1 Respondents own health

The distribution of own health state was as follows: 11111 (i.e. perfect health, 63.6%), 11112 (16.3%), 11122 (8.7%), 11121 (7.1%), 11222 (1.6%), 21121 (1.1%), and 11211, 11221, and 21111 (each 0.5%); level 3 was not observed. The average conviction about EQ-5D-3L level assigned decreases with level (p-value < 0.001, all dimensions pooled, t-Student test for levels 1 & 2). This result is probably due to the respondents with no health problems being entirely convinced (out of 7 respondents with VAS=100, only in one dimension in one respondent the conviction was 90, not 100). To account for this background health impact, we estimated the linear model explaining the conviction with VAS and a dummy if level equals 2 (standard errors in parentheses):

$$\widehat{\text{conviction}} = \underset{(2.88)}{85.3} - \underset{(0.97)}{12.95} \times [\text{level} = 2] + \underset{(0.03)}{0.13} \times \text{VAS}. \tag{1}$$

The results suggest that the act of selecting level 2 is associated with a decrease in conviction, the overall health state (as measured by VAS) held constant. This might mean that the wording of level 2 is less precise than of level 1. Due to few level 2 answers, we did not analyse the dimensions separately (or level 3). With more data and a less healthy population it might be possible to test how precise answers a given descriptive system allows.

The range of *epas* amounted to 78.9–90.4, on average[1]; the range of *spas* amounted to 70.8–93.4. The point VAS valuation on average was in the middle of the *epas* (p-value=0.6, t-Student test for the relative position, $H_0 : \mu = 0.5$)[2], and was located in the upper half of *spas* (mean of 0.57, p-value $<$ 0.001).

2 Data cleaning & quality checks

The interviews were entered into MS Excel by interviewers and pooled. The data quality checks (e.g. consistency checks, whether *spas* was larger than *epas*, and whether both contained the point answer) were performed; in doubts, paper forms were re-examined. Doubts appeared in 27 cases (out of 1840 non-warm-up TTO experiments, $< 1.5\%$) and were mostly (17 cases) solved. The remaining 10 exercises were removed from subsequent analysis (in 1 case the state was perceived implausible and no answer was provided; in 2 cases the respondent misinterpreted the question in several initial experiments; in 7 cases the ranges were inconsistent).

We checked the interviewer effect on point valuation and *epas*/*spas* length by random effects (RE) models (RE across interviewees). To test this effect on point valuation, we used a linear model with standard explanatory variables (i.e., dummies for dimensions/levels) plus dummies for individual pollsters. No

[1] Three cases were removed for ranges analysis, as the *epas* was not contained in *spas*.

[2] Zero-length *epas* removed (substituting 0.5 does not change the result)

Level	%	mean ± SD (min, max)
Mobility		
1: I have no problems in walking about	98.4%	98.1 ± 4.45 (75, 100)
2: I have some problems in walking about	1.6%	91.7 ± 8.50 (80, 100)
3: I am confined to bed	0.0%	—
Self-care		
1: I have no problems with self-care	100.0%	99.5 ± 2.00 (86, 100)
2: I have some problems washing or dressing myself	0.0%	—
3: I am unable to wash or dress myself	0.0%	—
Usual activities		
1: I have no problems with performing my usual activities	97.3%	97.3 ± 5.88 (70, 100)
2: I have some problems with performing my usual activities	2.7%	88.0 ± 11.66 (70, 100)
3: I am unable to perform my usual activities	0.0%	—
Pain / Discomfort		
1: I have no pain or discomfort	81.0%	94.3 ± 8.13 (50, 100)
2: I have moderate pain or discomfort	19.0%	82.0 ± 19.72 (15, 100)
3: I have extreme pain or discomfort	0.0%	—
Anxiety / Depression		
1: I am not anxious or depressed	73.4%	92.7 ± 10.59 (50, 100)
2: I am moderately anxious or depressed	26.6%	83.2 ± 17.98 (35, 100)
3: I am extremely anxious or depressed	0.0%	—

Table 1 Own-health assessment for five dimensions: % of respondents and mean conviction for the level selected (0–100 scale). SD = standard deviation

interviewer-effect on the point TTO result was found (the minimal p-value for an interviewer amounted to 0.175).

To test the interviewer-effect on the range length a similar model was used with two modifications: dummies for entire health states were used (not for dimension-level, as the effect of dimensions/levels on ranges was not verified to be additive), and also a semi-log model was used (to account for heavy positive skew in the endogenous variable)[3] Out of all the four models (hence, 32 p-values) in 5 cases the p-value was < 0.05 (minimal p-value of 0.009). Due to the multiplicity of hypotheses, we concluded no interviewer effect.

