



An EQ-5D-5L Value Set for Belgium

Nicolas Bouckaert¹ · Irina Cleemput¹ · Stephan Devriese¹ · Sophie Gerkens¹

Accepted: 5 July 2022 / Published online: 4 August 2022
© The Author(s) 2022

Abstract

Objective This study aimed to establish a Belgian EQ-5D-5L value set based on the preferences of the adult Belgian general population.

Methods The most recent EuroQol Valuation Technology (EQ-VT 2.1) protocol for EQ-5D-5L valuation studies was followed. Computer-assisted personal interviews were carried out in a representative sample of the adult Belgian population. Potential respondents were randomly selected from the National Register using a multistage, stratified, cluster sampling with unequal probability design. Each respondent valued 10 or 11 health states using composite time trade-off (cTTO) and 14 health states in seven paired choice tasks using a discrete choice experiment (DCE). Different model specifications were explored and assessed based on logical consistency, goodness of fit, predictive accuracy and theoretical considerations.

Results A total of 892 respondents were included in the analyses. The sample was representative of the Belgian adult population in terms of age, sex, region of residence, educational attainment, labour market status, self-assessed health status and health-related quality of life (HRQoL). The preferred model specification was a hybrid (DCE and cTTO data combined) multiplicative eight-coefficient model with intercept random effects and correction for heteroskedasticity. Values range from –0.532 to 1. Loss of HRQoL is highest in the dimension pain/discomfort, closely followed by anxiety/depression.

Conclusions This study developed a Belgian EQ-5D-5L value set, based on the preferences of the Belgian adult general population. It provides opportunities for future clinical and economic evaluations in healthcare, for the measurement of patient-reported outcomes and for population health assessments.

Key Points for Decision Makers

This study established an EQ-5D-5L value set for Belgium, based on the preferences of a random sample of the Belgian general public in 2018–2020.

The value set allows researchers and decision makers to value health states described by means of the EQ-5D-5L. It can be employed to assess the impact of health interventions on health-related quality of life in cost-utility analyses or to value patient-reported outcomes measured in trials using the EQ-5D-5L.

The current value set should replace the former cross-walk value set used in Belgium.

1 Introduction

In Belgium, the submission of a health economic evaluation has been mandatory since 2002 for reimbursement requests concerning drugs with a claimed added therapeutic value compared to existing alternatives [1, 2]. Guidelines for economic evaluations were first developed in 2008 and updated in 2012 by the Belgian Health Care Knowledge Centre (KCE). Belgium is one of many countries to recommend the use of the EQ-5D to measure health-related quality of life (HRQoL) and to calculate quality-adjusted life years (QALYs) [1, 3].

The EQ-5D is a generic multi-attribute utility instrument with five dimensions to measure health and HRQoL: mobility (MO), self-care (SC), usual activities (UA), pain/discomfort (PD) and anxiety/depression (AD). Depending on the version of the instrument, each health dimension has either three levels of severity (EQ-5D-3L; 3L: no, some/moderate, extreme problems/unable to) or five levels of severity (EQ-5D-5L; 5L: no, slight, moderate, severe, extreme problems/

✉ Nicolas Bouckaert
nicolas.bouckaert@kce.fgov.be

¹ Belgian Health Care Knowledge Centre (KCE), Boulevard du Jardin Botanique 55, 1000 Bruxelles, Belgium

unable to). By increasing the number of response categories per dimension, the five-level version can describe substantially more health states than the earlier three-level version (3125 vs 243), which has improved the discriminatory power of the EQ-5D and reduced its ceiling effects [4–8].

An EQ-5D value set assigns a utility value to each health state that can be described with the EQ-5D. A utility value reflects the HRQoL of a health state, based on the preferences of the general public. It is anchored on a scale with 1 equal to full health and 0 equal to dead [9, 10]. Health states considered worse than dead have negative utility values [11].

While a value set for the EQ-5D-3L has existed since 2003 [12], no value set specific to the EQ-5D-5L was available for Belgium. As an intermediary solution, an EQ-5D-5L “crosswalk” value set could be derived from the EQ-5D-3L value set using the mapping procedure proposed by van Hout et al. [13]. However, to benefit from the full potential of the EQ-5D-5L, a value set specific to the five-level version was necessary. The objective of this study was to derive a Belgian EQ-5D-5L value set based on preferences of a representative sample of the adult Belgian population.

2 Methods

This study followed the most recent EuroQol Valuation Technology (EQ-VT 2.1) protocol for EQ-5D-5L valuation studies [14, 15]. The research protocol was approved by the Ethical Committee of the University hospital of Ghent (B670201835610). The reporting of this study follows the CREATE guidelines for valuation studies of multi-attribute utility-based instruments [16].

2.1 Sample Selection

Following Oppe and van Hout [17], a sample of 1000 respondents with a valid interview was targeted to ensure precision in the model estimates. Potential respondents aged 18–100 years old were randomly selected using a multistage, stratified, cluster sampling with unequal probability design, comparable in design to Demarest et al. [18] (more details in the Electronic Supplementary Material: ESM1). In the first stage, a geographical stratification was made, with random selection of municipalities in each province proportional to the provincial population size. Municipalities can be drawn multiple times, and for each draw, ten interviews were allocated. Second, interviews were assigned to a combination of sex and age group (18–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, 90–99), proportional to the provincial demographic composition. Figure 1 presents the selected municipalities and the number of assigned interviews. A good geographic coverage was obtained.

To deal with non-participation and low-quality interviews, ten potential candidates were selected for each targeted interview from the National Register, which lists all persons officially residing in Belgium. Institutionalized individuals were excluded.

Post-stratification weights (detailed in ESM1) were used to correct for differences between planned and realised interviews.

2.2 Valuation Interview

Professional interviewers from a market research company conducted face-to-face interviews using computer-assisted personal interview (CAPI) software developed by EuroQol for EQ-5D-5L valuation studies [15]. The software included interview scripts translated to Dutch and French by a professional translation agency and reviewed by the research team. At the end of the study, because of a transfer to a new EQ-VT platform, it appeared that in the Dutch version, severity level 2 in dimension MO was incorrectly labelled, i.e. “lopen” was used instead of “rondwandelen”. This was due to the fact that “lopen” and “rondwandelen” are Dutch for “walking about” used in, respectively, the Netherlands and Belgium. The impact of the incorrect labelling was expected to be limited because interviewers also had a printed questionnaire with the correct labels to assist respondents in assessing the severity of the health states described on the screen.

Each interview followed the same structure: (1) introduction to the study, (2) basic background questions (age, sex, health status using EQ-5D-5L descriptive system and EuroQol visual analogue scale (VAS), experiences of severe illness), (3) valuation with composite time trade-off (cTTO), (4) valuation with discrete choice experiment (DCE), (5) country-specific background questions (educational attainment, labour market position, self-assessed health).

2.2.1 Techniques for Eliciting Preferences

Two elicitation techniques were used—cTTO and DCE—to gain distinct but complementary information on the respondent’s preferences.

