ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Preference-Based Assessments

Generating Utilities for the Château-Santé Base: A Novel, Generic, and Patient-Centered Health-Outcome Measure

Xin Zhang, PhD, Karin M. Vermeulen, PhD, Nic J.G.M. Veeger, PhD, Ruslan Jabrayilov, PhD, Paul F.M. Krabbe, PhD

ABSTRACT

Objectives: We have developed a new patient-centered, preference-based generic health-outcome measure, Château-Santé Base (CS-Base), which is based on a novel multiattribute preference response (MAPR) measurement framework. This study aimed to generate a first utility set for the CS-Base, making it suitable for use in health-economic evaluations.

Methods: CS-Base comprises 12 health attributes: mobility, vision, hearing, cognition, mood, anxiety, pain, fatigue, social functioning, daily activities, self-esteem, and independence, each with 4 levels. Our methodology to generate utilities for the CS-Base was 2-fold. First, we derived coefficients from patient MAPR data to calculate CS-Base values. Subsequently, these were normalized to a 0.0 to 1.0 utility scale, in which 0.0 signifies dead. The dead position was estimated using general population data from a discrete choice experiment (discrete choice experiment + dead), using a division-value strategy, which localize the position of states better or worse than dead.

Results: We analyzed MAPR data from 3222 patients and discrete choice experiment + dead data from 1995 respondents. All MAPR coefficients were negative, logically ordered, and significantly different from the reference level. The dead position was denoted by a division value of −148.385. Utility values spanned from −0.071 to 1.0, and only 53 of 16 777 216 states were deemed worse than dead.

Conclusions: This study introduced the first CS-Base utility set, underlining a 2-step utility derivation method. This method, blending societal and patient views, surpasses traditional preference-based approaches, yielding firmer results. However, improvement of the normalization procedure is expected. Estimating CS-Base utilities is an ongoing process that gains precision over time.

Keywords: health-outcome measure, preference-based, utility, value.

VALUE HEALTH. 2024; 27(10):1426–1435

Highlights

- A generic patient-centered preference-based health-outcome measure Château-Santé Base has been developed.
- This study introduces the first Château-Santé Base utility set using a 2-step approach.
- The 2-step approach applied a novel preference-based method, which might overcome traditional health-economic preference challenges and merged societal and patient views.

Introduction

Health technology assessment bodies and reimbursement authorities generally require studies to evaluate the value of health interventions. Many of the regulatory bodies recommend using a summary measure of health outcome, such as quality-adjusted life-years (QALYs), as the unit of health benefit.[1] Central to the computation of QALYs is the quality component, which is usually quantified in terms of concepts such as health status or health-related quality of life (HRQoL).[2] Such HRQoL measures suited for QALY computation are expressed in a single metric (utility) reflecting the overall quality of health status, anchored on a unidimensional scale with the position of “full health” at 1.0 and “dead” at 0.0.

Health-outcome measures that can be used to derive utilities are often categorized as preference-based measures. These measures typically consist of various attributes (items) covering different health domains (eg, physical, mental, and social health) to measure health status. Preference-based measures typically use preference or valuation methods to assign weights to specific health attributes levels. These weights are then used to calculate a single metric that represents the overall quality of a person’s health status.[3] Commonly used preference-based methods stem from the field of health economics, such as standard gamble (SG)[4] and time trade-off (TTO).[5] These methods can directly produce a utility, which can be applied in QALY computation. However, it is worth noting that these methods have faced criticism because of both theoretical and empirical limitations (eg, including time preference, being cognitively demanding, and loss aversion).[6]

