ORIGINAL RESEARCH ARTICLE

The Better than Dead Method: Feasibility and Interpretation of a Valuation Study

R. A. van Hoorn · A. R. T. Donders ·
M. Oppe · P. F. M. Stalmeier

Abstract

Background Traditionally, the valuation of health states worse than being dead suffers from two problems: [1] the use of different elicitation methods for positive and negative values, necessitating arbitrary transformations to map negative to positive values; and [2] the inability to quantify that values are time dependent. The Better than Dead (BTD) method is a health-state valuation method where states with a certain duration are compared with being dead. It has the potential to overcome these problems.

Objectives To test the feasibility of the BTD method to estimate values for the EQ-5D system.

Methods A representative sample of 291 Dutch respondents (aged 18–45 years) was recruited. In a web-based questionnaire, preferences were elicited for a selection of 50 different health states with six durations between 1 and 40 years. Random-effects models were used to estimate the effects of socio-demographic and experimental variables, and to estimate values for the EQ-5D. Test–retest reliability was assessed in 41 respondents.

Results Important determinants for BTD were a religious life stance [odds ratio 4.09 (2.00–8.36)] and the educational level. The fastest respondents more often preferred health-state scenarios to being dead and had lower test–retest reliability (0.45 versus 0.77 and 0.84 for fast, medium and slow response times, respectively). The results showed a small number of so-called maximal endurable time states.

Conclusion Valuating health states using the BTD method is feasible and reliable. Further research should explore how the experimental setting modifies how values depend on time.

R. A. van Hoorn (✉) · A. R. T. Donders · P. F. M. Stalmeier
Department for Health Evidence, Radboud University Nijmegen
Medical Center, PO Box 9101, 6500 HB Nijmegen,
The Netherlands
e-mail: Ralph.vanHoorn@radboudumc.nl

A. R. T. Donders
e-mail: Rogier.Donders@radboudumc.nl

P. F. M. Stalmeier
e-mail: Peep.Stalmeier@radboudumc.nl

M. Oppe
Department of Health Policy and Management, Institute
for Medical Technology Assessment, Erasmus University
Rotterdam, PO Box 1738, 3000 DR Rotterdam,
The Netherlands
e-mail: oppe@euroqol.org

Key Points for Decision Makers

The Better than Dead (BTD) method led to consistent weights for health attributes and duration.

A small number of maximal endurable time states was detected.

Decision makers might consider time-dependent values for analyses.

1 Introduction

Health-related quality of life (HRQoL) is a commonly used effect measure in economic and health care evaluations [1]. By expressing HRQoL as a single value anchored on full

health (=1) and being dead (=0), comparison of health effects across different diseases and interventions is possible [2]. These health values are often combined with survival data to compute quality-adjusted life-years (QALYs). A variety of instruments exist to measure these values, such as the visual analogue scale (VAS), standard gamble (SG), time trade-off (TTO) and discrete-choice experiment (DCE) [3, 4].

In TTO and SG instruments, states better than being dead are valued on a scale ranging from being dead to full health. But for states worse than being dead, such a scale is not adequate, and another task is used. A challenge arises as different tasks are used to elicit values for states judged to be better than being dead (positive values) and states worse than being dead (negative values) [5–7] Transformations are used to place these values on a single value scale, but there is no agreed method for doing so [7–10]. A second challenge lies in the assumption that health-state values are independent of their durations. Studies have shown that in some cases, health states are valued differently as their durations increases. For instance, Sutherland et al. and Stalmeier et al. [11, 12] described a phenomenon where a severe positive health state is less likely to be preferred over being dead when the health state lasts longer than a certain duration, the 'maximal endurable time' (MET). Time dependency has implications for the way health-state values are used in Markov models, for example, where each state is assigned a utility value that does not depend on time.

Traditional valuation methods (VAS, TTO, SG) have not provided satisfactory solutions for the two challenges stated above [13]. Recent studies have partly overcome these challenges by using discrete choices between health states with different durations, enabling the estimation of negative values, using probabilistic choice models [14–17]. Variants of the TTO method, e.g. the lead-time TTO method, have been developed to avoid transformation problems [18]. These methods tackle the first challenge described above but have not been capable of picking up MET states.

A valid method to assess health states worse than being dead has to be able to detect time-dependent states, because for such states, while they are initially judged as better than being dead, worse than being dead preferences may occur later on. To this end, the Better than Dead (BTD) method was developed, in which preferences are elicited between health states with specified durations (scenarios) and being dead [12]. Comparisons with being dead have been part of the elicitation procedure of traditional methods. However, the insight that comparisons with being dead can be the sole basis for deriving health-state values is new. The BTD method enables a straightforward valuation of states better and worse than being dead in a single question, and allows quantitative measurement of time-dependent values [12].

Our objective was to show that the BTD method can be used to estimate values for a multi-attribute health utility index in a web-based study. First, the social demographic and experimental determinants of BTD preferences were investigated. Second, the reliability and consistency of the BTD method were tested. Last, values were estimated for the EQ-5D system.

2 Methods

An agency for market research (Survey Sampling Europe BV) recruited respondents between 18 and 45 years of age from the Dutch population. The upper limit of 45 years avoided presenting unrealistic health states (e.g. confronting a 60-year-old respondent with a health state lasting 40 years). The respondents were offered tokens worth €6.00 for participation.