The cTTO valuation applies the standard time trade-off (TTO) for health states considered better than dead and shifts to lead-time TTO for health states valued worse than dead by the respondent [19] (more detailed descriptions in [11, 19–21]). The resulting cTTO values range between –1 (trading all lead time) and 1 (trading no time) with 0.05 increments. After receiving instructions, the interviewer familiarized the respondent with cTTO by means of an example state “being in a wheelchair”, followed by the evaluation of three practice health states and ten real EQ-5D-5L health states. In a feedback module after finalizing the cTTO task, the

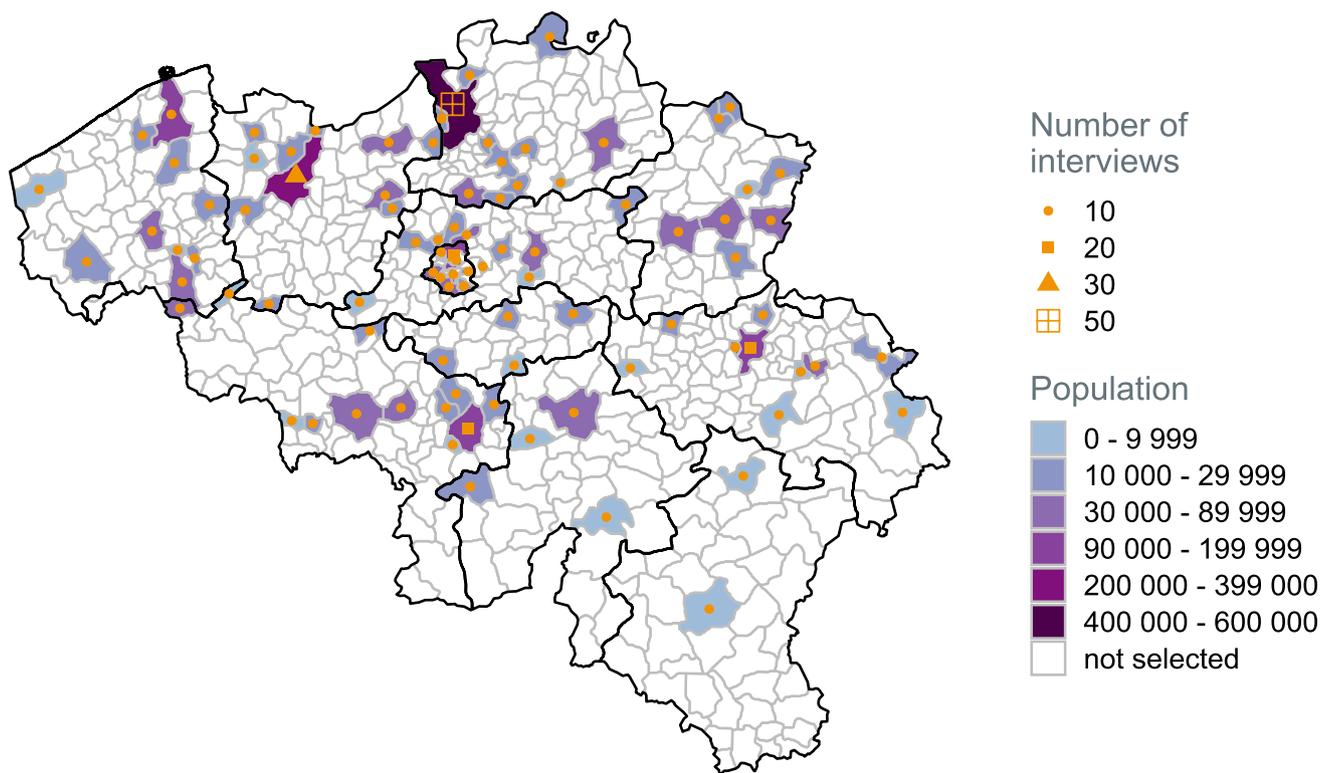


Fig. 1 Selected municipalities and number of interviews (map of Belgium)

respondent was presented with a ranking of the health states based on his/her responses and could mark health states for which the relative rank was deemed unsatisfactory.

In the DCE task, respondents were asked to indicate their preference between two EQ-5D-5L health states (forced choice). After receiving instructions, each respondent performed seven choice tasks.

2.2.2 Health States Valued

The EQ-VT experimental design includes a subset of health states to be evaluated by the respondent. These states were selected using a Monte Carlo simulation [17, 22].

For the cTTO, a subset of 86 health states were divided into ten blocks with ten health states per block [23, 24]. The blocks were balanced regarding the severity of the included states, with each block consisting of one mild state (i.e. four dimensions with severity level 1 and one dimension with severity level 2, e.g. 21111), the most severe state (i.e. state 55555) and eight intermediary states [14, 17, 24]. Each block was randomly assigned to respondents and states were presented in a random order.

The *unconscious state* is not part of the EQ-5D-5L descriptive system, and is not by default included in valuation [25]. However, the research team deemed it important for clinical research to obtain a utility value for this state,

which was also included in the Belgian EQ-5D-3L value set [12]. The unconscious state was added as an 11th state to two cTTO blocks as the last state to be valued to avoid interference with the standard cTTO procedure.

For the DCE, 196 choice pairs of EQ-5D-5L health states were divided over 28 blocks of seven pairs [23, 24]. The health states in a choice pair were comparable in severity and no choice pairs had to be evaluated where one alternative logically dominated the other. The assignment of a block to a respondent was randomized as was the order of paired choices and the left–right positioning of the states.

2.3 Data Collection and Quality Control

Interviewers were selected based on their experience in handling complex interviews. Prior to deployment, they received one-day training and performed five pilot interviews. A detailed account of the deployed interviewers and fieldwork is provided in ESM2.

Throughout the fieldwork, the quality of the collected data, the interviewer’s overall performance and the compliance with the interview script were monitored using a quality control (QC) tool integrated in the EQ-VT software [26]. At regular times, the EuroQol support team drafted QC reports and provided feedback. Interviews that did not meet specific criteria were flagged as “potentially poor quality” and were removed

according to pre-defined rules [15]. Interviewers with a high rate of flagged interviews were retrained or excluded.

2.4 Modelling and Data Analysis

The representativeness of the sample after data cleaning was assessed descriptively. For value set construction, a range of potential regression models were estimated. Multiple criteria were assessed to select a preferred model. The resulting EQ-5D-5L value set was then compared to the current EQ-5D-3L value set.

2.4.1 Final Sample

In addition to interviews excluded in the QC process, we excluded respondents who could not be linked to the sample plan due to an inaccurate registration ID. We also excluded respondents with implausible cTTO response patterns, based on three criteria drawn from previous valuation studies [19, 27–32]:

- (1) cTTO values were positively associated with the level sum score (misery index) of the health states in a linear regression.
- (2) All health states were given the same cTTO value.
- (3) The mild and most severe health state were given the same cTTO value.

In consultation with EuroQol, health states flagged in the feedback module were not excluded by default, this in deviation from the EQ-VT protocol and recent valuation studies [21, 33–35]. The main concern was to not further reduce the sample size (see “Results” section), which was smaller than targeted given the discontinuation of the data collection due to the coronavirus disease 2019 (COVID-19) crisis (see ESM2). The robustness of the results regarding this choice was tested (see Sect. 2.4.4). No restrictions were imposed based on the DCE data.

Summary statistics of the final sample were calculated to assess representativeness to the Belgian population. Belgian population characteristics were obtained from Statistics Belgium, the Belgian Health Interview Survey (HIS) and the EU Statistics on Income and Living Conditions (EU-SILC).

2.4.2 Modelling Approach

As only a subset of health states was valued directly by DCE or cTTO, a modelling approach was pursued to produce a value set for all EQ-5D-5L health states. The EQ-VT protocol does not prescribe a specific model.

The modelling approach consisted in the estimation and evaluation of a large number of potential regression models. Models were fitted on cTTO data only, DCE data only and

on the combined cTTO and DCE data (hybrid model) to maximize the available information [36].

Based on the literature, four main regression specifications were used, each with the same dependent and independent variables, but a different set of estimated coefficients [37–40]. Twenty independent variables were defined, i.e. four binary variables in each dimension indicating the severity level of the state, with level 1 being the reference. The dependent variable was different for the cTTO data, i.e. the reported disutility or deviation in utility from the full health state, and the DCE data, i.e. the choice made between two states. Disutility was preferred over utility because it is always positive (range 0–2), and thus easier to model.

In addition to the frequently used 20-coefficients additive model (ADD20)—with one coefficient for each binary variable—we also applied three multiplicative specifications with 8, 9 or 11 coefficients, allowing for different interactions between levels and dimensions as described in Rand-Hendriksen et al. [37], and referred to as MULT8, MULT9 and MULT11. While the additive model is more flexible and more widely used in valuation studies [21–23, 27, 28, 30–35, 41–47], research showed that the more restrictive, non-linear multiplicative models might perform better in terms of predictive accuracy [37, 38, 48]. Formal specifications are presented in ESM3.