An alternative to these conventional methods is the discrete choice experiment (DCE), which entails comparisons of 2 or more hypothetical health states. This method has been rapidly adopted because it is easier to perform and free from some of the limitations of conventional methods.[7] In its customary form, DCE lacks the inclusion of elements relating to time duration or the state of “dead.” Consequently, the weights generated are not anchored to "dead" and are thus not suitable for calculating QALYs. For this reason, researchers have explored adapted DCEs, such as DCE with duration[8] and with "dead" (DCE + Dead).[9] However, such valuation assessments that involve life years or "dead" are complex and can pose challenges. Such DCE tasks may place a high cognitive demand on respondents, making it difficult to come to valid decisions when dealing with composite description of health states.[10] This observation may be particularly relevant in instances where DCE tasks are characterized by designs that may be deemed inadequate, user interfaces that could be improved, or instructions that are overly scholarly in nature. Additionally, respondents might struggle with trading off between health states and dealing with sensitive issues, such as being dead. Therefore, the outcomes may lack robustness. For example, such DCEs might yield non-monotonic coefficients.[11,12] Moreover, if something goes wrong with such a compound valuation study, it is difficult to detect the origin of the disturbance.

Because healthcare is publicly funded and members of the general public are taxpayers and potential users of the healthcare system, the prevailing approach is to derive utilities from the general public, which represents the societal perspective[13] (National Institute for Health and Care Excellence. Guide to the methods of technology appraisal: London: National Institute for Health and Care Excellence (NICE); 2013 Apr 4; 2013 [Internet]. Process and methods guides No. 9.). However, empirical studies have highlighted that patient preferences (based on experience with their health conditions) differ from the general population's preferences and that these differences have an impact on utilities and cost-effectiveness analysis results.[14-16]

More recently, a new preference-based measurement framework has been introduced, with the multiattribute preference response (MAPR) model being central to it.[17–19] A number of health-outcome measures have been developed using this novel framework, one of which is the Château-Santé Base (CS-Base) used in this study.[19,20] The CS-Base is a patient-centered, preference-based generic patient-reported outcome (PRO) measure. It was fully patient centered in its development and construction, with the content of the outcome measure fully selected by patients and its values generated based on patient responses. The importance of patient involvement and the measurement of PROs have gained substantial recognition.[21] Regulatory agencies actively encourage the recording of PROs to supplement conventional clinical assessments.[13,22,23]

We have previously conducted a study in which health-state values were obtained from patients using the CS-Base.[24] However, these values cannot be directly used to calculate QALYs because they are not anchored to "dead." To create a utility scale anchored on 0.0-dead to 1.0-full health, a separate study is required to rescale these original values. Therefore, this study was designed to generate utilities for the CS-Base through a 2-step process. In the first step, we estimated weights for the attribute levels using the preference-based Drop-Down (DD®, patent pending) method of the MAPR model, based on the view of respondents familiar with the health states under evaluation (patients). In the second step these weights (and their corresponding values) were normalized to a utility scale and integrated the general population perspective.

Methods

Samples

This study involved a sample of patients with various health conditions and a sample of respondents from the general population, both from the United States. Both samples were nationally representative for age and sex. The samples were reached through a market research company (Survey Sampling International, based in Rotterdam, The Netherlands, now named Dynata). The general population sample was recruited in July 2018, the patient sample in November 2020 and January 2022. An online survey, containing a link to our study tasks, was distributed by Dynata, which continued to recruit until reaching the desired sample size. Participants received a small financial incentive from Dynata, based on prior agreements. The study was exempt from Medical Ethics review under Dutch law (the Medical Research Involving Human Subjects Act) because it involved low-risk survey research without identifiable information. Survey responses were anonymous. Dynata adheres to ethical standards according to the market research Society Code of Conduct. Informed consent was obtained through questionnaire completion, with the study's purpose explained beforehand.

Health-Outcome Measure

The CS-Base comprises 12 health attributes: mobility, vision, hearing, cognition, mood, anxiety, pain, fatigue, social functioning, daily activities, self-esteem, and independence. Each attribute consists of 4 levels ranked by severity (level 1—optimal level, indicating no problems; level 4—the most suboptimal level, indicating the most severe problems, see Fig. 1). According to the 12 attributes and 4 levels, a health state can be generated that forms an overall health description expressed as 12 digits. For example, the full health state is expressed as "111 111 111 111" and the worst health state as "444 444 444 444." The total number of possible health states that can be generated by CS-Base is 16 777 216 (4^{12}, 12 attributes each with 4 levels).