3 Raw results for somewhat plausible answers (spas)

Table 2 presents the descriptive statistics for spas.

4 Sub-additivity of imprecision

In each block, there was a pair of states which when compounded yield another state from that block (11311 and 11112 yielding 11312, 11131 and 11113

[3] With 0.004 added, half of the minimal strictly positive value, to avoid log(0) problem.

State	range of means	mean length	% zero length	% including 1 (0)	% spas=epas
11112	(0.800; 0.938)	0.138	16.5%	60.4% (1.1%)	38.5%
11113	(0.475; 0.704)	0.229	7.4%	27.7% (5.3%)	23.4%
11121	(0.829; 0.955)	0.126	25.0%	58.7% (0.0%)	38.0%
11131	(0.336; 0.543)	0.207	15.4%	25.3% (3.3%)	33.0%
11133	(0.148; 0.381)	0.233	12.1%	19.8% (6.6%)	19.8%
11211	(0.759; 0.927)	0.168	15.2%	55.4% (1.1%)	35.9%
11311	(0.468; 0.669)	0.201	11.1%	27.2% (2.8%)	24.4%
11312	(0.399; 0.607)	0.208	8.7%	18.5% (3.3%)	23.9%
12111	(0.774; 0.931)	0.157	14.1%	55.4% (1.1%)	41.3%
13311	(0.176; 0.413)	0.237	13.5%	12.4% (6.7%)	22.5%
21111	(0.834; 0.954)	0.119	20.7%	60.9% (0.0%)	37.0%
21121	(0.711; 0.885)	0.174	8.7%	42.4% (0.0%)	28.3%
22222	(0.427; 0.626)	0.199	7.1%	20.2% (1.1%)	24.0%
23232	(-0.135; 0.050)	0.185	12.1%	6.6% (2.2%)	28.6%
32211	(0.256; 0.474)	0.218	12.0%	14.2% (7.7%)	23.0%
32223	(-0.063; 0.150)	0.213	12.9%	5.4% (7.2%)	26.9%
32313	(-0.200;-0.014)	0.186	9.8%	7.6% (8.7%)	32.6%

Table 2 Raw results for somewhat plausible answers (spas). (epas = equally plausible answers)

yielding 11133). We checked whether the amount of imprecision is additive, i.e. if the length of epas/spas for the compound state equals on average the sum of the lengths of the building states (paired t-Student test). In all the cases (two blocks, epas and spas), the imprecision for the compound state was smaller: the mean difference in length amounted to -0.118 (11133, epas), -0.198 (11133, spas), -0.076 (11312, epas), and -0.138 (11312, spas).

We built linear models (four types: for epas and spas, for compound states 11312 and 11133) explaining the amount of imprecision for the compound state by (two specifications, no intercept) either the sum of epas/spas lengths for building states or the two individual lengths for building states. To remove the dilution effect (as the regressors are also observed with error; see Frost and Thompson, 2000) we used the total least square regression via principal component analysis. The results in Table 3 show the subadditivity: the imprecision in the compound state is lower than the combined imprecision in the building states. For 11133, the overall imprecision is smaller than the average imprecision of the combined states (the coefficient in the first specification is < 0.5); hence, is reduced in this sense (and the imprecision of 11131 weigh more than of 11113). The large coefficient by 11112 may result from the fact that the overall imprecision for this mild state is small.

Parameter	Explanatory variable			
	epas		spas	
	11312	11133	11312	11133
sum of lengths	0.716	0.378	0.517	0.437
length of 11311	0.143	—	0.631	—
length of 11112	1.576	—	0.513	—
length of 11131	—	0.977	—	0.740
length of 11113	—	0.088	—	0.404

Table 3 Explaining the length of equally/somewhat plausible answers (epas/spas) for a compound state (11312, 11133) by the lengths of building states.

References

Frost C, Thompson S (2000) Correcting for Regression Dilution Bias: Comparison of Methods for a Single Predictor Variable. Journal of the Royal Statistical Society Series A (Statistics in Society) 163:173–189

Golicki D, Jakubczyk M, Niewada M, Wrona W, Busschbach J (2010) Valuation of EQ-5D Health States in Poland: First TTO-Based Social Value Set in Central and Eastern Europe. Value in Health 13:289–297