For cTTO-only and hybrid models, different potential alternatives were tested within these four core specifications, again based on the literature: with/without *intercept*; with/without *random effects* (a respondent-specific component in the error term, which can be interpreted as individual variation around the intercept); with/without *heteroscedasticity* (a correction for increasing variability in reported cTTO values as health states worsen) following Ramos-Goñi et al. [36]; with/without *censoring* (correction for respondents who would like to value health states below the minimum cTTO value capped at –1 by design).

All models were fitted using maximum likelihood with SAS 9.4 procedure NLMIXED [49].

2.4.3 Model Selection

The final model should be able to fit the underlying data and predict values based on a subset of observed states. Model selection was based on four criteria (further elaborated in ESM4):

- (1) *Logical consistency*: Disutility in a health dimension does not decrease as severity increases.
- (2) *Goodness of fit (cTTO-only models)*: Ability of a model to fit the observed data.
- (3) *Predictive accuracy (cTTO-only models)*: Ability of a model to predict unobserved values.
- (4) *Theoretical considerations*: Evaluation of features (hybrid model, heteroscedasticity, censoring) that may

be desirable from a theoretical point of view, but may reduce the model's performance in terms of predictive accuracy or goodness of fit.

Criteria 1–3 were used to narrow down the selection regarding the core specification (ADD20, MULT8, MULT9, MULT11), the use of an intercept and random effects. Theoretical considerations were substantiated with suggestive evidence where possible.

2.4.4 Sensitivity Analysis

The robustness of the results was tested by re-estimating the model using three different data samples: (1) sample without exclusions based on implausible response patterns (see Sect. 2.4.1); (2) sample excluding interviews collected after the onset of the COVID-19 pandemic (about 5% of the respondents); and (3) sample excluding observations flagged in the feedback module (about 4% of the cTTO observations).

2.4.5 Comparing Value Sets

The main features and Kernel density distribution of the Belgian EQ-5D-5L value set was compared with the current Belgian EQ-5D-3L and related crosswalk value sets [12, 13].

3 Results

3.1 Sample Characteristics

Between May 2018 and September 2020, 22 interviewers performed 916 interviews that passed the EuroQoL QC process, while six interviewers performed poorly and had their interviews dropped ($n = 96$) (see ESM2). In addition, 24 further interviews were excluded due to inaccurate IDs ($n = 3$) and implausible responses ($n = 21$) (see Section 2.4.1), leaving a final sample of 892 respondents.

Sample characteristics are presented in Table 1. The sample was representative of the Belgian adult population in terms of age, sex, region of residence, educational attainment, labour market status, self-assessed health status and HRQoL. Three quarters of the respondents rated their health as good or very good and about 35% reported no health problem in any EQ-5D dimension, a slightly lower fraction than in the Belgian HIS. Problems related to PD were most frequently reported.

3.2 Data Characteristics

The full range of potential cTTO values was observed, although there was clustering at key values (see ESM5 for details on the distribution of cTTO values). No or a very

limited willingness to trade off life years to avoid health problems was observed in 9.5% (value 1) and 9.9% (value 0.95) of the responses. All available lead time was exhausted in 9.1% of the responses (value -1). Overall, 25.2% of cTTO values were negative. There was inter-interviewer variation regarding the fraction of responses clustered at -1 (range 0–32%) and the fraction of negatively valued states (range 4–50%) (see ESM5).

Figure 2 indicates that as health states worsen, cTTO values decrease and the variability is higher (difference 5th percentile [P5] and 95th percentile [P95] increases up to misery index 13, difference 25th percentile [P25] and 75th percentile [P75] increases up to misery index 18). The former supports the face validity of the data, while the latter is indicative of heteroscedasticity.

In support of the face validity of the DCE data, a correlation was found between the chosen state and the difference in misery index between both states (see Fig. 3). The health state with the lower misery index was favoured with the preference being more pronounced as the difference widened.

3.3 Preferred Model (Belgian EQ-5D-5L Model)

Taking all selection criteria into consideration, the hybrid version of the multiplicative eight-coefficient model with intercept for the cTTO data, with random effects and correction for heteroskedasticity, was selected as the preferred model (the full specification and details of the selection process are given in ESM6).

In summary, the additive cTTO-only model had the best fit, but was excluded because of logical inconsistencies in the coefficients of levels 2/3 in dimension UA. After enforcing logical consistency in dimension UA, inconsistencies showed up in levels 2/3 and 4/5 in dimension SC and levels 1/2 in dimension PD. No inconsistency was found in cTTO-only models for dimension MO, which was reassuring given the incorrect Dutch level 2 label. Second, the multiplicative models,—particularly MULT8 and MULT9, with intercept, random effects and correction for heteroscedasticity performed best on predictive accuracy, while also having a good fit. Third, as the data showed heteroskedasticity, a correction was applied. Fourth, following a high level of agreement between results from DCE-only and cTTO-only models, the combination of both data types in a hybrid model was justified to improve accurateness. In addition, data were not treated as censored given that predictions of the DCE-only models—by design uncensored—had a better fit with results from uncensored cTTO-only models [28]. Finally, as the additional coefficient in the MULT9 model was statistically insignificant, the MULT8 model was preferred.

Table 2 presents the coefficients and bootstrapped standard errors of the preferred model. The scatterplot (Fig. 4) between predicted and observed (cTTO) values shows

Table 1 General characteristics of respondents and the Belgian general adult population

Characteristics	Final sample (<i>N</i> = 892)		Belgian general adult population %	Difference sample – population Percentage points
	<i>N</i>	% (weighted)		
Demographics (sex, age group)*				
Male, aged 18–29	84	9.3	9.3	0.0
Male, aged 30–39	70	8.2	8.1	0.0
Male, aged 40–49	70	8.6	8.5	0.0
Male, aged 50–59	80	8.9	8.8	0.0
Male, aged 60–69	63	7.0	7.0	0.0
Male, aged 70–79	42	4.4	4.3	0.1
Male, aged 80–89	18	2.1	2.2	–0.1
Male, aged 90–99	1	0.1	0.3	–0.3
Female, aged 18–29	84	9.2	9.2	0.0
Female, aged 30–39	75	8.2	8.1	0.0
Female, aged 40–49	72	8.4	8.3	0.0
Female, aged 50–59	84	8.8	8.8	0.0
Female, aged 60–69	70	7.7	7.3	0.3
Female, aged 70–79	43	5.2	5.2	0.0
Female, aged 80–89	31	3.6	3.6	0.1
Female, aged 90–99	5	0.4	0.8	–0.4
Region of residence*				
Flanders	529	58.2	58.2	0.1
Wallonia	293	31.6	31.7	–0.1
Brussels	70	10.2	10.2	0.0
Educational attainment (highest degree)[§]				
No degree, primary or lower secondary	212	24.0	25.8	–1.8
Secondary or post-secondary	303	34.0	36.5	–2.5
Tertiary	370	42.0	37.7	4.3
Labour marker status[§]				
Employee	423	47.5	45.2	2.3
Self-employed	53	6.2	6.0	0.2
Unemployed	36	4.2	4.0	0.2
Inactive (including students)	136	15.5	19.0	–3.5
Retired	237	26.6	25.3	1.3
Self-assessed health[§]				
Very good	228	26.4	27.7	–1.3
Good	438	49.0	46.3	2.7
Fair	157	17.5	16.9	0.6
Bad	54	6.1	7.5	–1.4
Very bad	8	1.0	1.6	–0.6
Health status EQ-5D-5L[#]				
No problem reported (state 11111)	307	34.6	40.7	–6.1
Problem with mobility	199	22.7	17.4	5.3
Problem with self-care	59	6.2	5.7	0.5
Problem with usual activities	195	21.6	18.0	3.6
Problem with anxiety/depression	264	28.9	22.9	6.0
Problem with pain/discomfort	517	57.9	53.0	4.9
	<i>N</i>	Mean score	Mean score	Difference in score
Visual Analog Scale (EQ-VAS) by age[^]				
Aged 15–24	93	85.9	82.9	3.0