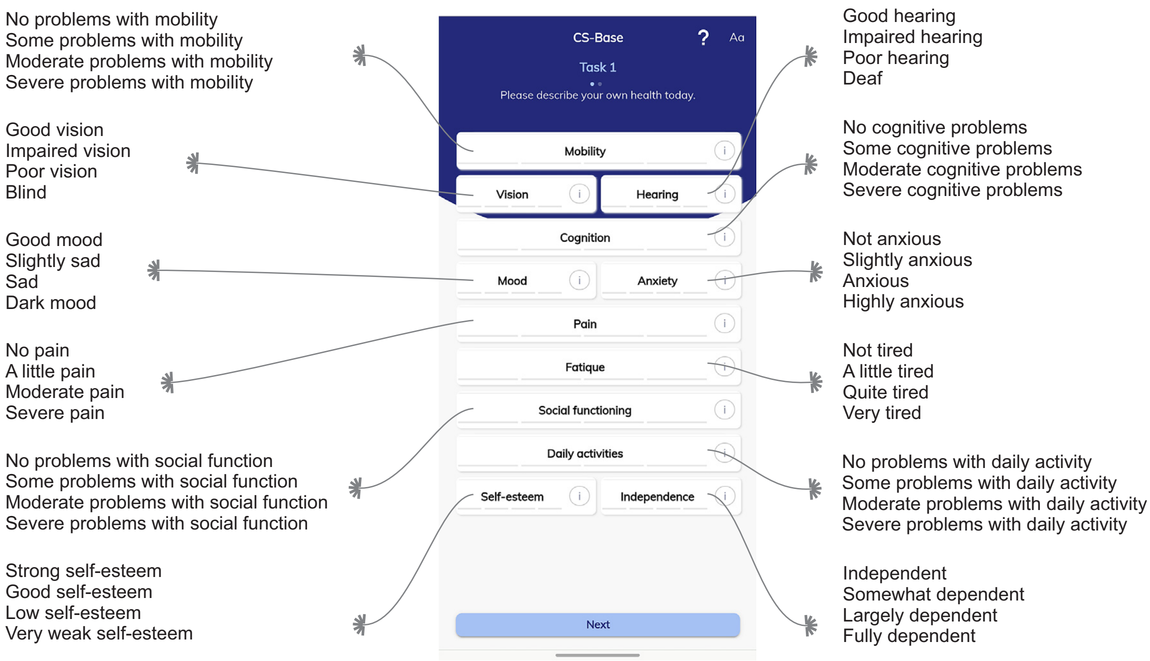
Response Tasks

Different preference-based tasks were applied to collect responses from patients and from respondents of the general population. For patients, the response tasks were designed according to the MAPR framework, which we called MAPR tasks (elaborated below). In the general population sample, the DCE + Dead method was used.

MAPR tasks

Founded on the MAPR framework, 2 response tasks were included: a descriptive task (task 1) and a preference task. In task 1, patients described their health today, by indicating the level of problems for each attribute. Thus, an overall description of their health states, expressed as 12 digits was generated. For task 2, the DD method was used as the preference task within the MAPR framework.[24,25] Patients were presented with their own health state as assessed in task 1 and instructed to select a suboptimal level (2, 3, or 4, not level 1) for the attribute that most hindered them. This was done by clicking or swiping this attribute and dropping down one level lower (eg, from level 3 to 2, indicating a transition from worse to better health status). Patients then repeated this process for the attribute that hindered them the second most. We limited the number of DD selections to a maximum of 5 (patients could make between 1 and 5 selections). If a full health state (111 111 111 111) was described in task 1, the patients will not proceed to task 2 because there are no attributes at suboptimal levels (levels 2, 3, or 4) that can be lowered down. Additional information regarding the execution of the 2 tasks can be found in Appendix Figure 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.06.013. The DD method produced ranked health states, which were treated as ordinal data for the analysis of coefficients (weights of attribute levels) estimation. These weights were subsequently used to generate an overall

Figure 1. The 12 attributes of the generic health-outcome measures CS-Base, each with 4 levels, as depicted in the HealthSnApp (an application for mobile phones).



CS indicates Château-Santé.

value for each health state. The full operation of the 2 MAPR tasks was run in a mobile application: HealthSnApp (www.healthsnapp.info). Because task 2 (DD) was fully dependent on the responses produced in task 1, no research design was needed (which was the case for the DCE, see below).

