ORIGINAL RESEARCH ARTICLE

Valuation of EQ-5D-5L Health States in Poland: the First EQ-VT-Based Study in Central and Eastern Europe

Dominik Golicki[1,3] · Michał Jakubczyk[2,3] · Katarzyna Graczyk[3] · Maciej Niewada[1,3]

Abstract

Objective Cost-utility analyses are becoming increasingly important in Central and Eastern Europe. We aimed to develop a Polish utility tariff for EQ-5D-5L health states.

Methods Face-to-face, computer-assisted interviews were collected in a representative sample. Each respondent followed a standardised protocol to collect ten composite time trade-off and seven discrete choice experiment observations. In the Bayesian approach, several model specifications were compared based on model fit, the usability of the final value set and how they reflect the elicitation procedure (e.g. censoring). A hybrid approach (using composite time trade-off and discrete choice experiment data) was employed in the final set, which was compared with the existing ones: EQ-5D-3L and EQ-5D-5L cross-walk.

Results Data from 1252 respondents (11,480 composite time trade-off valuations and 8764 discrete choice experiment pairs) were collected over the period June to October 2016. The final model accounted for random parameters, error scaling with fat tails, censoring at -1, unwillingness to trade in time trade-off by the religious people and Cauchy distribution in discrete choice experiments. Pain/discomfort impacts the utility most: the disutility equals 0.575 when at level 5. In the value set, 4.4% of EQ-5D-5L states are worse than dead. The new value set has a comparable range (minimum of -0.590 compared to -0.523) and the same ordering of the first three dimensions (pain/discomfort, mobility, self-care) as the EQ-5D-3L value set and the EQ-5D-5L cross-walk value set. Moreover, it is more sensitive to a moderate decline in health.

Conclusions The new value set supports consistency with past decisions in cost-utility studies, while offering a better assessment of even moderate improvements in health. It could represent an option for Central and Eastern Europe countries lacking their own value sets.

1 Introduction

Health technology assessment is developing rapidly in Central and Eastern Europe (CEE): e.g. in Bulgaria, Czechia, Hungary and Croatia [1–4]. In Poland, it is compulsory when applying for drug reimbursement [5]. The Polish Health Technology Assessment Agency (AOTMiT) has issued approximately 1800 recommendations since its

---

Dominik Golicki and Michał Jakubczyk participated in the study to a similar extent.

✉ Dominik Golicki
dominik.golicki@wum.edu.pl

Extended author information available on the last page of the article

---

Key Points For Decision Makers

The EQ-5D-5L value set was developed based on directly measured health preferences of a representative sample of Polish society

It should provide a substitute for a mapping-based cross-walk value set when calculating quality-adjusted life-years based on EQ-5D-5L results in Poland

Researchers from Central and Eastern European countries may consider it as an option when national health preferences data are lacking

The new value set provides ground for consistency with past decisions in cost-utility analyses while being sensitive even to moderate health improvements

foundation in 2006, and has assessed nearly 500 health technology assessment reports since the introduction of the current Reimbursement Act in 2012 [6]. Based on this regulation, cost utility is the preferred form of pharmacoeconomic analysis with the official threshold for the cost per quality-adjusted life-year updated yearly [7]. AOTMiT recommends EQ-5D for the purposes of valuing health states and calculating quality-adjusted life-years [8].

The EQ-5D questionnaire consists of a descriptive system and a visual analog scale [9]. The descriptive system contains five dimensions: mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD) and anxiety/depression (AD). In the original version (EQ-5D-3L), each dimension has three levels: *no*, *some* or *severe* problems; whereas there are five levels in the new version (EQ-5D-5L): *no*, *slight*, *moderate*, *severe* or *extreme* problems [10, 11]. The previous two mentioned questionnaires define 243 and 3125 health states, respectively. By attaching disutility to each of the levels in each dimension, it is possible to calculate a single value (EQ-Index) for every health state, forming a value set [12]. EQ-5D-5L demonstrates better measurement properties than EQ-5D-3L [13, 14].

There are only two published EQ-5D-3L value sets in CEE countries: Slovenian [15] and Polish [16], and no value set for EQ-5D-5L. Although it has been possible to use EQ-5D-5L in Poland [17–20], using the mapping-based cross-walk value set [21, 22], the lack of a directly measured EQ-5D-5L value set limited the implementation within the decision-making process.

Our objective was to derive a Polish tariff for the EQ-5D-5L descriptive system, using a standardised approach developed by the EuroQol Group. Such a value set could also be used by other CEE countries that are too small to finance valuation studies, yet are culturally similar and are likely to have congruent health preferences.

2 Methods

The methods and analyses reported in this paper comply with the CREATE guidelines for reporting valuation studies of multi-attribute utility-based instruments [23].

2.1 Study Design

Quota-based sampling was applied using Polish census data from November 2014, based on personal identification number registry (PESEL) and Central Statistical Office data on education [24]. A representative sample in terms of age, sex, education, geographical region and the size of the locality was obtained from Polish residents aged 18+ years. Individuals were recruited through a mixed strategy (public locations, personal contact). Interviews were conducted in public

venues or at participants' homes. Respondents received a financial incentive (voucher of value equivalent to €8).

The study design followed a valuation protocol: EuroQol Valuation Technology (EQ-VT 2.0). It includes software for conducting computer-assisted personal interviews, an interviewer script, standardised training materials, data quality-control procedures and an Excel-based quality-control tool enabling monitoring of protocol compliance, interviewer effects and the validity of the collected data [25].

2.2 Valuation Interview

Computer-assisted personal interviews consisted of four main parts: introduction, composite time trade-off (TTO) valuation, discrete choice experiment (DCE) valuation and country-specific background questions. After a general introduction and explanation of the purpose of the study, the respondents self-reported their health using the EQ-5D-5L questionnaire and answered basic background questions (about age, sex and experiences of severe illness).

In the composite time trade-off valuation, a composite approach was used: starting with the standard TTO (to find the number of years in full health equivalent to 10 years in an impaired EQ-5D-5L state) and shifting to a 'lead time' TTO when participants considered the state to be *worse than dead* (see detailed descriptions [26–29]). The resulting TTO values range from − 1 to 1 in 0.05 increments (the smallest tradable unit being 6 months in duration).