Table 1 (continued)

	<i>N</i>	Mean score	Mean score	Difference in score
Aged 25–34	151	82.9	80.7	2.2
Aged 35–44	126	80.5	78.3	2.2
Aged 45–54	147	75.6	77.3	–1.7
Aged 55–64	162	75.2	76.1	–0.9
Aged 65–74	119	75.2	73.5	1.7
Aged 75 or more	94	67.7	67.3	0.4

Six respondents had missing values regarding country-specific background questions (educational attainment, labour market status, self-assessed health)

EU-SILC EU Statistics on Income and Living Conditions, *HIS* Health Interview Survey

*Population characteristics from population data on 1 January 2017 provided by Statistics Belgium

§Population characteristics from EU-SILC data wave 2019

#Population characteristics from Belgian HIS wave 2018

^Population characteristics from Belgian HIS wave 2013

a good fit for the 86 evaluated health states. The utility decrements for the different dimensions at severity level 5 amount to: 0.227 (MO), 0.166 (SC), 0.181 (UA), 0.482 (PD) and 0.439 (AD). This needs to be multiplied with a rescaling factor of 0, 0.139, 0.258, 0.788 or 1 for severity levels 1–5, respectively. The overall utility loss for deviating from full health (intercept) equals 0.038. Loss of HRQoL is highest in the dimension PD, closely followed by AD.

The results are robust to the choices made with respect to the sample. There is an average absolute difference of 0.004, 0.002 and 0.005 between the coefficients of the preferred model and those from, respectively, the full sample, the sample limited to pre-COVID interviews and the sample excluding flagged interviews. The largest deviation was observed in dimension PD. In all robustness checks, the preference ranking of the dimensions was identical.

To ensure that 1 year in full health (state 11111) is valued as 1 QALY in later applications of the value set, its utility value was standardized at 1. To generate utility values for other states in the value set, utility decrements are calculated for each combination of dimension and severity level, summed together with the intercept and subtracted from 1. For example, the utility value for health state 15342 is calculated as follows:

$$\begin{aligned} \text{Utility } 15342 &= 1 - (\text{Intercept} + \beta_{\text{MO}} \times 0 + \beta_{\text{SC}} \times 1 \\ &\quad + \beta_{\text{UA}} \times L_3 + \beta_{\text{PD}} \times L_4 + \beta_{\text{AD}} \times L_2) \\ &= 1 - (0.038 + 0 + 0.166 + 0.047 \\ &\quad + 0.380 + 0.061) = 1 - 0.691 = 0.309 \end{aligned}$$

The utility for the unconscious state was obtained directly from the survey and amounts to –0.454. The full value set can be downloaded in CSV format (see ESM7).

3.4 Comparison Value Sets

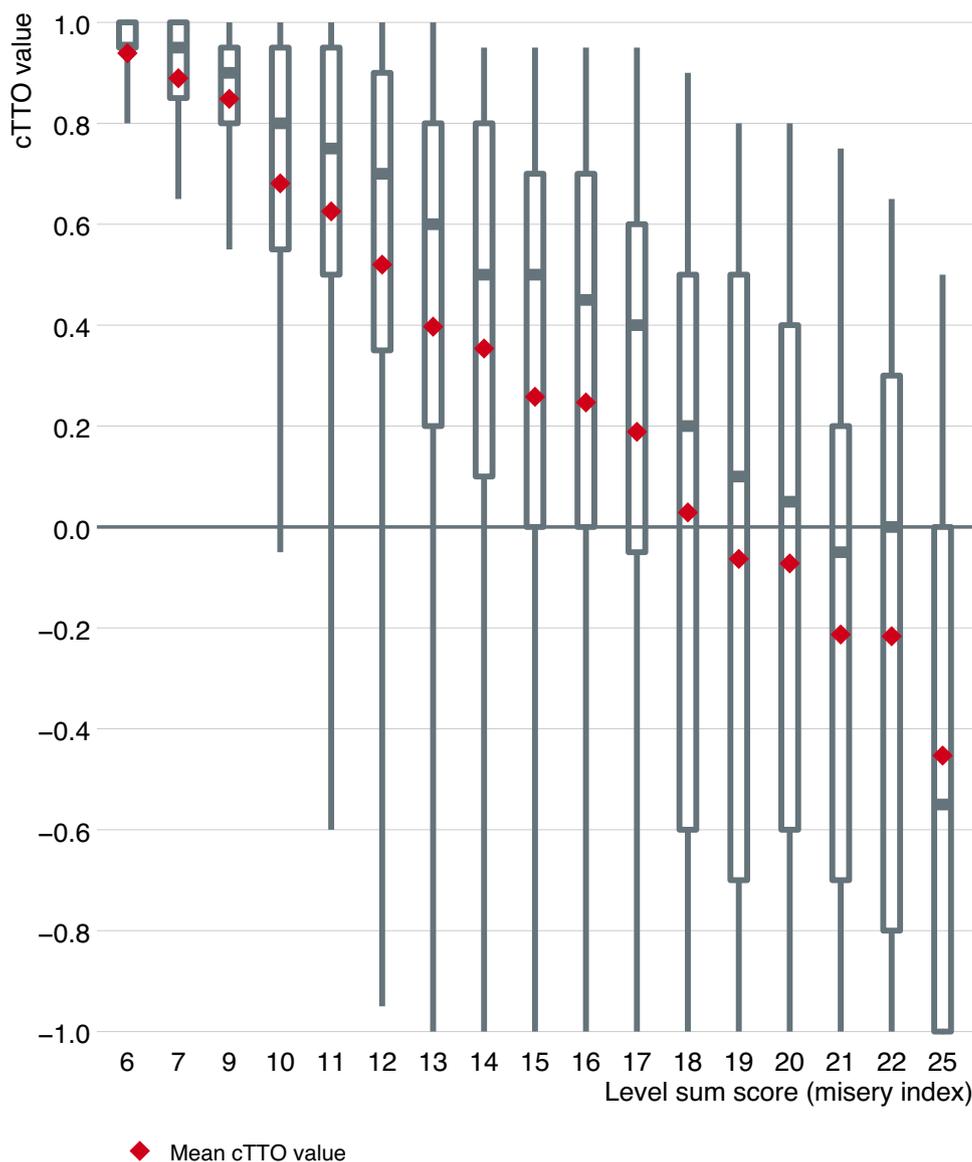
Kernel density plots (Fig. 5) and a summary of key features (Table 3) illustrate the discrepancy between the new Belgian EQ-5D-5L, the EQ-5D-3L and the crosswalk value sets. The EQ-5D-3L value set is characterized by two peaks (bimodal distribution). In the crosswalk set, roughly the same utility range is covered, but with a smoother transition between the peaks. In the EQ-5D-5L value set, the peak value is relatively similar, but the distribution of utility values has shifted to the right, i.e. higher valuation of mild states, and to the left, i.e. lower valuation of severe states and a higher share of negative values. The ranking of dimensions differs, although AD and PD still have the highest impact on HRQoL.

4 Discussion

This paper reports on the development of the Belgian EQ-5D-5L value set based on preferences from the Belgian general public and produced following the most recent EQ-VT protocol and QC process to ensure high data quality. The preferred underlying model was the hybrid multiplicative eight-coefficient model with intercept for the cTTO data, random effects and correction for heteroskedasticity.

Given the high level of agreement between results from DCE-only and cTTO-only models, the choice for a hybrid model is justifiable as one can argue that the same value function underlies the responses in both valuations. A hybrid model maximizes the available data, increasing precision. It was preferred in ten out of 15 valuation studies using EQ-VT protocol 2.0 or 2.1 [21, 22, 32–35, 41, 43, 44, 46–48, 51–53].