DCE

Regarding the DCE design in this study, each DCE task comprised 3 comparisons: a comparison of 2 hypothetical CS-Base health states (state A versus B), followed by a comparison of state A with "dead" and state B with "dead" (Appendix Fig. 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.06.013). Based on several criteria (Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.06.013), 437 pairs were selected encompassing 865 health states. Of this selected set, 33 were severe states (only levels 3 or 4 on all attributes) and were presented in 20 pairs.

Analysis

To generate CS-Base utilities, we applied a 2-step approach.

Step 1: estimation of coefficients—we estimated the coefficients of the attribute levels of the CS-Base using data from the MAPR tasks performed on the patient sample. Subsequently, we computed values based on these coefficients. The ordinal response data (ranked health states) obtained with the MAPR tasks were analyzed using a rank-ordered logit choice model (cmrologit, Stata 17.0). This analysis yielded parameter estimates presented as regression coefficients, which reflect the weights of the attribute levels. We designated the first level (no problems) of each attribute as the reference category. The coefficients for the remaining 3 levels were then estimated using 36 dummy variables (12 × 3).

Step 2: normalization of coefficients and values—we normalized (rescaled) the coefficients and values derived in step 1 to fit a utility scale ranging from 0.0 to 1.0, in which 0.0 represents "dead." The position of "dead" was determined using data gathered via the "DCE + Dead" method from a general population sample. We implemented 2 strategies to estimate this position.

As a first and basic strategy, we used the "division value" method to differentiate states that are considered better than dead (BTD) from those regarded as worse than dead (WTD). Sullivan et al used a similar strategy in creating EQ-5D-5L value sets. They applied an adaptive DCE and a binary search algorithm to identify the range of health states in which "dead" lies; the average value of these 2 states was then taken as the value for "dead."[10] In this study, we calculated the proportion of all severe states that were perceived as WTD. States with a 50% proportion of being considered as WTD were assigned a value of 0. This marked the point at which the likelihood of choosing "dead" or a health state in the DCE task became equal.[26] This point was used as the division value, representing the position of "dead." To derive normalized coefficients, we divided the estimated MAPR coefficients by this division value. Finally, CS-Base utilities for the health states were determined using the formula: one minus the sum of all normalized coefficients.

The second extended strategy involved estimating coefficients for the 36 attribute levels and one for "dead." To obtain the utilities, we first divided the remaining coefficients by the estimated "dead" coefficient, subsequently, the utility of a health state was calculated as 1 minus the sum of coefficients for the corresponding levels on the health attributes. We have used this strategy to generate utilities for an infant HRQoL instrument.[26] For estimating the 37 coefficients we analyzed the ordinal response data, which included 3 ranked states: states A, B, and "dead," using also the rank-ordered logit choice model (cmrologit, Stata 17.0). For states A and B, we designated the first level (no problems) of each attribute as the reference category. The coefficients for the remaining 3 levels were then estimated using 36 dummy variables

Table 1. The number of patients and the general population for each demographic subgroup.

Sociodemographic factors	Patients (N = 3222)	General population (N = 1995)
Gender, *n* (%)	3196 (99)	1995 (100)
Male	1434 (45)	946 (47)
Female	1762 (55)	1049 (53)
Age, year	3197 (99)	1995 (100)
Mean	47	43
Age groups, *n* (%)	3197 (99)	1995 (100)
18-27	476 (15)	337 (17)
28-37	656 (20)	533 (27)
38-47	567 (18)	315 (16)
48-57	469 (15)	378 (19)
58-67	552 (17)	432 (22)
≥ 68	477 (15)	NA
Education, *n* (%)	2638 (82)	1995 (100)
More than high school	1374 (52)	1642 (82)
High school graduate	1067 (40)	321 (16)
Less than high school	197 (7)	32 (2)
Ethnicity, *n* (%)	3082 (96)	1995 (100)
Asian/Asian-American	105 (3)	99 (5)
Black/African American	285 (9)	165 (8)
Hispanic or Latino American	171 (6)	88 (4)
Native American/Inuit/Alaskan	51 (2)	13 (1)
Native Hawaiian/Pacific Islander	17 (1)	3 (0)
White American/Caucasian	2426 (79)	1599 (80)
Other	27 (1)	28 (2)

(12 × 3). Another dummy was set for “dead” (0 if health state A or B was selected, 1 if “dead” was selected). With this extended strategy, it is possible to compare the estimated regression coefficients based on responses from the general population with the coefficients from the patients (MAPR model). All computations and the visualization of the results were performed using Stata 17.0 and CorelDRAW 23.0.