Health states in the questionnaire were expressed in the EQ-5D three-level system [19]. This system uses five attributes (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) to describe health states. Each attribute has three levels [no problems (1), some problems (2) and extreme problems (3)]. Thus the worst health state the EQ-5D can describe has levels of 33333 and the best health state has levels of 11111.

The web-based questionnaire consisted of three parts. Part 1 consisted of questions regarding socio-demographic characteristics, and the respondents gave a self-description of their health status, using the EQ-5D. Part 2 started with a 2-minute video introducing how the EQ-5D system is used to describe health states and discussed an example choice between states 11111 and 11112. This was followed by a warm-up question consisting of a comparison between states 11113 and 31111 without a specified duration. An audio fragment pointed out the differences between these states. In the second warm-up question, health state 33312 was compared with death, described as being quick and painless. The notion of comparing states with death was introduced by audio fragments. It was explained that being dead should not be associated with a violent ending (some pilot study respondents held such interpretations). A final audio fragment introduced the notion of duration, and the notion that some may dislike a bad state lasting too long. The third warm-up question was between (11123, 10 years) and being dead. The respondents were asked to make a forced choice between two scenarios: 'living for a number of years in a certain health state followed by dead', or 'dead'. Part 3 consisted of 108 of these so-called BTD questions (see Fig. 1). The respondents were not limited in the time allowed for them to answer the questions, but they were unable to skip questions.

Fig. 1 Better than Dead (BTD) preference question: in the web-based questionnaire, respondents were asked to indicate whether they preferred state A or state B by clicking on the preferred option

Imagine you still have 1 year to live in state A below, after which you decease. Below B is a state which equals dead. Please indicate whether you would prefer to live for 1 year in state A or prefer state B.

Please choose the state you prefer. For 1 of the 2 states a time span is stated, which indicates how long you have to live <u>in the described health</u> state before you decease. Choose now A or B.

A

Some problems in walking about
Some problems washing or dressing myself
Unable to perform my usual activities
Extreme pain or discomfort
Extremely anxious or depressed

1 year

B

DEAD

A total of 50 health states, including 17 health states from Macran and Kind [20], were selected for the experiment, using Bayesian optimal design techniques. These were divided into three blocks of 108 scenarios (each block contained 18 health states, each with six different durations of 1, 2, 5, 10, 20 and 40 years). Level balance in each block was ensured. Respondents were randomly assigned to these blocks. The side on which 'dead' was presented (left or right) was balanced within each block. The presentation order of the scenarios was randomized for each respondent separately. Response times were recorded for each BTD question separately. By including 70 respondents per block, the maximum standard error for any proportion would be 0.06.

Respondents from the first block were asked to participate voluntarily in a second session to assess test–retest properties. Those who agreed were given 30 BTD questions combining ten states (11211, 11312, 12111, 23232, 32211, 13213, 33113, 13232, 22333, 21233) with three durations (2, 10 and 40 years). These states spanned the value range according to a previous Dutch valuation study [21]. Again, presentation order was randomized.

3 Analysis

The analysis covered the determinants, test–retest characteristics, consistency of BTD preferences and predicted values. Analyses were performed in R version 3.0.1 with lme4 and psych packages [22].

3.1 Determinants of BTD Preferences

The effect of self-description—that is, how respondents described their health on the EQ-5D attributes—and socio-demographic variables on BTD preferences were estimated using a binomial logistic random-effect (RE) model, including the 50 presented health states and the six durations as fixed effects. It included a random effect for

respondents to take into account correlated valuations as each respondent rated multiple scenarios. The effects of the presentation order of the questions, presentation side of 'dead' and response time were estimated in a similar but separate model. For each respondent, the median of the response times to the 108 BTD questions was used as measure of response speed.

3.2 Test–Retest Properties

The test–retest properties of the BTD method were evaluated using tetrachoric correlations between the test and retest answers to matched BTD questions. Tetrachoric correlations are similar to Pearson correlations and correlate dichotomous choices by assuming these choices are based on normally distributed intrinsic values [23]. The tetrachoric correlations were calculated as measures for overall agreement aggregated on the 30 BTD questions asked in the test and retest across all respondents, and additionally in strata (tertiles) based on the median response time per respondent.

3.3 Comparison with Previous Work

The results were compared with previous work [12]. To assess the occurrence of MET states, MET states were defined by (5 years X) preferred to being dead but (20 years X) not preferred to being dead, as in previous work [12]. This time dependency of preferences is considered to be acceptable. The percentage of respondents with MET preferences was determined. To assess the consistency of the BTD method, two tests were performed: (1) if a state lasting 20 years was preferred to being dead, shorter (5- or 10-year) durations of that state should also be preferred to being dead; and (2) if being dead was preferred to a state lasting 5 years, being dead should also be preferred to longer (10- or 20-year) durations. These two tests assumed that health-state values were independent of duration, and violations were not acceptable. These

analyses were limited to health states 11121, 11312, 13311, 11113 and 23232, as these were identical to the states used in previous work [12].