Fig. 2 Distribution of cTTO value by level sum score. The whiskers of the boxplot indicate the 5th percentile (P5) and 95th percentile (P95). *cTTO* composite time trade-off



A particular strength of this study was the representativeness of the sample. Although the EQ-VT protocol does not require representative sampling, this was deemed essential for the validity of the value set and its usability in health policy in the Belgian context where different language communities coexist (Dutch, French, German) and federal and regional governments share health competences. Respondents were randomly selected from the National Register using a multistage, stratified, cluster sampling with unequal probability design. Using comparative information from administrative sources and reference surveys, we concluded that the final sample was representative of the Belgian population in terms of age, sex, geographic spread, health status, HRQoL, education and employment status. Although representativeness in these fields does not necessarily imply representativeness in terms of health preferences, it at least

reduces the possible impact of lack of representativeness on these factors on the utility values obtained. The stringent sample selection procedure in combination with the QC process led to an extended data collection period. This was, however, not considered problematic as it makes the collected utility values less prone to time specific effects [35].

Great effort was also put into the model selection process. Given the nature of the value set where utility values are predicted for 3125 health states based on a limited number of observed states, the criterion of predictive accuracy was considered highly relevant. It was a main argument to prefer the multiplicative eight-parameter model.

The new value set is based on a robust and valid method and provides up-to-date utilities, proper to the EQ-5D-5L instrument. There are several major methodological improvements compared to the EQ-5D-3L and crosswalk

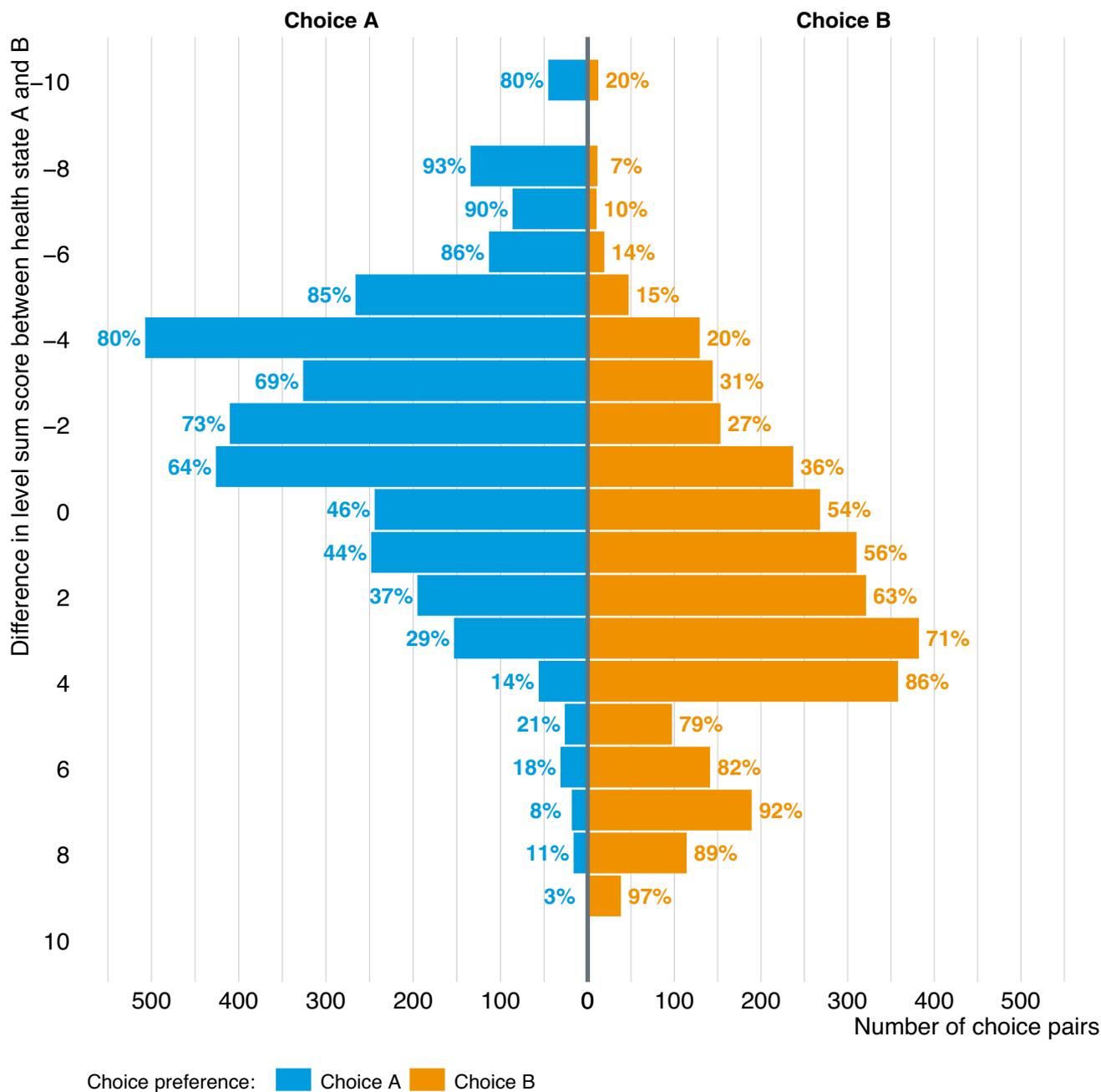


Fig. 3 Preference for alternative A or B in DCE task by difference in level sum score (misery index). DCE discrete choice experiment

value sets. [12]. First, the five-level version improved the discriminatory power and sensitivity of the earlier three-level version [7]. Second, a clear protocol and QC process improved the quality, consistency and accurateness of the valuation data [14, 15]. Third, in comparison with the VAS used for the EQ-5D-3L, cTTO and DCE are more robust valuation techniques for utility measurement, embedded in economic random utility theory. Fourth, face-to-face interviews in a one-to-one setting were conducted instead of postal surveys, which is more appropriate given the cognitive

challenging nature of the cTTO task [15, 20]. Fifth, for the EQ-5D-3L value set, the sample was geographically limited to one Belgian region, while preferences for the new value set were surveyed from a sample of the entire adult Belgian population. Such extension was desirable since the value set is used for national decision making.

For these reasons, we recommend the use of the EQ-5D-5L instrument in combination with the newly developed value set. Inevitably, this will generate different results than the previous value set. First, the EQ-5D-5L

Table 2 Coefficients and bootstrapped standard errors of the preferred model and results of the sensitivity analysis

	Preferred model				Sensitivity analysis		
	Coefficient value	Standard error	T statistic	<i>P</i> value	Sample, no exclusions on cTTO values	Restricted sample pre-COVID	Restricted sample feedback module
Intercept	0.038	0.0148	2.55	0.005	0.037	0.041	0.033
β_{MO}	0.227	0.0102	22.19	< 0.0001	0.220	0.228	0.230
β_{SC}	0.166	0.0108	15.32	< 0.0001	0.165	0.166	0.170
β_{UA}	0.181	0.0098	18.37	< 0.0001	0.181	0.180	0.184
β_{PD}	0.482	0.0138	34.84	< 0.0001	0.466	0.481	0.492
β_{AD}	0.439	0.0132	33.13	< 0.0001	0.434	0.446	0.447
L_2	0.139	0.0154	9.02	< 0.0001	0.144	0.136	0.139
L_3	0.258	0.0158	16.34	< 0.0001	0.262	0.258	0.254
L_4	0.788	0.0157	50.26	< 0.0001	0.787	0.789	0.781

Standard errors of the coefficients were derived using the Rao-Wu bootstrap with 1000 replications [50]