The TTO part of the interview consisted of an explanation of the TTO procedure (the 'being in a wheelchair' example and three practice states: mild, severe and difficult to imagine), proper TTO valuation of ten EQ-5D-5L health states, a structured TTO debriefing and the TTO feedback module. Each respondent was presented with the rank ordering of health states derived from previous responses, to indicate states for which they were not happy with the ranking (though there was no possibility of re-evaluation).

The TTO experimental design included 86 EQ-5D-5L health states distributed into ten blocks that were balanced in terms of severity of states. The health states used in EQ-VT 2.0 were selected using a Monte Carlo simulation [30]. Each block included one of five very mild states (only one dimension at level 2 and all others at 1), the most severe state ('55555') and eight intermediate states. Respondents were randomised into one of the ten blocks; the health states were presented in a random order.

In the DCE valuation task, participants were presented with a pair of EQ-5D-5L health states with no duration specified (labelled A and B) and asked to indicate which they consider 'better' [30–32]. This part of the interview consisted of instructions regarding the task, the valuation of seven pairs and a structured debriefing.

The DCE experimental design included 196 pairs of states randomly divided into 28 blocks, which were identified using an efficient Bayesian design. The blocks were similar in terms of severity, assessed by the sum of the level scores of the health states (i.e. the misery index). Participants were randomly assigned to one of the blocks. The question order and left-right positioning of states were randomised.

The set of Polish country-specific questions covered: priorities in TTO valuations (length or quality of life), general health using an SF-1 question from the SF-36 questionnaire [33], comorbidities, potential concerns during severe illness, religiosity and beliefs, relationship status, childcare responsibilities, professional status and financial situation. In accordance with the EQ-VT protocol, the minimal recommended sample size for EQ-5D-5L valuation studies is $N = 1000$ (see the detailed description [30]). Given a planned experimental arm of our research, we established the basic target sample size at $N = 1250$ (the methods and results of the experimental substudy will be reported elsewhere).

2.3 Quality Control and Data Analysis

We excluded (1) interviews of suspicious quality ('flagged' interviews; for a detailed description of quality-control procedures see Electronic Supplementary Material [ESM] 1), (2) the first ten interviews conducted by an interviewer not meeting the minimum quality criteria (at least seven unflagged interviews) and (3) individual TTO valuations when marked by the respondent in the Feedback Module as not adequately representing their health preferences. No individual DCE valuations were excluded. Descriptive statistics were used to summarise the respondent's characteristics and responses to the TTO and DCE tasks.

2.4 Modelling

2.4.1 General Approach

Below, we present the general approach (dependent/independent variables, model-selection criteria, estimation technique and the building blocks of the model specification under consideration). The formal specification is presented in ESM 1, Online Resource 2.

We based the final model on data from both elicitation techniques (often referred to as a hybrid approach). In the recent literature, all three approaches are used: TTO only [34], DCE only [35] or both [36–39]. As it remains unknown if one clearly outperforms the other, we deemed it safest to have both of them impact the value set (which necessarily worsens the model fit). Therefore, there are two dependent variables: the reported utility of a state (for TTO) and the choice made from a pair of states (for DCE). The states' dimensions are taken as independent variables. In the process of constructing the final model, several specifications were tested: the choices were based on statistical criteria, pragmatic reasons (what the estimation results are used for) or our beliefs concerning how the elicitation tasks work.

In the estimation process, we used a Bayesian approach [40], as we find it more intuitive and flexible to work with a code (JAGS model run from within R, the code in ESM 2) directly describing the data generation process. To let the data speak, we used non-informative priors. In the estimation, we used a Markov-chain Monte Carlo simulation with, respectively, 2000, 30,000 and 20,000 adaptive, burn-in and actual iterations (2000, 20,000 and 10,000 for the intermediate models), no thinning and four chains. The medians of posterior distributions were used as point estimates, and 2.5 and 97.5 percentiles to construct 95% credible intervals. The model fit was assessed based on deviance and penalised deviance (deviance information criterion [DIC]). Potential scale reduction factors were monitored to diagnose convergence for individual parameters [41].

We only used main effects, i.e. no interactions between dimensions. This was a pragmatic decision, undertaken to ensure the final model may also be useful when only partial information is available (e.g. marginal distributions of levels for each dimension separately) [42]; for similar reasons, models with no constant term were preferred (also supported by results).

We tested (and utilised in the final model) the random parameters approach: the disutilities of dimensions/levels differ between individuals. Not only do we find this assumption intuitive but in addition the usefulness of random parameters (and the choice of specific distribution) was confirmed by DIC. Nevertheless, to limit random noise and the number of parameters, and also to avoid technical assumptions (the logical ordering of levels), we assumed it is the importance of each dimension (the disutility of level 5) that is distinctive for each individual, while the relative importance of each level is fixed across individuals (somewhat resembling the idea of simplifying how relative level importance is modelled [43]).

It is not possible in TTO to report a utility lower than −1. Hence, we tested (and used in the final model) censoring: the observed −1s are treated as ≤−1. Some authors use censoring at 0 (where TTO is changed for lead time TTO) or at 1 (in TTO, a value greater than 1 cannot be reported) [38], which we find unconvincing. Regarding censoring at 0, negative values are possible in the protocol used, and modelling an endogenous self-censoring process would require assumptions (is a given zero the true utility or the effect of censoring?). Being unable to decide if a state is worse than dead is not equivalent to being unable to report <0 utility. Regarding censoring at 1, values above 1 are impossible, not only owing to the protocol but also

because of the logical construction of the descriptive system and how the utility values were normalised.

Typically (and in our dataset), there is more variability in responses to more severe states (with lower utility, on average). This may be explained by the random parameters approach, as used in the present paper. Nonetheless, we find it plausible that for a given individual (the importance of dimensions known) there is an additional error term in TTO responses, and that this error tends to be larger for more severe health states (intuitively, for a state whose true utility for a given individual is close to 1, there is little room for a larger error). Therefore, we assumed that the scale parameter of the distribution increases with the theoretical disutility. Specifically, we used a generalised t-Student distribution with the scale and the number of degrees of freedom treated as parameters, allowing for fat tails (but also having a normal distribution as an asymptote).

In the DCE part, we assumed the probability of one state being chosen is a function of the difference in utilities, as is typically done. In the standard approach, this dependence is given by the cumulative distribution function of the logit distribution. Instead, based on the previous findings [44] and the DIC, we used the Cauchy distribution.