3.4 Modelling Health-State Values

The most commonly used version of the QALY model defines the QALY U for health state Q with duration t as $U(Q,t) = V(Q) \cdot t$ [24]. The existence of time-dependent MET states violates this model, as the effect of t is different for different values of Q [24]. Thus, in order to model MET states, a generalization is needed:

$$U(Q,t) = V(Q(t)) \cdot t \tag{1}$$

where $V(Q(t))$ allows for time dependence for the value of Q. For mild states, BTD preferences do not depend on time. Therefore, BTD preferences are assumed not to measure $U(Q,t)$ but to measure $V(Q(t))$ [12]. $V(Q(t))$ is estimated by taking the latent value of the binary logistic random-effects model estimated on the data:

$$\text{pr(BTD)} = \frac{1}{1 + e^{-V(Q(t))}} \tag{2}$$

where pr(BTD) is the percentage of respondents preferring (Q,t) to being dead.

Preliminary analysis of the data, stratified by the six durations, showed that a model consisting of 11 dummy variables (two for each of the five EQ-5D attributes, with level 1 (no problems) as the base value, and an 'N3' term equalling 1 if any of the EQ-5D attributes are at level 3, otherwise it is 0) fitted the data best, according to likelihood ratio (LR) tests.

Because the relation between duration and health state is unknown, a full model containing all interactions between attributes and duration was chosen. To predict values, a random-effects model was used to model whether or not a respondent preferred a scenario to being dead. The full model is:

$$V(Q(t))_{ij} = (\alpha + \underbrace{\beta_1' x_{ij}}_{\text{Attributes}}) + (\underbrace{\beta_2 t + \beta_3 t^2}_{\text{Duration}})$$
$$+ (\underbrace{\beta_4' x_{ij} t + \beta_5' x_{ij} t^2}_{\text{Attributes} \times \text{Duration}}) + (\underbrace{\varepsilon_{ij} + \varepsilon_{ij} t + \varepsilon_{ij} t^2}_{\text{Random \quad effects}}) \tag{3}$$

where $i = 1,\ldots,n$ represents individual respondents, and $j = 1,\ldots,m$ represents health states. α is the intercept that sets $V(Q,(t))$ to zero for pr(BTD) = 50 % (see below), and $\beta_1', \beta_4', \beta_5'$ are matrices with weight estimates. x_{ij} is a matrix containing ten dummy variables of the EQ-5D attributes and the 'N3' term, β_2 and β_3 are weight estimates for durations t and t^2, and $\varepsilon_{ij} + \varepsilon_{ij} t + \varepsilon_{ij} t^2$ are the random effects for α, t and t^2, respectively. The β-weights are the values associated with the attributes, 'duration' terms, and

'attribute × duration' terms. The three random effects in the model allow, per respondent, a different intercept, as well as different linear and quadratic effects for duration. Thus, for each respondent, health states may have different values on the $V(Q(t))$ scale, and different time-dependent effects may be modelled. In subsequent analyses, simplifications of this model were considered by dropping the quadratic and linear 'attribute × duration' terms in steps. Models were compared using LR tests, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) [25, 26].

When respondents were on average indifferent between a certain scenario and being dead, that is pr(BTD) = 0.5, this meant that $V(Q^d(t)) \sim V(\text{dead})$. By formula 2, if $\text{pr(BTD)} = \frac{1}{1+e^{-V(Q^d,t)}} = 0.5$, this implied that $V(Q^d, t) = -\ln\left(\frac{1}{0.5} - 1\right) = 0$. Thus, on the $V(Q(t))$ scale, the value of being dead equalled 0.

To anchor the value so that full health was equal to 1, the latent values were divided by the latent value of state 11111, for which x_{ij} was 0, at $t = 10$ years. This duration was chosen because it was in the centre of our range and has commonly been used in other valuation methods. In Eq. 3, when $x_{ij} = 0$, the only remaining terms were α, $\beta_2 t$ and $\beta_3 t^2$, so the anchored value of health state Q_j with duration t would become:

$$V(Q_j(t))_{\text{anchored}} = \frac{V(Q_j(t))}{\alpha + 10\beta_2 + 10^2\beta_3} \tag{4}$$

To generate a visual representation of the time dependency assessed with the BTD preferences, we plotted the QALY model $V(Q(t)) \cdot t$ for a number of health states [24]. This analysis assumed zero time preferences. A selection of ten health states, spread out evenly over the value range at 40 years, was made to avoid the visual clutter from depicting all 50 health states.

4 Results

4.1 Determinants of BTD Preferences

The study population was comparable to the Dutch population aged 18–45 years in terms of gender, religion, educational status and marital status, compared with public data from the Dutch Bureau of Statistics (http://www.cbs.nl). The characteristics of the participants are listed in Table 1. Of the 291 respondents, 4 (1.4 %) did not complete the questionnaire for reasons unknown, resulting in 197 missing responses. The partially completed questionnaires were included in the analyses, resulting in a total of 291×108 questions $- 197 = 31{,}231$ answered BTD questions. Preferences better than being dead occurred in