Results

Completion

The survey of MAPR tasks was sent to patients of which in total 4104 responded. We had to exclude 882 patients because our routine to deal with respondents who had dropped down an attribute more than once was not operating well in all cases. This left 3222 patients whose responses were included in the regression model. For the estimation of the coefficients, 681 patients reported a full health state (task 1), thus no DD response was generated, whereas 7 reported impaired health states but had no records of a DD response because of technical issues of the software. Therefore, the coefficients were estimated based on the responses of 2534 patients (with 16 465 ranked health states generated, Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.06.013).

The survey of DCE tasks was sent to 2000 respondents, each respondent performed 10 DCE tasks (3 comparisons). One hundred percent completion would generate 20 000 DCE responses: specifically, 20 000 comparisons of state A with B; 40 000 comparisons of state “dead” D with A or B (20 000 comparisons each with A or B); and 20 000 responses of ranks (ranked as 1, 2, and 3) of 3 states. Finally, 19 999 DCE responses were collected (1 missing response from the expected 20 000). However, 1724 responses led to illogical ranking of the 3 health states (eg, state A > B, A < D, and B > D) and were excluded (Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.06.013). This specifically excludes 5 respondents (representing 50 responses) and other illogical responses (1674) from other respondents (their logical responses were included). Thus, 1995 respondents remained, with 18 275 DCE responses and 36 550 comparisons including “dead” (18 275 comparisons each with state A or B, Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.06.013). The 18 275 ranks of 3 states were included in the analysis to estimate the coefficients for the general population.

General Information About Respondents

The analysis of the patient sample included 3222 respondents (Table 1). The mean and median ages of the patients were 47 and 46 years, with ages ranging from 18 to 94 years. More than half were women (1762; 55%), and the majority were white American/Caucasian (1187, 81%). More than half of the patients (1374, 52%) had more than high school education. The patients reported 14 kinds of health conditions, including pain (1511, 47%), fatigue/sleep problems (1237, 38%), mental health problems (865, 27%), diabetes (587, 18%), hearing or vision loss (531, 16%), eczema (374, 12%), gastrointestinal disease (316, 10%), heart disease (244, 8%), cancer (181, 6%), rheumatism (149, 5%), epilepsy (88, 3%), and other diseases (352, 11%).

The analysis of the general population included 1995 respondents (Table 1). The mean and median ages were 43 and 41 years, with ages ranging from 18 to 65 years. There were 1049 (53%) females, 1599 (80%) were white American/Caucasian, and 1642 (82%) respondents had more than a high school education.

States Worse Than “Dead”

In the DCE + Dead tasks, each of the 865 states were considered WTD at least once (in 1 comparison). The probabilities of a state being considered WTD ranged from 3% (state 114 112 122 432 was considered WTD in 1 out of 40 presented comparisons) to 63% (state 344 444 444 443 was considered WTD in 26 out of 41 presented comparisons). The probabilities of the 33 severe states being considered WTD ranged from 21% (state 434 444 344 444) to 63% (state 344 444 444 443) (Table 2).

Among the 33 severe states, 2 were considered WTD at a probability of 50%, with the mean value of the 2 states −148.405. Because the limited number (2) of states considered WTD with a probability of exactly 50%, the precision of the estimated position of “dead” may have been diminished. Therefore, we included a window by enlarging the probability from 50% to a range of 45% to 55%. Within this window, 12 states were included. The mean value of these 12 states was −148.385. This division value was used for the normalization procedure.