AD anxiety/depression, *COVID* coronavirus disease, *cTTO* composite time trade-off, *MO* mobility, *PD* pain/discomfort, *SC* self-care, *UA* usual activities, β_x coefficient for dimension x , L_x coefficient for severity level x

value set has a higher fraction of health states valued worse than dead (15.0% in 5L vs 6.6% in 3L) and covers a larger range of utility values than the EQ-5D-3L set (−0.532 to 0.939 in 5L and −0.158 to 0.817 in 3L, excluding state 11111). The higher valuation of mild health states reflects the very limited willingness to trade off life years to avoid mild health problems that was observed among survey respondents. A wider range of utility values allows for a better discrimination between patients, both for mild and (very) severe health conditions, which improves the applicability of the EQ-5D instrument. Also, a low HRQoL value (−0.454, only five EQ-5D-5L states in the new value set are valued lower) was derived directly from survey responses for the state of unconsciousness (coma). This state cannot be described by the EQ-5D system, but might nevertheless be important in some assessments (e.g. of interventions where patients are or go into coma). Second, the preference ranking of the dimensions has altered. In the 3L value set, the AD dimension is associated with the highest utility loss, followed by PD, SC, MO and UA. In the 5L value set, the dimension PD is associated with the highest loss in HRQoL, closely followed by AD and then MO, UA and SC. Pain and discomfort is also the dimension for which most respondents reported problems regarding their own health. It is expected that policy priorities and resource allocation decisions will reflect the importance attached to pain relief and mental health.

One limitation of the study was that our sample did not reach the targeted 1000 respondents. The COVID-19 pandemic was one reason, but in general, it proved difficult to find sufficient individuals willing to participate, with only a short list of randomly selected, geographically dispersed potential candidates at our disposal. Sampling in fewer

locations or from a market research panel would have been faster and less resource intensive, but would come at a cost of reduced representativeness. However, given that the target of 1000 respondents was set to have sufficient power to estimate a 20-parameter additive model, it can be argued that our smaller sample still has sufficient power to accurately estimate a more parsimonious model. A second limitation was that more interviewers worked in the field than put forward in the protocol, this because of replacements and drop-outs. Although there is evidence to support the face validity of the collected data, it seems that part of the variability in the data, and in particular in health states valued worse than dead in the cTTO task, can be attributed to differences between interviewers. However, the random effects included in the model allow for individual variation in the valuation and can (at least partly) absorb such interviewer-specific effects.

A final concern was an incorrect labelling of severity level 2 in the MO dimension in the Dutch version of the EQ-VT software. The impact of this was expected to be limited because interviewers were provided a correctly labelled printed overview to help respondents locate the health state described on the screen. Moreover, the data analysis did not reveal any unexpected pattern in valuation scores in the MO dimension.

5 Conclusion

This study established an EQ-5D-5L value set, based on the preferences of a random sample of the Belgian general public in 2018–2020. The EQ-5D-5L value set should replace the currently used EQ-5D-3L value set from 2003. Instead