Table 1 Population characteristics

Characteristic	Test ($n = 291$)	Retest ($n = 41$)
Gender		
Male	143 (49)	21 (51)
Female	148 (51)	20 (49)
Age		
18–24 years	84 (29)	17 (41)
25–29 years	35 (12)	2 (5)
30–34 years	45 (15)	4 (10)
35–39 years	47 (16)	7 (17)
40–45 years	80 (28)	11 (27)
Marital status		
Not married	131 (45)	21 (51)
Married	146 (50)	19 (46)
Divorced	14 (5)	1 (3)
Educational level		
Low	17 (6)	1 (2)
Middle	177 (61)	25 (61)
High	97 (33)	15 (37)
Religious life stance		
Yes	106 (36)	16 (39)
No	185 (64)	25 (61)
Belief in life after death		
Yes	147 (51)	24 (59)
No	144 (49)	17 (41)
Problems with mobility[a]		
Yes	40 (14)	5 (12)
No	251 (86)	36 (88)
Problems with self-care[a]		
Yes	13 (4)	0 (0)
No	278 (96)	41 (100)
Problems with daily activities[a]		
Yes	55 (19)	6 (15)
No	236 (81)	35 (85)
Problems with pain[a]		
Yes	108 (37)	16 (39)
No	183 (63)	25 (61)
Problems with depression[a]		
Yes	78 (27)	8 (20)
No	213 (73)	33 (80)
Time of day[b]		
0000–0600 hours	4 (1)	1 (2)
0600–1200 hours	27 (9)	6 (15)
1200–1800 hours	173 (60)	20 (49)
1800–2400 hours	87 (30)	14 (34)

The reported values are n (%). The percentages are rounded to sum to 100 %

[a] Self-described health status at the time of the interview

[b] Time of day when the questionnaire was completed

69 % of all BTD questions. Fifty-one responders (18 %) had lexicographic preferences, of whom 50 preferred all scenarios over being dead and one respondent preferred being dead over all scenarios. The median response time over all 108 BTD questions was 4 s, ranging from 2 to 20 s. Of the lexicographic respondents, 78 % were found to be in the fastest tertile of response times.

Gender, age and marital status were not associated with preferences. Having a religious life stance [raw BTD preferences 76 versus 65 % for a non-religious life stance; adjusted odds ratio (OR) 4.086 (95 % CI 1.996–8.363)] and a low educational level [80 % versus 67 % for medium and 71 % for high; ORs 6.942 (95 % CI 1.645–29.297) for low and 1.276 (95 % CI 0.607–2.680) for high, respectively, versus the medium educational level] were associated with more states being preferred to being dead. Scenarios were preferred to being dead more often by those indicating that they had problems with daily activities [76 versus 68 %; OR 4.340 (95 % CI 1.637–11.504)] and less often by those with problems on the pain attribute of the EQ-5D [68 versus 70 %; OR 0.420 (95 % CI 0.195–0.905)]. The presentation order of a scenario and the side on which 'dead' was presented (left or right) were not significant, nor was the time of day when the questionnaire was completed. Respondents in the slowest and medium tertiles of the median response time preferred fewer health states to being dead [ORs 0.12 (95 % CI 0.05–0.25) and 0.06 (95 % CI 0.03–0.13), respectively, versus the fastest tertile]. This relation remained after exclusion of respondents with lexicographic preferences.

The number of participants in block 1 was larger than those in the other blocks (138 participants versus 78 and 72) to yield enough retest respondents. Significant differences between the blocks existed in raw BTD preferences (65, 77 and 69 %), gender, age and response times but not in the responses to the warm-up questions. Full data are available in the Electronic Supplementary Material.

4.2 Test–Retest Properties

In the retest data, 15 out of 1,230 responses (41 respondents × 30 questions) were missing. The tetrachoric correlations showed overall agreement of 0.78 between the test and retest. For the fastest tertile of responders, the agreement was lower (0.45) compared with the medium and slowest tertiles (0.77 and 0.84, respectively).

4.3 Comparison with Previous Work

MET states were found in 15 % of all respondents and varied from 2 % to 6, 8, 9 and 17 % across the five health

states 11121, 11312, 13311, 11113 and 23232, respectively. These percentages are lower than the previously found percentages of 4, 13, 28, 22, and 26 %, respectively [12].

Of the 488 cases in which (Q, 20 years) was preferred to being dead, 44 cases (9 %) did not prefer both shorter durations over being dead. Of the 122 cases in which being dead was preferred to (Q, 5 years), 43 cases (35 %) preferred both longer durations over being dead. Both tests showed higher percentages of inconsistencies than the previous study, which showed 1 and 5 % of inconsistencies for these two tests, respectively [12]. For the second test, there were significantly more inconsistencies in the tertile with the fastest respondents (OR 2.78; 95 % CI 1.27–6.20).