Previous research suggests that people with religious beliefs may misrepresent their preferences in TTO tasks, owing to an unwillingness to trade life-years—interpreted as a reporting bias, rather than a difference in preference [45]. For this reason, we introduced a parameter that scaled down the disutilities for religious respondents (separately for TTO and DCE), to disentangle the underlying and the reported preferences. In the final model, the scaling was not found in the DCE part, confirming the above interpretation.

2.4.2 Intermediate Models

We constructed several models sequentially, introducing additional building blocks in succession, and controlling for the DIC improvement, potential scale reduction factors and for whether the 95% credible interval contained a neutral value (i.e. a form of statistical significance). In this paper, we present the results of some of the intermediate steps (all based solely on TTO data):

- M1—panel random-effects approach, with heteroscedasticity-robust standard errors;
- M2—fixed parameters Bayesian model, with no constant term;
- M3—random parameters Bayesian model;
- M4—as M3, with error depending on the theoretical disutility via a t-Student distribution;
- M5—as M4, with scaling as a result of religiosity.

We decided not to present the intermediate steps of the DCE-only part, as the parameters would require some anchoring (for more details on this issue, see [46]). However, as in the DCE part, we monitored the impact of modelling assumptions on DIC.

2.5 Value Set Comparison

There are three EQ-5D value sets available for Poland: EQ-5D-3L [16], EQ-5D-5L mapping-based cross-walk [22] and the present, directly measured EQ-5D-5L value set. To compare the utility values, we used three methods. First, we estimated the kernel density function of the utility values. Second, we identified the median and the worst levels between the EQ-5D-3L and EQ-5D-5L systems and we presented the utilities for all states. In the ESM 1, Online Resource 6, we additionally present the scatter plot to illustrate the relationship between the EQ-5D-5L value set and the other value sets.

3 Results

3.1 Sample Characteristics

From June to October 2016, 15 interviewers conducted 1570 interviews. The mean interview time was 41.1 minutes. In total, 2.3% of interviews were flagged. After excluding interviews with experimental TTO blocks (the results of the study will be reported elsewhere), 29 flagged interviews and six interviews that failed to meet the minimum quality criteria, data from 1252 respondents (52.5% female) aged 18–91 years (mean 46.2; standard deviation 17.6) were available (Table 1).

The sample was representative of the Polish population in terms of age, sex, educational background, employment status, size and geographical location of the place of residence (Fig. 1). It was also similar to the Polish population in terms of health as measured by the EQ-5D-5L descriptive system, EQ visual analog scale and SF-1 [20].

3.2 Data Characteristics

In total, 12,520 individual TTO valuations were available, with a mean number of 250 (standard deviation 6.7) observations per mild health state (misery index 6) and a mean number of 125 (standard deviation 5.3) observations for other 80 health states. In TTO, in 10.7% of the experiments, the time was not traded, and eight respondents (0.6%) did not trade for any state (an additional four respondents valued all the states at the same level, in each case with a utility of 0.95). In 1552 (13.5%) experiments, the valuations were considered worse than dead. In 271 (2.4%) and 784 (6.1%) cases, a utility of 0 and -1 was reported, respectively.

Table 1 General characteristics of respondents

Characteristics	Study sample (N = 1252)		Polish general adult population (30.7 million) [19, 23]
	N	%	%
Age group (years)			
18–34	378	30.2	30.2
35–49	313	25.0	25.1
50–64	332	26.5	25.6
65+	229	18.3	19.2
Sex			
Female	657	52.5	52.6
Male	595	47.5	47.4
Size of place of residence			
Rural area	501	40.1	39.5
Town of less than 100,000 inhabitants	404	32.3	32.9
City of 100,000 and more inhabitants	345	27.6	27.6
Geographical location of residence (macro-region)			
Central	242	19.4	20.3
Southwest	136	10.9	10.3
South	245	19.6	20.6
Northwest	199	15.9	16.0
North	187	15.0	15.0
East	241	19.3	17.9
Education			
Primary or middle school	221	17.7	17.9
Vocational school	328	26.2	24.9
Secondary school	428	34.2	35.9
Higher	273	21.8	21.3
Employment status			
Employed/self-employed	637	51.2	49.7
Unemployed (able to work)	90	7.2	8.4
Unemployed (unable to work, annuitant)	77	6.2	6.7
Student (full time)	114	9.2	7.2
Homemaker, housewife	32	2.6	3.4
Retired person	295	23.7	24.7
Responsibility for children	429	34.3	
Number of persons in a household, mean, SD	2.96	1.4	2.69
Considering himself/herself as a religious person	1127	90.1	92.3
Religion (among religious persons)			
Catholicism	1106	98.1	92.0
Other	21	1.9	
Believe in life after death			
Definitely yes	390	31.2	44.0
Rather yes	375	30.0	31.0
I don't know	228	18.2	7.0
Rather no	127	10.1	18.0
Definitely no	113	9.0	
Experience with serious illness			
In self	382	30.5	
In family	892	71.2	
In caring for others	626	50.0	
Comorbidity confirmed by doctor	533	42.6	

Table 1 (continued)

Characteristics	Study sample (N = 1252)		Polish general adult population (30.7 million) [19, 23]
	N	%	%
General perception of health (SF-1)			
Excellent	89	7.1	6.2
Very good	384	30.7	25.3
Good	566	45.2	44.3
Fair	190	15.2	20.3
Poor	22	1.8	3.9
Self-rated health using EQ-5D-5L			
11111	437	34.9	38.5
Any other health state	815	65.1	61.5
Self-rated health using EQ-VAS			
100	109	8.7	8.1
90–99	432	34.5	23.8
80–89	300	24.0	22.0
< 80	411	32.8	46.4
EQ VAS, mean (SD)	79.9	(16.9)	73.7 (19.9)
Any health problems within EQ-5D-5L dimension			
Mobility (MO)	360	28.8	25.8
Self-care (SC)	124	9.9	9.1
Usual activities (UA)	258	20.6	17.4
Pain/discomfort (PD)	668	53.4	52.2
Anxiety/depression (AD)	537	42.9	41.5
Household income (monthly, per person, €)			
≤ 200	207	16.5	Average 340
201–320	306	24.4	
321–500	296	23.6	
> 500	200	16.0	
Refuse to answer	243	19.4	

SD standard deviation, *VAS* visual analog scale

The average utility of the 55555 state in TTO was − 0.408 (33.5% at − 1). ESM 1, Online Resource 3 and Fig. 2 for the observed TTO values.