Coefficients

The estimated coefficients derived from the MAPR tasks collected from patients were all negative and in logical order (eg, no problems, slight problems, moderate problems, and severe problems, Table 3). All the coefficients were statistically significant compared with the reference level, and all had small confidence intervals, with only level 4 of the cognition attribute showing a large confidence interval. The coefficients for each of the 3 levels (levels 2, 3, and 4) showed consistency across 12 attributes (Fig. 2). Variation in coefficients were evident between all levels across all

Table 2. States with the highest proportion being considered WTD.

Health state	Proportion WTD (%)	Number of times states presented	Number of times states considered WTD	WTD state	Severe state
344 444 444 443	63	41	26	1	1
444 443 444 434	55	44	24	1	1
444 411 314 411	53	43	23	1	0
344 444 434 444	53	38	20	1	1
444 434 443 444	51	39	20	1	1
444 344 443 444	50	42	21	1	1
444 434 344 444	50	42	21	1	1
314 224 412 333	50	40	20	1	0
444 444 443 443	49	41	20	1	1
344 434 444 444	49	41	20	1	1
444 443 443 444	48	48	23	1	1
342 344 112 223	46	39	18	1	0
444 444 444 334	46	39	18	1	1
421 241 414 243	46	39	18	1	0
144 421 232 133	46	39	18	1	0
444 444 443 434	46	82	38	1	1
342 131 313 434	46	39	18	1	0
141 243 133 422	45	40	18	1	0
344 444 444 344	45	42	19	1	1
444 434 434 444	45	42	19	1	1
231 214 411 244	45	40	18	1	0
444 344 434 444	44	79	35	0	1
444 434 444 344	44	48	21	0	1
344 444 443 444	44	119	52	0	1
334 444 444 444	41	37	15	0	1
444 444 344 434	39	38	15	0	1
444 444 443 344	38	39	15	0	1
443 443 444 444	38	42	16	0	1
444 443 444 443	36	44	16	0	1
344 444 444 434	36	44	16	0	1
444 444 444 343	36	45	16	0	1
443 444 444 443	35	40	14	0	1
444 444 334 444	34	41	14	0	1
434 443 444 444	33	33	11	0	1
444 444 344 344	32	40	13	0	1
434 344 444 444	31	39	12	0	1
443 444 444 344	30	118	35	0	1
434 434 444 444	30	84	25	0	1
444 434 444 443	26	42	11	0	1
443 444 444 434	24	42	10	0	1
434 444 344 444	21	43	9	0	1

Note. Proportion WTD (%) = Number of times that a state was considered WTD/Number of times that a state was presented. WTD state = state is considered WTD and belongs to the top 20 states with the highest proportions: 1 = yes, 0 = no. Severe state = state with only levels 3 or 4 on all attributes: 1 = yes, 0 = no.
WTD indicates worse than dead.

Table 3. Coefficients of CS-Base attribute levels.