4.4 Modelling Health-State Values

Figure 2 shows the raw data—that is, the proportion of respondents with BTD preferences for a selection of health states. If values are independent of duration, these lines should be horizontal. Table 2 shows estimates of the full model (Eq. 3) in the last three columns. The relevant attribute weights were consistent—that is, larger coefficients for level 3 than for level 2—and were significant, except for MO2. This model had the best performance of the two models presented. Table 2 also presents the reduced model containing only the EQ-5D dummy variables and 'duration' plus 'duration^2' terms. The negative
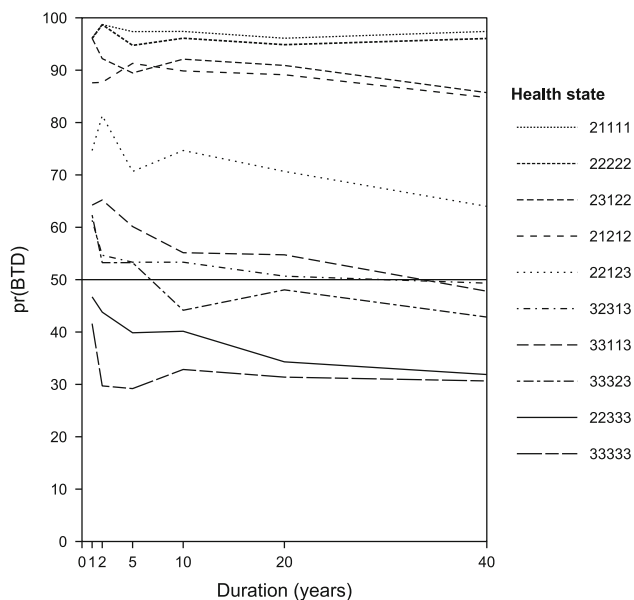
Fig. 2 Percentage of respondents who rated a scenario as better than being dead [pr(BTD)] for ten selected health states. Data were collected for $t = 1, 2, 5, 10, 20$ and 40 years. The *lines* were added for clarity

coefficient for duration in the reduced model reflects that preference strength declines with duration (see Fig. 2). The similarities between the estimated coefficients of the full and reduced model led to very similar predictions for $V(Q(t))$ (results not shown).

Figure 3 shows the estimates for $V(Q(t)) \cdot t$ and their relation with the duration of a health state, based on the full model. Independence of utility and duration would show up in this graph as straight lines fanning from the origin. Though some curvature can be seen in Fig. 3 (especially for the more severe health states), visual inspection suggests that the overall dependence on duration is small. Out of all 243 possible EQ-5D health states, the full model yielded 13 MET states (12233, 12333, 13233, 22233, 23233, 31133, 31332, 32132, 32232, 32331, 33132, 33231 and 33331), where MET is defined as (5 years X) preferred to being dead but (20 years X) not preferred to being dead. A list of QALY values $V(Q(t)) \cdot t$ for all 50 health states calculated using the full model can be found in Table 3 in Appendix 1.

5 Discussion

The BTD valuation method was used in a web-based questionnaire to value health states from the EQ-5D system. A consistent set of values was derived for the EQ-5D system. The values did not depend strongly on time, but, nevertheless, some MET states were detected. Test–retest properties were good but deteriorated for the fastest responders. The number of non-completers was small (1.4 %). The results indicated that, in the presence of introductory video and audio fragments, the respondents understood the tasks without help from interviewers or researchers.

Preferences depended on socio-demographic characteristics. The relation between religious life stance and preferring states to being dead was to be expected, as religiousness has been linked with non-trading in TTO [27] and difficulties in conceptualizing states worse than being dead, inadvertently linking them with euthanasia [28]. Self-described health states and a low educational level were also significant factors, consistent with findings in other studies [27, 29]. Other socio-demographic characteristics were not significant. Age has been found to be a determinant in other studies [30] but not here, possibly because of the restricted age range of our participants (18–45 years).

A comparison with the previous Dutch valuation study showed that weights for mobility, self-care and usual activities were similar—that is, the weights were between 0.123 and 0.184; pain/distress had the largest weight in both studies (0.414 and 0.419); however, the weight for

Table 2 Parameter estimates from modelled Better than Dead (BTD) preferences

Parameter	Reduced model			Full model		
	Estimate	SE	Pr(>|z|)	Estimate	SE	Pr(>|z|)
Intercept	1.054	0.038	<0.001	0.919	0.040	<0.001
MO2	−0.020	0.010	0.045	−0.023	0.017	0.177
MO3	−0.215	0.010	<0.001	−0.184	0.017	<0.001
SC2	−0.092	0.009	<0.001	−0.082	0.016	<0.001
SC3	−0.123	0.009	<0.001	−0.115	0.016	<0.001
UA2	−0.067	0.009	<0.001	−0.061	0.015	<0.001
UA3	−0.127	0.009	<0.001	−0.100	0.016	<0.001
PD2	−0.071	0.010	<0.001	−0.051	0.018	0.003
PD3	−0.414	0.010	<0.001	−0.373	0.017	<0.001
AD2	−0.076	0.009	<0.001	−0.064	0.016	<0.001
AD3	−0.241	0.009	<0.001	−0.214	0.016	<0.001
N3	−0.135	0.015	<0.001	−0.083	0.026	0.001
Duration	-0.604×10^{-2}	0.204×10^{-2}	0.003	1.077×10^{-2}	0.432×10^{-2}	0.013
Duration^2	0.063×10^{-4}	0.037×10^{-4}	0.095	-0.262×10^{-4}	0.097×10^{-4}	0.007
MO2 × duration				0.104×10^{-2}	0.279×10^{-2}	0.708
MO3 × duration				-0.539×10^{-2}	0.269×10^{-2}	0.045
SC2 × duration				-0.209×10^{-2}	0.262×10^{-2}	0.425
SC3 × duration				-0.047×10^{-2}	0.254×10^{-2}	0.854
UA2 × duration				-0.057×10^{-2}	0.248×10^{-2}	0.819
UA3 × duration				-0.327×10^{-2}	0.266×10^{-2}	0.219
PD2 × duration				-0.203×10^{-2}	0.275×10^{-2}	0.460
PD3 × duration				-0.745×10^{-2}	0.271×10^{-2}	0.006
AD2 × duration				-0.212×10^{-2}	0.254×10^{-2}	0.403
AD3 × duration				-0.313×10^{-2}	0.262×10^{-2}	0.233
N3 × duration				-0.562×10^{-2}	0.412×10^{-2}	0.173
$\text{MO2} \times \text{duration}^2$				-0.025×10^{-4}	0.066×10^{-4}	0.699
$\text{MO3} \times \text{duration}^2$				0.123×10^{-4}	0.063×10^{-4}	0.051
$\text{SC2} \times \text{duration}^2$				0.053×10^{-4}	0.062×10^{-4}	0.390
$\text{SC3} \times \text{duration}^2$				0.003×10^{-4}	0.060×10^{-4}	0.957
$\text{UA2} \times \text{duration}^2$				0.011×10^{-4}	0.058×10^{-4}	0.848
$\text{UA3} \times \text{duration}^2$				0.053×10^{-4}	0.063×10^{-4}	0.397
$\text{PD2} \times \text{duration}^2$				0.027×10^{-4}	0.064×10^{-4}	0.677
$\text{PD3} \times \text{duration}^2$				0.189×10^{-4}	0.064×10^{-4}	0.003
$\text{AD2} \times \text{duration}^2$				0.050×10^{-4}	0.060×10^{-4}	0.407
$\text{AD3} \times \text{duration}^2$				0.053×10^{-4}	0.062×10^{-4}	0.390
$\text{N3} \times \text{duration}^2$				0.083×10^{-4}	0.097×10^{-4}	0.391
Degrees of freedom		20			42	
Deviance		21,290			21,242	
Log likelihood		−10,645			−10,621	
Akaike information criterion		21,330			21,326	
Bayesian information criterion		21,497			21,677	