In total, 1040 health states (8.3%) were indicated by the respondents in the feedback module as not revealing their true preferences in hindsight and were removed, leaving 11,480 TTO valuations for modelling. Using the feedback module reduced the number of respondents with inconsistencies related to health state 55555, from an initial 49 (3.9%) to 22 (1.8%). In the DCE data (8764 DCE pairs), 36 respondents (2.9%) presented with a suspicious response pattern (choosing states on the left or right or regularly alternately) but were not excluded from the modelling.

3.3 Preferred Model (Polish EQ-5D-5L Value Set)

In the final model, the estimated decrease of utility for level 5 amounts to: 0.314 (MO), 0.264 (SC), 0.205 (UA),

0.575 (PD) and 0.232 (AD). For example, the relative weights of levels 2–4 in UA are: 0.112 (i.e. 0.023/0.205 in Table 2), 0.195 and 0.471, while in PD: 0.052, 0.087 and 0.455. The intermediate models differ slightly (Table 2). The disutilities increase when accounting for the impact of religion and censoring (both motivated by statistical criteria). In the final value set, we get u(22222) = 0.873, u(33333) = 0.800, u(44444) = 0.296, and u(55555) = − 0.590, as compared to u(22222) = 0.716 and u(33333) = − 0.523 in the Polish EQ-5D-3L tariff. We present the complete results, alongside more technical parameters, in ESM 1, Online Resource 4, a practical example of how to use a scoring algorithm to estimate the value for a health state in ESM 1, Online Resource 5 and all 3125 values for the Polish EQ-5D-5L value set, as well as an index calculator, in ESM 3.
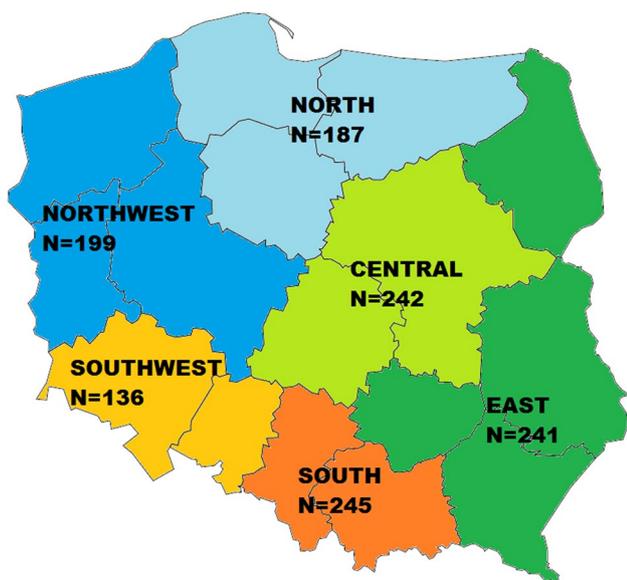
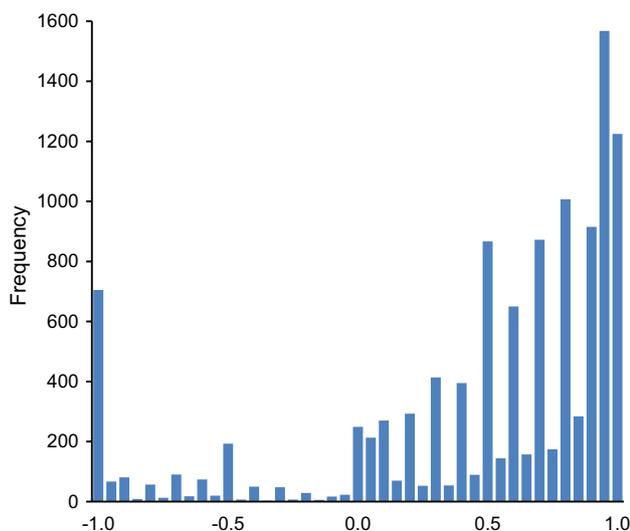**Fig. 1** Geographical representation of respondents in the Polish EQ-5D-5L valuation study

Fig. 2 Distribution of observed time trade-off (TTO) values

3.4 Comparison of Polish Value Sets

The kernel density plots (Fig. 3) and the utility values for the individual states (Fig. 4) illustrate the high degree of similarity between the three Polish value sets. The new descriptive system is also more sensitive to a slight worsening in health: in the results, there are more utility values close to 1. In all likelihood, because of a higher number of health states, the distribution of the utility value is unimodal for EQ-5D-5L, while bimodal for EQ-5D-3L (though this comparison is made within the domain of health states, not of individuals,

and thus it is a property of the descriptive system more than the value set itself). The new value set has a slightly lower worst utility (−0.590; −0.523, for both the EQ-5D-3L and the cross-walk), which is intuitive in view of the five-level system, but also in terms of accounting for the censoring and the bias from religious respondents in the present modelling.

The importance of the dimensions scarcely changed: PD, MO and SC are the most significant, followed by AD and UA in the present value set, whereas by UA and AD in the previous two value sets. In Table 3, additional descriptive statistics are presented. Importantly, some of them are primarily influenced by the descriptive system, rather than the value set (e.g. the percentage of states with negative utility depends on the utilities assigned to health states, and also on how many severe health states are present in a descriptive system).

4 Discussion

In this study, we followed an official EQ-VT protocol, performed over 1200 computer-assisted face-to-face interviews, collected TTO values for 86 EQ-5D-5L health states and DCE choices for 196 pairs of states, and estimated the Polish EQ-5D-5L value set using both elicitation tasks. Our final model accounted for random parameters (respondent heterogeneity), error scaling (greater noise for more severe states), censoring at −1, unwillingness to trade in TTO by religious participants and non-logit distribution in DCE. All these elements of the model were added in response to the statistical considerations. To the best of our knowledge, two elements are novel: the impact of religiosity and error scaling. We find the latter one rather intuitive; the variance of noise increasing with severity may partially explain why there is a weak relationship between the misery index and the disutility for the negative utility values [47]. The former element is probably the most controversial assumption in our model, and our decision to use it followed the reasoning presented in [45]. It is important to stress that correcting for the impact of religiosity does not aim at neglecting the preferences of religious individuals, but at correcting for how they may be biased in the TTO task (and how the elicitation task differs from what the resulting utilities are used for; not to actually shorten an individual's life but to trade-off benefits between different individuals).