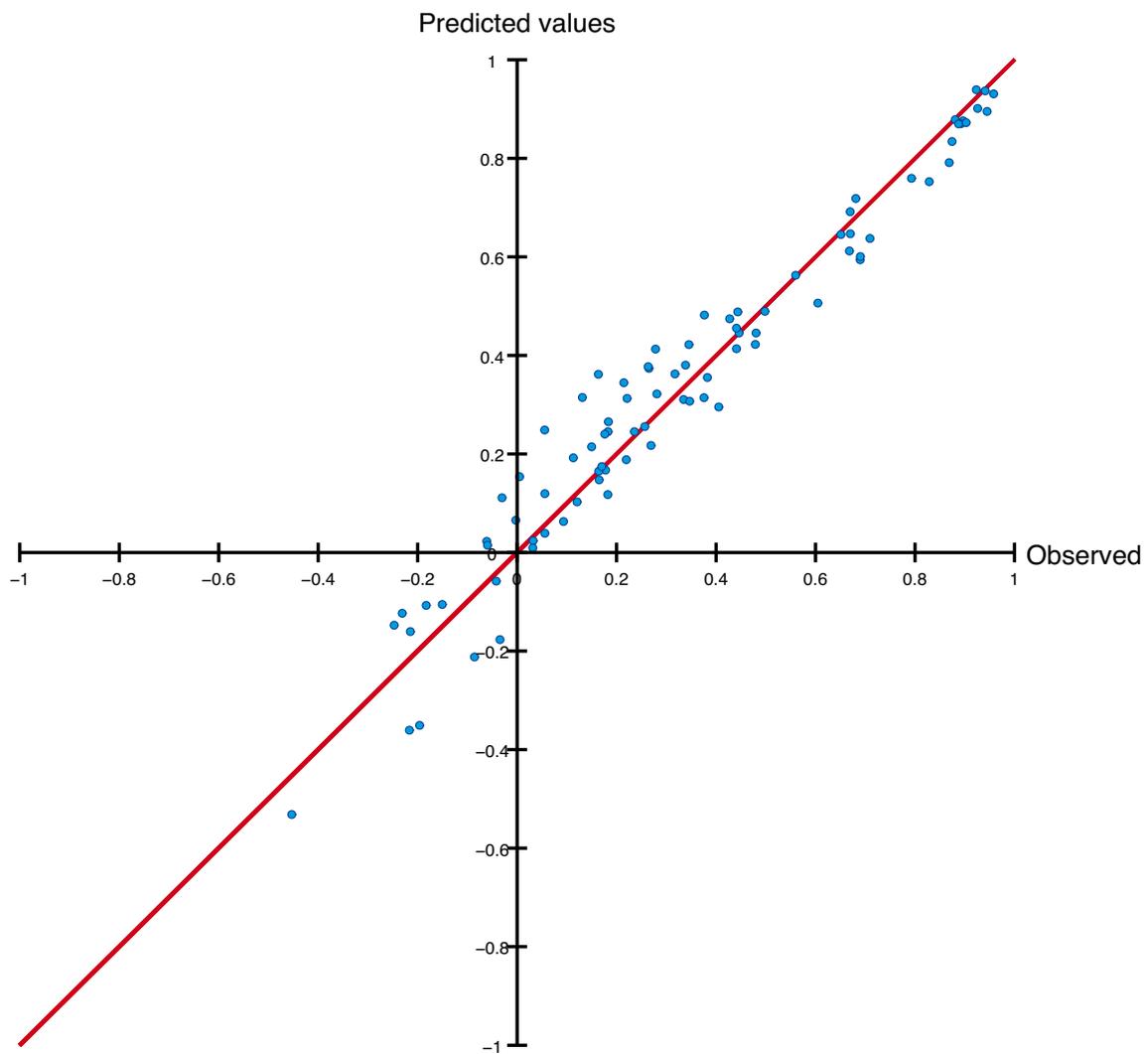


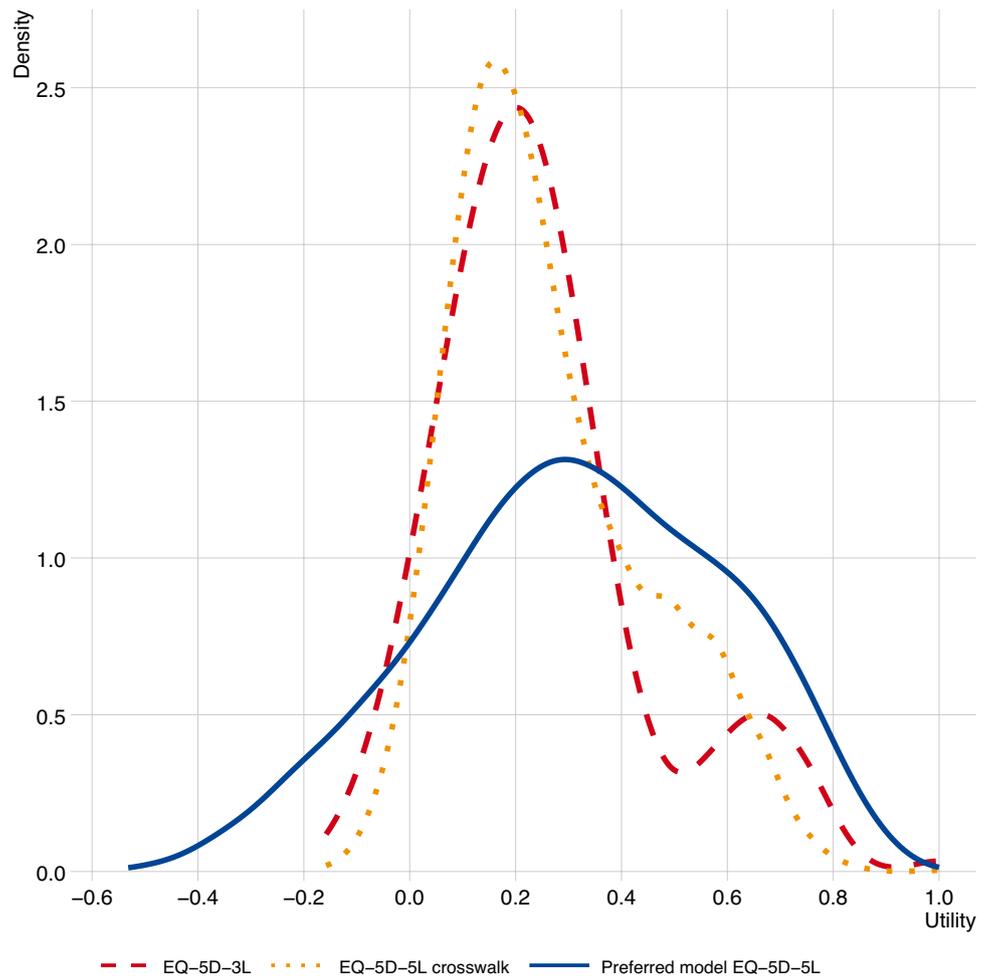
Fig. 4 Scatterplot of predicted (preferred model) and observed (cTTO) utility values for the 86 evaluated states. The diagonal is presented by the red line and indicates an equal predicted and observed utility values. cTTO composite time trade-off

Table 3 Comparison of key characteristics of the Belgian EQ-5D-5L and EQ-5D-3L value sets

	Preferred model	EQ-5D-3L value set
% health states valued worse than dead	15.0%	6.6%
Preference ranking of dimensions (ordered from highest to lowest utility loss at level 5)	PD AD MO UA SC	AD PD SC MO UA
Minimum value	-0.532 (state 55555)	-0.158 (state 33333)
Maximum value (except full health)	0.939 (state 12111)	0.817 (state 11211)

AD anxiety/depression, MO mobility, PD pain/discomfort, SC self-care, UA usual activities

Fig. 5 Kernel density plots of three Belgian EQ-5D value sets: the new EQ-5D-5L, and the current EQ-5D-3L and crosswalk value sets



of describing and valuing 243 health states, the current value set values 3125 health states. The value set provides opportunities for future clinical and economic evaluations in healthcare, for the measurement of patient-reported outcomes and for population health assessments.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s41669-022-00353-3>.

Acknowledgements The authors are grateful for the work of Profacts, the interviewers and the EuroQol Research Foundation for support provided throughout the conduct of the study.

Declarations

Funding This research was co-funded by the Belgian Health Care Knowledge Centre (KCE) and the EuroQol Research Foundation (EQ Project 20170490 and 20170491). The KCE is a federal institution that is financed by the National Institute for Health and Disability Insurance (NIHDI, RIZIV – INAMI), the Federal Public Service of health, food chain safety and environment, and the Federal Public Service of social security. The development of health services research studies is part of the legal mission of the KCE. Although the development of the studies is paid by the KCE budget, the sole mission of the KCE is providing scientifically valid information.

Conflict of interest Nicolas Bouckaert, Irina Cleemput, Stephan Devriese and Sophie Gerkens declare that they have no conflicts of interest directly relevant to the content of this article.

Ethics approval The study received approval from the ethics committee of the university hospital of Ghent (B670201835610) and was conducted in accordance with the Declaration of Helsinki.

Consent to participate Informed consent was obtained from all individual participants included in the study. Participants were informed about their freedom of refusal. Anonymity and confidentiality were maintained throughout the research process.

Consent for publication The study participants consented to the collection, storage and analysis of their data and the publication of anonymous final results.

Data sharing Anonymous datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. Requests for personal data at the individual level will be judged case by case to assure compliance with informed consent clauses on confidentiality of personal data as approved by the ethics committee.

Code availability Code for data cleaning and analysis are available from the corresponding author on reasonable request.

Author contributions All authors contributed to the study conception and design. SD was responsible for the sampling design. SG and IC supervised the data collection and quality control process. NB was responsible for the data analysis. The first draft of the manuscript was written by NB and SG. All authors played a role in the review of the analysis, the interpretation of the results and the review and editing of the final version of the manuscript.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

- Thiry N, Neyt M, Van De Sande S, Cleemput I. Belgian guidelines for economic evaluations: second edition. *Int J Technol Assess Health Care*. 2014;30(6):601–7. <https://doi.org/10.1017/s0266462314000725>.
- Royal Decree laying down the procedures, periods and conditions regarding the reimbursement of the costs of pharmaceutical specialities by the compulsory health and disability insurance. Belgian official journal I February 2018.
- Kennedy-Martin M, Slaap B, Herdman M, van Reenen M, Kennedy-Martin T, Greiner W, et al. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. *Eur J Health Econ*. 2020;21(8):1245–57. <https://doi.org/10.1007/s10198-020-01195-8>.
- 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(10):1727–36. <https://doi.org/10.1007/s11136-011-9903-x>.
- 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(7):1717–27. <https://doi.org/10.1007/s11136-012-0322-4>.
- Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. *Med Care*. 2007;45(3):259–63. <https://doi.org/10.1097/01.mlr.0000254515.63841.81>.
- 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(6):675–97. <https://doi.org/10.1007/s40273-018-0623-8>.
- Feng Y, Devlin N, Herdman M. Assessing the health of the general population in England: how do the three- and five-level versions of EQ-5D compare? *Health Qual Life Outcomes*. 2015;13(1):171. <https://doi.org/10.1186/s12955-015-0356-8>.
- Whitehead SJ, Ali S. Health outcomes in economic evaluation: the QALY and utilities. *Br Med Bull*. 2010;96(1):5–21. <https://doi.org/10.1093/bmb/ldq033>.
- Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. 2nd ed. Oxford: Oxford University Press; 2017.
- Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goñi JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics*. 2016;34(10):993–1004. <https://doi.org/10.1007/s40273-016-0404-1>.
- Cleemput I. A social preference valuations set for EQ-5D health states in Flanders, Belgium. *Eur J Health Econ*. 2010;11(2):205–13. <https://doi.org/10.1007/s10198-009-0167-0>.
- van Hout B, Janssen MF, Feng Y-S, 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 in Health*. 2012;15(5):708–15. <https://doi.org/10.1016/j.jval.2012.02.008>.
- 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(4):445–53. <https://doi.org/10.1016/j.jval.2014.04.002>.
- Stolk E, Ludwig K, Rand K, van Hout B, Ramos-Goñi JM. Overview, update, and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. *Value Health*. 2019;22(1):23–30. <https://doi.org/10.1016/j.jval.2018.05.010>.
- Xie F, Pickard AS, Krabbe PFM, Revicki D, Viney R, Devlin N, et al. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). *Pharmacoeconomics*. 2015;33(8):867–77. <https://doi.org/10.1007/s40273-015-0292-9>.
- Oppe M, van Hout B. The “power” of eliciting EQ-5D-5L values: the experimental design of the EQ-VT. Rotterdam: EuroQol Group; 2017.
- Demarest S, Van der Heyden J, Charafeddine R, Drieskens S, Gisle L, Tafforeau J. Methodological basics and evolution of the Belgian health interview survey 1997–2008. *Arch Public Health*. 2013;71(1):24. <https://doi.org/10.1186/0778-7367-71-24>.
- Janssen BMF, 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(1):5–13. <https://doi.org/10.1007/s10198-013-0503-2>.
- Shah KK, Lloyd A, Oppe M, Devlin NJ. One-to-one versus group setting for conducting computer-assisted TTO studies: findings from pilot studies in England and the Netherlands. *Eur J Health Econ*. 2013;14(1):65–73. <https://doi.org/10.1007/s10198-013-0509-9>.
- Jensen CE, Sørensen SS, Gudex C, Jensen MB, Pedersen KM, Ehlers LH. The Danish EQ-5D-5L value set: a hybrid model using cTTO and DCE data. *Appl Health Econ Health Policy*. 2021;19(4):579–91. <https://doi.org/10.1007/s40258-021-00639-3>.
- Golicki D, Jakubczyk M, Graczyk K, Niewada M. Valuation of EQ-5D-5L health states in Poland: the first EQ-VT-based study in central and eastern Europe. *Pharmacoeconomics*. 2019;37(9):1165–76. <https://doi.org/10.1007/s40273-019-00811-7>.
- Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabases JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Med Care*. 2017;55(7):e51–8. <https://doi.org/10.1097/MLR.000000000000283>.
- van Hout B, Mulhern B, Feng Y, Shah K, Devlin N. The EQ-5D-5L value set for England: response to the “Quality Assurance.” *Value Health*. 2020;23(5):649–55. <https://doi.org/10.1016/j.jval.2019.10.013>.
- Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: past, present and future. *Appl Health Econ Health Policy*. 2017;15(2):127–37. <https://doi.org/10.1007/s40258-017-0310-5>.

26. Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJ, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value Health*. 2017;20(3):466–73. <https://doi.org/10.1016/j.jval.2016.10.012>.
27. 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(1):7–22. <https://doi.org/10.1002/hec.3564>.
28. 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(4):343–52. <https://doi.org/10.1016/j.jval.2016.01.003>.
29. Xie F, Pullenayegum E, Gaebel K, Bansback N, Bryan S, Ohinmaa A, et al. A time trade-off-derived value set of the EQ-5D-5L for Canada. *Med Care*. 2016;54(1):98–105. <https://doi.org/10.1097/MLR.0000000000000447>.
30. 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(2):323–33. <https://doi.org/10.1007/s11136-015-1086-4>.
31. Kim S-H, 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(7):1845–52. <https://doi.org/10.1007/s11136-015-1205-2>.
32. Ferreira PL, Antunes P, Ferreira LN, Pereira LN, Ramos-Goñi JM. A hybrid modelling approach for eliciting health state preferences: the Portuguese EQ-5D-5L value set. *Qual Life Res*. 2019;28(12):3163–75. <https://doi.org/10.1007/s11136-019-02226-5>.
33. Andrade LF, Ludwig K, Ramos-Goñi JM, Oppe M, de Pourville G. A French value set for the EQ-5D-5L. *Pharmacoeconomics*. 2020;38(4):413–25. <https://doi.org/10.1007/s40273-019-00876-4>.
34. Ludwig K, Graf von der Schulenburg JM, Greiner W. German value set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36(6):663–74. <https://doi.org/10.1007/s40273-018-0615-8>.
35. Hobbins A, Barry L, Kelleher D, Shah K, Devlin N, Ramos-Goñi JM, et al. Utility values for health states in Ireland: a value set for the EQ-5D-5L. *Pharmacoeconomics*. 2018;36(11):1345–53. <https://doi.org/10.1007/s40273-018-0690-x>.
36. Ramos-Goñi JM, Craig B, Oppe M, van Hout B. Combining continuous and dichotomous responses in a hybrid model. Rotterdam: EuroQol Research Foundation; 2016.
37. 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(7):945–52. <https://doi.org/10.1016/j.jval.2017.03.013>.
38. 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(4):662–9. <https://doi.org/10.1016/j.jval.2016.11.016>.
39. Feng Y, Devlin N, Shah K, Mulhern B, van Hout B. New methods for modelling EQ-5D-5L value sets: an application to English data. London: Office of Health Economics; 2016.
40. Oppe M, Ramos-Goñi JM, van Hout B. Modeling EQ-5D-5L valuation data. In: Busschbach JJ, editor. 29th Scientific Plenary Meeting of the EuroQol Group. Rotterdam; 2012. p. 61–91.
41. Purba FD, Hunfeld JAM, Iskandarsyah A, Fitriana TS, Sadarjoen SS, Ramos-Goñi JM, et al. The Indonesian EQ-5D-5L value set. *Pharmacoeconomics*. 2017. <https://doi.org/10.1007/s40273-017-0538-9>.
42. Shiroya 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(5):648–54. <https://doi.org/10.1016/j.jval.2016.03.1834>.
43. Lin H-W, Li C-I, Lin F-J, Chang J-Y, Gau C-S, Luo N, et al. Valuation of the EQ-5D-5L in Taiwan. *PLoS ONE*. 2018;13(12):e0209344. <https://doi.org/10.1371/journal.pone.0209344>.
44. Rencz F, Brodszky V, Gulácsi L, Golicki D, Ruzsa G, Pickard AS, et al. Parallel valuation of the EQ-5D-3L and EQ-5D-5L by time trade-off in Hungary. *Value Health*. 2020;23(9):1235–45. <https://doi.org/10.1016/j.jval.2020.03.019>.
45. Pattanaphesaj J, Thavorncharoensap M, Ramos-Goñi JM, Tongsir S, Ingrisawang L, Teerawattananon Y. The EQ-5D-5L valuation study in Thailand. *Expert Rev Pharmacoecon Outcomes Res*. 2018;18(5):551–8. <https://doi.org/10.1080/14737167.2018.1494574>.
46. Mai VQ, Sun S, Minh HV, Luo N, Giang KB, Lindholm L, et al. An EQ-5D-5L Value Set for Vietnam. *Qual Life Res*. 2020;29(7):1923–33. <https://doi.org/10.1007/s11136-020-02469-7>.
47. Al Shabasy S, Abbassi M, Finch A, Roudijk B, Baines D, Farid S. The EQ-5D-5L valuation study in Egypt. *Pharmacoeconomics*. 2021. <https://doi.org/10.1007/s40273-021-01100-y>.
48. Shafie AA, Vasan Thakumar A, Lim CJ, Luo N, Rand-Hendriksen K, Md Yusof FA. EQ-5D-5L valuation for the Malaysian population. *Pharmacoeconomics*. 2019;37(5):715–25. <https://doi.org/10.1007/s40273-018-0758-7>.
49. SAS Institute Inc. SAS/STAT®14.1 user's guide: the NLMIXED procedure. Cary: SAS Institute Inc.; 2015.
50. Heeringa SG, West BT, Berglund PA. Applied survey data analysis. 2nd ed. Boca Raton: Chapman and Hall; 2017. <https://doi.org/10.1201/9781315153278>.
51. Augustovski F, Belizán M, Gibbons L, Reyes N, Stolk E, Craig BM, et al. Peruvian Valuation of the EQ-5D-5L: a direct comparison of time trade-off and discrete choice experiments. *Value Health*. 2020;23(7):880–8. <https://doi.org/10.1016/j.jval.2020.05.004>.
52. Pickard AS, Law EH, Jiang R, Pullenayegum E, Shaw JW, Xie F, et al. United States valuation of EQ-5D-5L health states using an international protocol. *Value in Health*. 2019;22(8):931–41. <https://doi.org/10.1016/j.jval.2019.02.009>.
53. Gutierrez-Delgado C, Galindo-Suárez R-M, Cruz-Santiago C, Shah K, Papadimitropoulos M, Feng Y, et al. EQ-5D-5L health-state values for the Mexican population. *Appl Health Econ Health Policy*. 2021;19(6):905–14. <https://doi.org/10.1007/s40258-021-00658-0>.