Parameter estimates divided by $U(11111, 10)$ to anchor full health on 1. For the Akaike information criterion and Bayesian information criterion, lower values mean better fit

All durations are in years. 2 and 3 stand for level of severity: 2 = some problems, 3 = severe problems. Example calculation: utility for health state 12111 using model B at 2 years is $0.919 - 0.082 + 2(1.077 \times 10^{-2} - 0.209 \times 10^{-2}) + 2^2(-2.62 \times 10^{-4} + 0.53 \times 10^{-4}) = 0.85$

AD anxiety/depression, *MO* mobility, *PD* pain/discomfort, *SC* self-care, *SE* standard error, *UA* usual activities
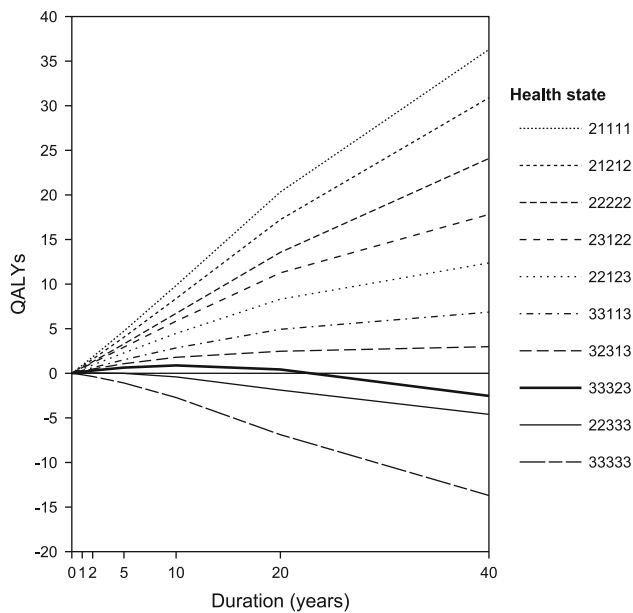
Fig. 3 Quality-adjusted life-years (QALYs) estimated using the full model for ten selected health states. State 33323 (*bold, solid line*) is a maximal endurable time state. Calculations were performed for t = 1, 2, 5, 10, 20 and 40 years. The *lines* were added for clarity

anxiety/depression was lower in this study (0.241 versus 0.408). While these converging results are encouraging, a comparison with our previous study [12] caused concerns about strategic behaviour—that is, skimping on the task. Response times in this study (median 4 s per BTD question) were much faster than the 20 s in the earlier study. While the number of lexicographic respondents was similar (18 and 19 % [12]), this study went on to show that of the lexicographic responders, 78 % were found to be in the fastest response time tertile. In the same vein, more inconsistencies were found in the present web-based study compared with the previous study, and this was, again, more frequent for the fast responders. These findings suggest that strategic behaviour may have affected the data quality. The causes may lie in any of the differences between the earlier study and the present one, such as the presence of an interviewer in the previous study, the use of a paper-based questionnaire in the previous study versus a stand-alone web-based questionnaire in this study, the smaller number of BTD questions in the previous study (15 versus 108) and the fact that the previous study followed the MVH protocol more closely [31]. This study also found fewer MET states. This suggests that 'how strongly values depend on duration' may be more sensitive to task characteristics than anticipated.

This concurs with the view that preferences are not pre-existing but are constructed by task demands [32]. It is our personal impression that in the web-based questionnaire used in this study, the data quality was lower—a view supported by mixed evidence [33–35].