There are two more arbitrary decisions we made in the modelling. First, we decided to combine TTO and DCE data. We believe that provided there is no consensus on whether one method is clearly better (not in terms of cost or ease of application but the quality of the results) using both is the safest approach. Second, we decided to use a simple model with no constant and no interaction terms. As mentioned above, that makes the final results more applicable to

Table 2 Modelling results

	Model 1 panel, random effects	Model 2 Bayesian	Model 3 M2 + random parameters	Model 4 M3 + error scaling with t-Student	Model 5 M4 + religion scaling	Final model M5 + DCE, censoring
Const.	0.005 (−0.010; 0.019)	Not used	Not used	Not used	Not used	Not used
MO2	0.021 (0.002; 0.039)	0.023 (0.001; 0.044)	0.058 (0.013; 0.073)	0.017 (0.014; 0.022)	0.019 (0.014; 0.023)	0.025 (0.020; 0.029)
MO3	0.012 (−0.007; 0.031)	0.016 (0.000; 0.036)	0.077 (0.021; 0.094)	0.015 (0.005; 0.026)	0.016 (0.005; 0.028)	0.034 (0.026; 0.042)
MO4	0.098 (0.077; 0.118)	0.101 (0.074; 0.129)	0.159 (0.071; 0.181)	0.101 (0.085; 0.116)	0.107 (0.090; 0.124)	0.126 (0.113; 0.141)
MO5	**0.262 (0.238; 0.285)**	**0.263 (0.239; 0.289)**	**0.303 (0.271; 0.330)**	**0.251 (0.228; 0.274)**	**0.267 (0.242; 0.293)**	**0.314 (0.286; 0.342)**
SC2	0.030 (0.014; 0.046)	0.037 (0.015; 0.059)	0.015 (0.003; 0.087)	0.029 (0.024; 0.034)	0.031 (0.026; 0.036)	0.031 (0.027; 0.036)
SC3	0.038 (0.017; 0.059)	0.042 (0.014; 0.071)	0.005 (0.000; 0.119)	0.037 (0.028; 0.047)	0.040 (0.029; 0.050)	0.047 (0.040; 0.055)
SC4	0.122 (0.098; 0.146)	0.116 (0.089; 0.143)	0.042 (0.027; 0.180)	0.108 (0.094; 0.123)	0.115 (0.099; 0.131)	0.111 (0.099; 0.123)
SC5	**0.276 (0.254; 0.298)**	**0.269 (0.244; 0.295)**	**0.242 (0.193; 0.268)**	**0.258 (0.237; 0.282)**	**0.273 (0.249; 0.299)**	**0.264 (0.243; 0.286)**
UA2	0.031 (0.014; 0.048)	0.034 (0.011; 0.058)	0.002 (0.000; 0.007)	0.033 (0.026; 0.039)	0.034 (0.028; 0.042)	0.023 (0.019; 0.027)
UA3	0.032 (0.009; 0.054)	0.041 (0.015; 0.067)	0.005 (0.000; 0.014)	0.050 (0.040; 0.060)	0.053 (0.043; 0.063)	0.040 (0.032; 0.048)
UA4	0.092 (0.070; 0.115)	0.088 (0.062; 0.115)	0.024 (0.010; 0.038)	0.104 (0.091; 0.117)	0.110 (0.095; 0.125)	0.097 (0.087; 0.107)
UA5	**0.186 (0.167; 0.206)**	**0.183 (0.157; 0.209)**	**0.180 (0.161; 0.201)**	**0.180 (0.161; 0.200)**	**0.190 (0.169; 0.212)**	**0.205 (0.188; 0.224)**
PD2	0.028 (0.012; 0.044)	0.033 (0.012; 0.054)	0.041 (0.028; 0.054)	0.025 (0.021; 0.028)	0.026 (0.022; 0.030)	0.030 (0.026; 0.034)
PD3	0.034 (0.014; 0.053)	0.035 (0.007; 0.063)	0.053 (0.036; 0.071)	0.030 (0.022; 0.039)	0.032 (0.022; 0.041)	0.050 (0.043; 0.058)
PD4	0.229 (0.208; 0.251)	0.228 (0.204; 0.254)	0.253 (0.224; 0.276)	0.223 (0.208; 0.239)	0.235 (0.217; 0.253)	0.261 (0.244; 0.280)
PD5	**0.467 (0.440; 0.494)**	**0.473 (0.446; 0.499)**	**0.490 (0.464; 0.518)**	**0.492 (0.463; 0.520)**	**0.519 (0.485; 0.555)**	**0.575 (0.538; 0.613)**
AD2	0.024 (0.006; 0.041)	0.032 (0.010; 0.054)	0.049 (0.015; 0.061)	0.019 (0.016; 0.023)	0.020 (0.017; 0.024)	0.018 (0.015; 0.021)
AD3	0.034 (0.011; 0.056)	0.033 (0.006; 0.058)	0.085 (0.038; 0.101)	0.037 (0.026; 0.049)	0.039 (0.027; 0.052)	0.029 (0.022; 0.037)
AD4	0.114 (0.094; 0.135)	0.114 (0.088; 0.139)	0.160 (0.116; 0.181)	0.119 (0.106; 0.132)	0.126 (0.113; 0.142)	0.108 (0.097; 0.119)
AD5	**0.224 (0.203; 0.244)**	**0.226 (0.201; 0.251)**	**0.176 (0.153; 0.231)**	**0.211 (0.194; 0.229)**	**0.223 (0.204; 0.243)**	**0.232 (0.213; 0.252)**
Deviance	61.2% (R^2 used instead)	11,866	−777	−13,781	−13,780	−9215
DIC		11,886	2597	−9704	−9704	−9215[a]
PSRF	n.a.	All <1.01	Maximum = 15	All <1.01	All <1.01	All <1.01
Maximum u (not 11111)	0.983	0.984	0.998	0.985	0.984	0.982
u (22222)	0.862	0.841	0.834	0.877	0.870	0.873
u (33333)	0.847	0.833	0.775	0.830	0.821	0.800
u (44444)	0.340	0.352	0.361	0.345	0.307	0.296
u (55555)	−0.420	−0.415	−0.391	−0.392	−0.471	−0.590
% states $u < 0$	2.85	2.88	2.69	2.78	4.26	6.66
Dimension order	PD, SC, MO, AD, UA	PD, SC, MO, AD, UA	PD, MO, SC, UA, AD	PD, SC, MO, AD, UA	PD, SC, MO, AD, UA	PD, MO, SC, AD, UA
Levels consistency	MO3 < MO2	MO3 < MO2	SC3 < SC2	MO3 < MO2	MO3 < MO2	Consistent

AD anxiety/depression, *DCE* discrete choice experiment, *DIC* deviance information criterion, *M* model, *MO* mobility, *n.a.* PD pain/discomfort, *PSRF* potential scale reduction factor, *SC* self-care, *u* utility, *UA* usual activities

[a]Failed to calculate penalty in JAGS ("support of observed nodes is not fixed")

situations where only limited information is available (e.g. only marginal distributions of levels in individual dimensions). To represent respondents' answers more accurately (in the sense of predictive validity), a more complex model would probably have to be used (e.g. accounting for a non-linear time preference [44]). In this sense, there is a trade-off between trying to represent the data faithfully and using a specification that can be subsequently easily used.

