the third level of "mobility" [7]. In comparing dimensions (based on level 3 coefficients and assuming no other problems), individuals in Poland, similar to those in the United Kingdom and Zimbabwe, valued problems associated with "pain/discomfort" as the worst and not the problems associated with "mobility," as in Argentina, Denmark, Germany, Japan, Spain, the United States, and Spanish-speaking Hispanic US residents [2]. In Poland, similar to Argentina and the Spanish-speaking Hispanic US

Table 8 Comparison of coefficients in Polish and selected European models

	Polish TTO	UK TTO (MVH A1)	German TTO	European VAS	Slovenian VAS
Constant	0.049	0.081	0.001	0.1279	0.128
MO2	0.052	0.069	0.099	0.0659	0.206
MO3	0.331	0.314	0.327	0.1829	0.412
SC2	0.054	0.104	0.087	0.1173	0.093
SC3	0.235	0.214	0.174	0.1559	0.186
UA2	0.046	0.036		0.0264	0.054
UA3	0.212	0.094		0.0860	0.108
PD2	0.057	0.123	0.112	0.0930	0.111
PD3	0.489	0.386	0.315	0.1637	0.222
AD2	0.026	0.071		0.0891	0.093
AD3	0.207	0.236	0.065	0.1290	0.186
N3		0.269	0.323	0.2288	
Mean absolute difference		0.245	0.117	0.167	0.146
No. (of 243) > 0.05 vs. Polish		238 (98%)	178 (73%)	210 (86%)	193 (79%)
No. (of 243) > 0.10 vs. Polish		219 (90%)	122 (50%)	178 (73%)	147 (60%)
R^2 vs. Polish		0.90	0.81	0.74	0.73

MVH, Measurement and Value of Health; TTO, time trade-off; VAS, visual analog scale.

residents, the "anxiety/depression" domain was ranked as the least important [21,22]. This differs from other countries where the "usual activities" domain was judged the least important.

Conclusions

This is the first EQ-5D value set based on TTO in Central and Eastern Europe so far. Because the values differ considerably from those elicited in Western European countries, its use should be recommended for studies in Poland. Increasing the number of health states that each respondent is asked to value using TTO seems feasible and justifiable.

We thank Paul Kind, Aki Tsuchiya, and Benjamin Craig for useful comments and guidance. We also thank Rosalind Rabin for editorial support.

Source of financial support: This study was sponsored by unrestricted grants from AstraZeneca Pharma Poland, GSK Commercial, and Pfizer Poland.

References

1 The Euroqol Group. EuroQol—a new facility for the measurement of health-related quality of life. Health Policy 1990;16:199–208.
2 Rabin R, Herdman M, Fox-Rushby J, Badia X. Exploring the results of translating the EQ-5D into 11 European languages. In: Brooks R, Rabin R, de Charro F, eds. The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective. Dordrecht: Kluwer, 2003.
3 Agencja Oceny Technologii Medycznych. Wytyczne przeprowadzania Oceny Technologii Medycznych (HTA). Kraków-Warszawa, 2007. Available from: http://www.aotm.gov.pl/assets/files/wytyczne_hta/Wytyczne_HTA_w_AOTM.pdf [Accessed August 14, 2009].
4 Greiner W, Weijnen T, Nieuwenhuizen M, et al. A single European currency for EQ-5D health states. Results from a six-country study. Eur J Health Econ 2003;4:222–31.
5 Drummond MF, Sculpher MJ, Torrance GW, et al. Methods for the Economical Evaluation of Health Care Programmes (3rd ed.). Oxford: Oxford University Press, 2005.
6 Brooks R, the EuroQol Group. EuroQol: the current state of play. Health Policy 1996;37:53–72.
7 Szende A, Oppe M, Devlin N, eds. EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Dordrecht: Springer, 2007.
8 Central Statistical Office. Available from: http://www.stat.gov.pl [Accessed November 15, 2007].
9 Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
10 Lamers LM, McDonnell J, Stalmeier PF, et al. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006;15:1121–32.
11 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.
12 Kind P. Improving the MVH protocol: learning form the past. Work in progress.
13 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.
14 Agency for Healthcare Research and Quality, Rockville, MD. EQ-5D Health State Valuation Survey Conducted by Research Triangle Institute on Behalf of the University of Arizona Center for Health Outcomes and PharmacoEconomic Research—Interviewer Book. 2005. Available from: http://www.ahrq.gov/rice/EQ-5D/forms/EQ5Dbook.pdf [Accessed November 15, 2007].
15 Greiner W, Claes C, Busschbach JJ, Graf von Schulenburg JM. Validating the EQ-5D with time trade off for the German population. Eur J Health Econ 2005;6:124–30.
16 Busschbach JJV, McDonnell J, Essink-Bot ML, van Hout BAV. Estimating parametric relations between health description and health valuation with an application to the EuroQol instrument. J Health Econ 1999;18:551–71.
17 Gnu Regression, Econometrics and Time-Series Library. GRETL 1.7.5. Available from: http://gretl.sourceforge.net/ [Accessed April 23, 2008].
18 WinBUGS with DoodleBUGS. Version 1.4. Imperial College & Medical Research Council, UK. 2000. Available from: http://www.mrc-bsu.cam.ac.uk/bugs [Accessed April 23, 2008].
19 Prevolnik Rupel V, Rebojl M. The Slovenian VAS tariff based on valuations of EQ-5D health states from the general population. EuroQol Plenary Meeting. Pamplona, Spain, 2000.
20 Badia X, Roset M, Monserrat S, Herdman M. The Spanish VAS tariff based on valuation of EQ-5D health states from the general population. In: Rabin RE, Busschbach JJV, Charro FT, et al., eds. EuroQol Plenary Meeting Rotterdam 1997. Discussion papers. Rotterdam: Centre for Health Policy & Law, Erasmus University, October 2–3, 1998.
21 Augustovski FA, Irazola VE, Velazquez AP, et al. Argentine valuation of the EQ-5D health states. Value Health 2009;12:587–96.
22 Zarate V, Kind P, Chuang LH. Hispanic valuation of the EQ-5D health states: a social value set for Latin Americans. Value Health 2008;11:1170–7.