Nevertheless, progress has been made. In the BTD method, positive and negative health states are valued using a single question. The method combines the advantages of a discrete-choice experiment (DCE), i.e. a low cognitive burden [36], and includes duration in the preference choices [37]. Values were estimated using a random-effects model, as has previously been used by others [15, 38]. In the resulting scale, the value of zero corresponds to the value of being dead, solving previous problems with the estimation of the health state 'dead' [17]. The random-effects model led to valid estimates, even though 18 % of the respondents had lexicographic preferences, mentioned earlier as being problematic [39]. The random factors per respondent appear to solve this problem.

6 Conclusion

The main goal of this study was to test the feasibility of using the BTD method to estimate values for multi-attribute health-state classification systems, such as the EQ-5D. The results of this study support the feasibility of the BTD method. Specifically, the method yielded consistent scores for the EQ-5D health attributes. The results showed that the values were mostly independent from duration, though some MET states were detected. Further research should explore how the experimental setting modifies the effect of time on values.

Appendix

See Table 3.

Table 3 Quality-adjusted life-years (QALYs) for all 50 health states by duration, using the full model

Health state	BTD QALYs by duration						Dutch EQ-5D value set, 10 years
	1 year	2 years	5 years	10 years	20 years	40 years	
21111	0.91	1.84	4.73	9.84	20.33	36.31	8.93
11211	0.87	1.75	4.51	9.34	19.22	34.56	8.97
11121	0.88	1.77	4.53	9.31	18.96	33.62	8.43
11112	0.86	1.74	4.46	9.20	18.85	34.43	8.05
12111	0.85	1.71	4.37	9.02	18.53	33.98	8.47
21212	0.78	1.57	4.05	8.38	17.23	30.89	7.37
21122	0.79	1.59	4.06	8.35	16.97	29.95	6.83
22112	0.76	1.53	3.91	8.07	16.55	30.32	6.87
12221	0.73	1.47	3.75	7.68	15.54	27.75	7.29
22221	0.71	1.43	3.66	7.52	15.29	26.86	6.93
22222	0.64	1.29	3.29	6.72	13.56	24.09	5.69
11312	0.67	1.34	3.34	6.62	12.73	21.64	5.14
11113	0.62	1.25	3.14	6.29	12.24	20.07	3.70
13311	0.62	1.24	3.12	6.22	11.99	19.25	4.86
21113	0.60	1.21	3.05	6.14	11.98	19.18	3.34
23221	0.59	1.18	2.98	5.99	11.63	17.93	3.89
23122	0.58	1.17	2.93	5.85	11.26	17.81	2.97
22321	0.58	1.15	2.85	5.60	10.53	16.71	4.34
32211	0.51	1.01	2.47	4.80	9.08	16.29	4.20
13322	0.50	1.00	2.45	4.73	8.64	12.90	2.76
32112	0.50	0.99	2.42	4.66	8.72	16.17	3.28
31321	0.50	0.98	2.37	4.48	8.00	12.76	3.91
22123	0.46	0.93	2.29	4.47	8.31	12.37	1.66
13213	0.45	0.89	2.24	4.44	8.40	12.27	1.86
11131	0.46	0.91	2.25	4.40	8.40	15.47	3.66
12313	0.44	0.87	2.11	4.04	7.31	11.04	2.31
33122	0.42	0.82	1.99	3.75	6.68	10.65	1.72
31322	0.43	0.84	2.00	3.68	6.27	9.99	2.67
32321	0.41	0.81	1.91	3.50	5.95	9.54	3.09
32222	0.39	0.76	1.80	3.31	5.73	9.94	2.10
33113	0.32	0.63	1.52	2.85	4.93	6.86	0.57
13132	0.28	0.55	1.30	2.40	4.20	7.54	0.90
11133	0.24	0.48	1.11	2.00	3.30	5.32	0.41
32313	0.25	0.48	1.08	1.79	2.47	2.98	0.70
23331	0.22	0.43	1.00	1.77	2.80	3.60	1.21
13232	0.22	0.42	0.98	1.74	2.84	4.90	0.58
32223	0.24	0.46	1.03	1.72	2.36	2.57	0.09
23232	0.19	0.38	0.88	1.59	2.58	4.00	0.22
33223	0.20	0.39	0.89	1.50	1.94	0.63	−0.61
22332	0.18	0.35	0.76	1.19	1.49	2.77	0.67
21233	0.16	0.31	0.70	1.19	1.68	1.78	−0.27
33131	0.16	0.29	0.63	0.95	1.10	2.25	0.53
33323	0.16	0.31	0.64	0.88	0.43	−2.53	−0.86
11333	0.14	0.26	0.54	0.73	0.42	−0.49	−0.16
32132	0.12	0.22	0.39	0.37	−0.22	1.42	−0.01
33231	0.09	0.17	0.31	0.29	−0.27	−0.39	0.21
31332	0.10	0.18	0.28	0.07	−1.05	−1.17	0.24

Table 3 continued

Health state	BTD QALYs by duration						Dutch EQ-5D value set, 10 years
	1 year	2 years	5 years	10 years	20 years	40 years	
32331	0.08	0.14	0.19	−0.10	−1.36	−1.62	0.66
22333	0.03	0.05	−0.01	−0.40	−1.88	−4.60	−1.34
33333	−0.17	−0.36	−1.09	−2.72	−6.88	−13.70	−3.29