The assumptions resulted in the theoretical value of u(55555) = −0.590, visibly lower than the average utility elicited in TTO, i.e. −0.408. This difference stems from three elements of our model. First, censoring leads to interpreting observed −1s as effectively possibly much lower than −1 (33.5% of TTO tasks for 55555 ended by assigning −1). Second, introducing the impact of religiosity in TTO tasks results in effectively assuming that the true disutility is larger than the observed one. Third, by considering the random noise as having a larger variance for severe states, we make the parameters less driven by the actual observations for the severe states. Nevertheless, the final utility for the pit state is similar to the one in the
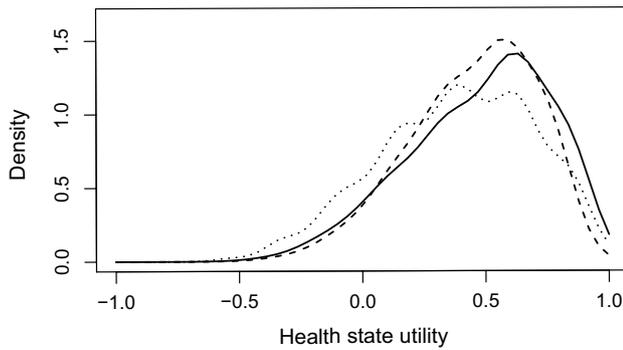
Fig. 3 Kernel density functions for the three Polish value sets (EQ-5D-5L directly measured is indicated by a solid line; EQ-5D-5L cross-walk is indicated by a dashed line; EQ-5D-3L is indicated by a dotted line)
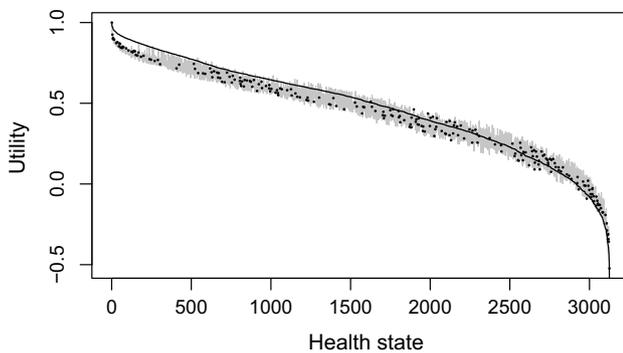

Fig. 4 Utility values for all states from three Polish value sets ordered by EQ-5D-5L (EQ-5D-5L directly measured is indicated by a solid line; EQ-5D-5L cross-walk is indicated by a solid light grey line, EQ-5D-3L is indicated by dots)

EQ-5D-3L value set (hence, the cross-walk), and the slight decrease is intuitive in view of the larger number of levels.

Regarding the final value set, despite the fact that it describes significantly more possible health states (3125 vs. 243), it is similar to the Polish EQ-5D-3L value set in terms of a minimum utility, the range of values and the order of three most important dimensions [16]. The resemblance between the general characteristics of both value sets should support the comparability of Health State Utility Values obtained with these two types of EQ-5D questionnaire, and consequently the comparability of the results from economic analyses and the reimbursement decisions made, what was questioned in some other countries, such as the UK [48, 49]. What differentiates our study from the previous Polish valuation is greater attention to sampling, which resulted in a study group similar to Polish society as a whole, in terms of a higher number of demographic features (geographical spread in the first instance, but also employment status and size of locality).

In similarity to some other EQ-5D-5L valuations performed in developed countries, we noted the relative increase in the importance of the anxiety/depression dimension, in comparison to former EQ-5D-3L valuation studies. We suppose that this is primarily a consequence of a change in health state preferences over the period of one or two decades separating the valuation studies, rather than the effect of different wording in the EQ-5D-5L questionnaire. We may observe this phenomenon in England, Germany, the Netherlands, Spain and Japan [34, 36–38, 50, 51], whereas in lower income countries, such as Uruguay or The Philippines, anxiety/depression remains the least important domain [52, 53]. In addition to this observation, the predominance of the mobility dimension in Asian countries (China, Hong Kong, Indonesia, Japan, South Korea and Thailand) merits further investigation [39, 54–57]. Some changes in the dimension weightings may also be subject to change in the descriptive system: in the Polish version, the wording for mobility has been changed from 'confined to bed' to 'extreme problems'.

Taking into account the number of CEE countries (20) and the relatively low gross domestic product these countries have, the objective of searching for simpler and inexpensive valuation protocols acquires further significance. Discrete choice experiment-based valuations performed online constitute a potential solution, although certain methodological challenges still have to be dealt with [58, 59]. In the meantime, researchers from the CEE region frequently face the dilemma: 'what EQ-5D value set should I use in the absence of a national value set?' According to the results of the recent review, in the case of EQ-5D-3L, CEE researchers mostly prefer the UK Measurement and Valuation of Health study tariff [60, 61]. In the case of EQ-5D-5L, the choice will be harder, as the EQ-5D-5L value set for England has faced criticism and is still not supported by the National Institute for Health and Care Excellence [48]. Slovenian researchers may use the cross-walk approach based on their visual analog scale-based EQ-5D-3L value set, but recommendations for scientists from other CEE countries are far from straightforward [21]. Nevertheless, they should at least consider using either the Polish or the forthcoming Hungarian EQ-5D-5L value sets, as CEE countries share some common cultural and historical background.

5 Conclusions

We developed the TTO and DCE-based EQ-5D-5L value set for Poland. It will complement the existing Polish EQ-5D-3L value set and will further stimulate the development of local quality-of-life research and the use of health technology assessment in decision making within the healthcare sector. While the new EQ-5D-5L value set offers more sensitivity, it also provides ground for consistency

Table 3 Comparison of three Polish EQ-5D value sets

	Polish EQ-5D-5L value set	Polish EQ-5D-5L crosswalk value set	Polish EQ-5D-3L value set
Valuation method	Hybrid (TTO/DCE)	Crosswalk (TTO)	TTO
Dimensions ordering, from the most to the least important (disutility for the worse level within a dimension)	PD (-0.575) MO (-0.314) SC (-0.264) AD (-0.232) UA (-0.205)	PD (-0.489)[a] MO (-0.331) SC (-0.235) UA (-0.212) AD (-0.207)	PD (-0.489)[a] MO (-0.331) SC (-0.235) UA (-0.212) AD (-0.207)
Number of health states	3125	3125	243
Maximum value (11111)	1	1	1
Second highest value (health state)	0.982 (11112)	0.940 (11112)	0.925 (11112)
Mean value (SD)	0.476 (0.286)	0.448 (0.253)	0.382 (0.310)
Median value (Q1–Q3)	0.523 (0.286–0.692)	0.483 (0.282–0.642)	0.406 (0.155–0.630)
Minimum value (health state)	-0.590 (55555)	-0.523 (55555)	-0.523 (33333)
Value for 22222	0.862	0.760	0.716
Value for 33333	0.721	0.716	-0.523
Value for 44444	0.173	0.336	n.a.
Health states ≥ 0.8, n (%)	160 (5.1)	163 (5.2)	22 (9.1)
Health states worse than dead (<0), n (%)	137 (4.4)	124 (4.0)	32 (13.2)

AD anxiety/depression, *DC*E discrete choice experiment, *MO* mobility, *n.a.*, *PD* pain/discomfort, *SC* self-care, *SD* standard deviation, *TTO* time trade-off, *UA* usual activities

[a]Disutilities for dimensions, not including the constant (-0.049)

of past and future decisions. The presented EQ-5D-5L value set may be considered as an option by researchers from CEE countries who lack their own national health preference data.

Compliance with Ethical Standards

Conflict of Interest Dominik Golicki and Michał Jakubczyk are members of the EuroQol Research Foundation (the copyright holders of the EQ-5D-5L) and have received grants from the EuroQol Research Foundation. Katarzyna Graczyk and Maciej Niewada have no conflicts of interest that are directly relevant to the content of this article.

Ethics Approval The study was approved by the Ethics Committee of the Medical University of Warsaw, Warsaw, Poland (AKBE/137/16).

Data Sharing The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

1. Dimova A, Rohova M, Atanasova E, Kawalec P, Czok K. Drug policy in Bulgaria. Value Health Reg Issues. 2017;13:50–4.
2. Skoupá J. Drug policy in the Czech Republic. Value Health Reg Issues. 2017;13:55–8.
3. Inotai A, Csanádi M, Harsányi A, Németh B. Drug policy in Huny. Value Health Reg Issues. 2017;13:16–22.
4. Culig J, Antolic S, Szkultecka-Dębek M. Drug policy in Croatia. Value Health Reg Issues. 2017;13:27–30.
5. Jahnz-Różyk K, Kawalec P, Malinowski K, Czok K. Drug policy in Poland. Value Health Reg Issues. 2017;13:23–6.
6. Biuletyn Informacji Publicznej Agencji Oceny Technologii Medycznych i Taryfikacji. Available from: http://bipold.aotm.gov.pl/. Accessed 10 Jan 2019.
7. Jakubiak-Lasocka J, Jakubczyk M. Cost-effectiveness versus cost-utility analyses: what are the motives behind using each and how do their results differ? A Polish example. Value Health Reg Issues. 2014;4:66–74.

8. Health Technology Assessment Guidelines (Version 3.0). Warsaw: Agencja Oceny Technologii Medycznych i Taryfikacji; 2016.

9. The Euroqol Group. EuroQol: a new facility for the measurement of health-related quality of life. Health Policy. 1990;16:199–208.

10. Brooks R. EuroQol: the current state of play. Health Policy. 1996;37:53–72.

11. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20:1727–36.

12. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. 4th ed. New York: Oxford University Press; 2015.

13. Buchholz I, Janssen MF, Kohlmann T, Feng YS. A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. Pharmacoeconomics. 2018;36:645–61.

14. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. Pharmacoeconomics. 2018;36:675–97.

15. Prevolnik Rupel V, Ogorevc M. The EQ-5D health states value set for Slovenia. Zdravstveno Varstvo. 2012;51:128–40.

16. Golicki D, Jakubczyk M, Niewada M, Wrona W, Busschbach JJ. Valuation of EQ-5D health states in Poland: first TTO-based social value set in Central and Eastern Europe. Value Health. 2010;13:289–97.

17. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2013;22:1717–27.

18. Golicki D, Niewada M, Karlinska A, Buczek J, Kobayashi A, Janssen MF, et al. Comparing responsiveness of the EQ-5D-5L, EQ-5D-3L and EQ VAS in stroke patients. Qual Life Res. 2015;24:1555–63.

19. Golicki D, Niewada M, Buczek J, Karlinska A, Kobayashi A, Janssen MF, et al. Validity of EQ-5D-5L in stroke. Qual Life Res. 2015;24:845–50.

20. Golicki D, Niewada M. EQ-5D-5L Polish population norms. Arch Med Sci. 2017;13:191–200.

21. van Hout B, Janssen MF, Feng YS, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health. 2012;15:708–15.

22. Golicki D, Niewada M, Hout BV, Janssen MF, Pickard AS. Interim EQ-5D-5L value set for Poland: first crosswalk value set in Central and Eastern Europe. Value Health Reg Issues. 2014;4:19–23.

23. Xie F, Pickard AS, Krabbe PF, Revicki D, Viney R, Devlin N, et al. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). Pharmacoeconomics. 2015;33:867–77.

24. Central Statistical Office. Demographic yearbook of Poland. Warsaw: Zakład Wydawnictw Statystycznych; 2015.

25. Ramos-Goni JM, Oppe M, Slaap B, Busschbach JJ, Stolk E. Quality control process for EQ-5D-5L valuation studies. Value Health. 2017;20:466–73.

26. Janssen BM, Oppe M, Versteegh MM, Stolk EA. Introducing the composite time trade-off: a test of feasibility and face validity. Eur J Health Econ. 2013;14(Suppl. 1):S5–13.