The Dutch EQ-5D value set [21] can be compared to the Better than Dead (BTD) QALYs for a duration of 10 years in column 5. The table is sorted by BTD QALYs at 10 years

References

1. Drummond MF. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press; 2005.
2. Bijlenga D, Birnie E, Bonsel GJ. Feasibility, reliability, and validity of three health-state valuation methods using multiple-outcome vignettes on moderate-risk pregnancy at term. Value Health J Int Soc Pharmacoecon Outcomes Res. 2009;12(5):821–7.
3. Froberg DG, Kane RL. Methodology for measuring health-state preferences—II: scaling methods. J Clin Epidemiol. 1989;42(5):459–71.
4. Louviere JJ, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. Health Econ Policy Law. 2009;4(Pt 4):527–46.
5. Robinson A, Spencer A. Exploring challenges to TTO utilities: valuing states worse than dead. Health Econ. 2006;15(4):393–402.
6. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. Med Decis Making Int J Soc Med Decis Making. 1994;14(1):9–18.
7. Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. Med Care. 2007;45(3):238–44.
8. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care. 2005;43(3):203–20.
9. Dolan P, Roberts J. To what extent can we explain time trade-off values from other information about respondents? Soc Sci Med. 2002;54(6):919–29.
10. Macran S, Kind P. "Death" and the valuation of health-related quality of life. Med Care. 2001;39(3):217–27.
11. Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. Attitudes toward quality of survival. The concept of "maximal endurable time". Med Decis Making Int J Soc Med Decis Making. 1982;2(3):299–309.
12. Stalmeier PF, Lamers LM, Busschbach JJ, Krabbe PF. On the assessment of preferences for health and duration: maximal endurable time and better than dead preferences. Med Care. 2007;45(9):835–41.
13. Tsuchiya A, Dolan P. The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature. Med Decis Making Int J Soc Med Decis Making. 2005;25(4):460–7.
14. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012;31(1):306–18.
15. Viney R, Norman R, Brazier J, et al. An Australian discrete choice experiment to value EQ-5D health states. Health Econ. 2014;23(6):729–42.
16. 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 J Int Soc Pharmacoecon Outcomes Res. 2010;13(8):1005–13.
17. Arons AM, Krabbe PF. Probabilistic choice models in health-state valuation research: background, theories, assumptions and applications. Exp Rev Pharmacoecon Outcomes Res. 2013;13(1):93–108.
18. 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(3):348–61.
19. Brooks R. EuroQol: the current state of play. Health Policy. 1996;37(1):53–72.
20. Macran S, Kind P. Valuing ED-5D health states using a modified MVH protocol. In: 16th Plenary Meeting of the EuroQoL Group, Sitges, 6–9 November 1999, discussion papers; 2000; Institut de Salut Pública de Catalunya, Spain
21. Lamers LM, McDonnell J, Stalmeier PF, Krabbe PF, Busschbach JJ. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ. 2006;15(10):1121–32.
22. R: A Language and Environment for Statistical Computing [computer program]. Version 3.0.1. Package psych: procedures for psychological, psychometric, and personality research. Version 1.3.2. Northwestern University, Evanston, Illinois, USA2013. Package lme4: linear mixed-effects models using S4 classes. Version 0.999999-0. 2012. Vienna, Austria: R Foundation for Statistical Computing; 2012.
23. Rupp AA, Templin J, Henson RA. Diagnostic measurement: theory, methods, and applications. New York: Guilford; 2010.
24. Miyamoto JM, Wakker PP, Bleichrodt H, Peters HJ. The zero-condition: a simplifying assumption in QALY measurement and multiattribute utility. Manag Sci. 1998;44(6):839–49.
25. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974;19(6):716–23.
26. Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6(2):461–4.
27. van Nooten FE, Koolman X, Brouwer WB. The influence of subjective life expectancy on health state valuations using a 10 year TTO. Health Econ. 2009;18(5):549–58.
28. Devlin NJ, Hansen P, Selai C. Understanding health state valuations: a qualitative analysis of respondents' comments. Qual Life Res Int J Qual Life Aspects Treat Care Rehabil. 2004;13(7):1265–77.
29. Dolan P. The effect of experience of illness on health state valuations. J Clin Epidemiol. 1996;49(5):551–64.
30. Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ. 1996;5(2):141–54.
31. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095–108.
32. Slovic P. The construction of preference. Am Psychol. 1995;50(5):364.

33. Mulhern B, Longworth L, Brazier J, et al. Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. Value Health J Int Soc Pharmacoecon Outcomes Res. 2013;16(1):104–13.

34. Norman R, King MT, Clarke D, Viney R, Cronin P, Street D. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. Qual Life Res Int J Qual Life Aspects Treat Care Rehabil. 2010;19(4):499–508.

35. 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 HEPAC Health Econ Prev Care. 2013;14(Suppl 1):65–73.

36. Salomon J. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. Popul Health Metr. 2003;1(1):12.

37. Dolan P. Modelling valuations for health states: the effect of duration. Health Policy. 1996;38(3):189–203.

38. 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 heath states. Eur J Health Econ HEPAC Health Econ Prev Care. 2013;14(Suppl 1):33–42.

39. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: what it can do for health care research and how to do it. J Health Econ. 2007;26(1):171–89.