27. Devlin NJ, Krabbe PF. The development of new research methods for the valuation of EQ-5D-5L. Eur J Health Econ. 2013;14(Suppl. 1):S1–3.

28. Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goni JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics. 2016;34:993–1004.

29. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. Health Econ. 2011;20:348–61.

30. Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health. 2014;17:445–53.

31. Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. Value Health. 2010;13:1005–13.

32. Ramos-Goni JM, Rivero-Arias O, Errea M, Stolk EA, Herdman M, Cabases JM. Dealing with the health state 'dead' when using discrete choice experiments to obtain values for EQ-5D-5L health states. Eur J Health Econ. 2013;14(Suppl. 1):S33–42.

33. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30:473–83.

34. Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level Version of EQ-5D. Value Health. 2016;19:343–52.

35. Craig BM, Rand K. Choice defines QALYs: a US valuation of the EQ-5D-5L. Med Care. 2018;56:529–36.

36. Ramos-Goni JM, Craig BM, Oppe M, Ramallo-Farina Y, Pinto-Prades JL, Luo N, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. Value Health. 2018;21:596–604.

37. Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. Pharmacoeconomics. 2018;36:663–74.

38. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. Health Econ. 2018;27:7–22.

39. Purba FD, Hunfeld JAM, Iskandarsyah A, Fitriana TS, Sadarjoen SS, Ramos-Goni JM, et al. The Indonesian EQ-5D-5L value set. Pharmacoeconomics. 2017;35:1153–65.

40. Kruschke J. Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan. 2nd ed. Oxford: Academic Press; 2014.

41. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. J Comput Graph Stat. 1997;7:434–55.

42. Jia YX, Cui FQ, Li L, Zhang DL, Zhang GM, Wang FZ, et al. Comparison between the EQ-5D-5L and the EQ-5D-3L in patients with hepatitis B. Qual Life Res. 2014;23:2355–63.

43. Rand-Hendriksen K, Ramos-Goñi JM, Augestad LA, Luo N. Less is more: cross-validation testing of simplified nonlinear regression model specifications for EQ-5D-5L health state values. Value Health. 2017;20:945–52.

44. Jakubczyk M, Craig BM, Barra M, Groothuis-Oudshoorn CGM, Hartman JD, Huynh E, et al. Choice defines value: a predictive modeling competition in health preference research. Value Health. 2018;21:229–38.

45. Jakubczyk M, Golicki D, Niewada M. The impact of a belief in life after death on health-state preferences: true difference or artifact? Qual Life Res. 2016;25:2997–3008.

46. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012;31:306–18.

47. Rand-Hendriksen K, Augestad LA, Dahl FA, Kristiansen IS, Stavem K. A shortcut to mean-based time tradeoff tariffs for the EQ-5D? Med Dec Mak. 2012;32:569–77.

48. Hernandez Alava M, Wailoo A, Grimm S, Pudney S, Gomes M, Sadique Z, et al. EQ-5D-5L versus EQ-5D-3L: the impact on cost effectiveness in the United Kingdom. Value Health. 2018;21:49–56.

49. Devlin N, Brazier J, Pickard AS, Stolk E. 3L, 5L, What the L? A NICE conundrum. Pharmacoeconomics. 2018;36:637–40.

50. Shiroiwa T, Ikeda S, Noto S, Igarashi A, Fukuda T, Saito S, et al. Comparison of value set based on DCE and/or TTO data: scoring for EQ-5D-5L health states in Japan. Value Health. 2016;19:648–54.

51. Selivanova A, Buskens E, Krabbe PFM. Head-to-head comparison of EQ-5D-3L and EQ-5D-5L health values. Pharmacoeconomics. 2018;36:715–25.

52. Augustovski F, Rey-Ares L, Irazola V, Garay OU, Gianneo O, Fernandez G, et al. An EQ-5D-5L value set based on Uruguayan population preferences. Qual Life Res. 2016;25:323–33.

53. Lam H, Purba F, Rivera A, Miguel RT, Cheng KJ. Same person, different languages, different health preferences. Discrete choice experiments (DCE) comparison from the bilinguals of the EQ-5D-5L valuation in the Philippines. In: Presented at the 35th Euro-Qol Group Scientific Plenary, Lisbon, 21 Sept 2018.

54. Luo N, Liu G, Li M, Guan H, Jin X, Rand-Hendriksen K. Estimating an EQ-5D-5L value set for China. Value Health. 2017;20:662–9.

55. Wong ELY, Ramos-Goni JM, Cheung AWL, Wong AYK, Rivero-Arias O. Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. Patient. 2018;11:235–47.

56. Kim SH, Ahn J, Ock M, Shin S, Park J, Luo N, et al. The EQ-5D-5L valuation study in Korea. Qual Life Res. 2016;25:1845–52.

57. Pattanaphesaj J, Thavorncharoensap M, Ramos-Goni JM, Tongsiri S, Ingsrisawang L, Teerawattananon Y. The EQ-5D-5L valuation study in Thailand. Expert Rev Pharmacoecon Outcomes Res. 2018;18:551–8.

58. Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. Soc Sci Med. 2014;114:38–48.

59. Mulhern B, Longworth L, Brazier J, Rowen D, Bansback N, Devlin N, et al. Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. Value Health. 2013;16:104–13.

60. Rencz F, Gulacsi L, Drummond M, Golicki D, Prevolnik Rupel V, Simon J, et al. EQ-5D in Central and Eastern Europe: 2000–2015. Qual Life Res. 2016;25:2693–710.

61. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35:1095–108.

Affiliations

Dominik Golicki[1,3] · Michał Jakubczyk[2,3] · Katarzyna Graczyk[3] · Maciej Niewada[1,3]

Michał Jakubczyk
michal.jakubczyk@sgh.waw.pl

Katarzyna Graczyk
katarzyna.graczyk88@gmail.com

Maciej Niewada
maciej.niewada@wum.edu.pl

[1] Department of Experimental and Clinical Pharmacology, Medical University of Warsaw, Banacha 1b St., 02-097 Warsaw, Poland

[2] Decision Analysis and Support Unit, SGH Warsaw School of Economics, Al. Niepodległości 162, 02-554 Warsaw, Poland

[3] HealthQuest Spółka z ograniczoną odpowiedzialnością Sp. k., 63 Mickiewicza Street, Megadex A Building, 01-625 Warsaw, Poland