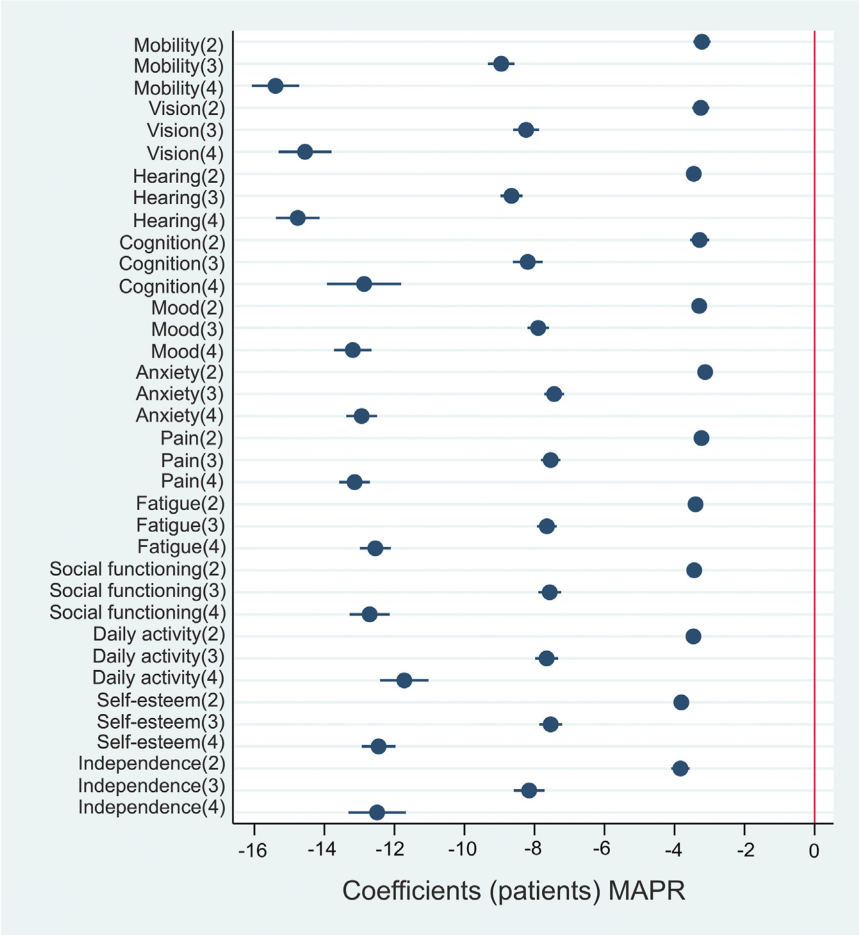
Attribute levels	MAPR (patients)			Normalized MAPR (utility scale: combination of patients and general population)
	Coefficients	SE	*P* value	Coefficients
Mobility (2)	−3.22	0.13	<.001	−0.022
Mobility (3)	−8.95	0.19	<.001	−0.060
Mobility (4)	−15.40	0.35	<.001	−0.104
Vision (2)	−3.25	0.12	<.001	−0.022
Vision (3)	−8.24	0.19	<.001	−0.056
Vision (4)	−14.55	0.39	<.001	−0.098
Hearing (2)	−3.45	0.10	<.001	−0.023
Hearing (3)	−8.66	0.14	<.001	−0.058
Hearing (4)	−14.76	0.32	<.001	−0.099
Cognition (2)	−3.28	0.14	<.001	−0.022
Cognition (3)	−8.19	0.22	<.001	−0.055
Cognition (4)	−12.87	0.54	<.001	−0.087
Mood (2)	−3.30	0.10	<.001	−0.022
Mood (3)	−7.89	0.16	<.001	−0.053
Mood (4)	−13.19	0.22	<.001	−0.089
Anxiety (2)	−3.13	0.09	<.001	−0.021
Anxiety (3)	−7.44	0.14	<.001	−0.050
Anxiety (4)	−12.94	0.22	<.001	−0.087
Pain (2)	−3.23	0.09	<.001	−0.022
Pain (3)	−7.54	0.14	<.001	−0.051
Pain (4)	−13.14	0.22	<.001	−0.089
Fatigue (2)	−3.40	0.09	<.001	−0.023
Fatigue (3)	−7.65	0.14	<.001	−0.052
Fatigue (4)	−12.55	0.23	<.001	−0.085
Social function (2)	−3.44	0.10	<.001	−0.023
Social function (3)	−7.56	0.17	<.001	−0.051
Social function (4)	−12.71	0.29	<.001	−0.086
Daily activities (2)	−3.46	0.10	<.001	−0.023
Daily activities (3)	−7.65	0.17	<.001	−0.052
Daily activities (4)	−11.72	0.35	<.001	−0.079
Self-esteem (2)	−3.81	0.11	<.001	−0.026
Self-esteem (3)	−7.54	0.17	<.001	−0.051
Self-esteem (4)	−12.45	0.25	<.001	−0.084
Independence (2)	−3.83	0.13	<.001	−0.026
Independence (3)	−8.15	0.22	<.001	−0.055
Independence (4)	−12.50	0.42	<.001	−0.084

Note. Negative coefficients implied that a particular level was worse than the reference level, which in our study was the first level of each health attribute. Moreover, the less preferable a level was considered, the more its coefficient had a negative direction.
Examples of calculating utility: 212 311 123 112 = 1 − (0.022 + 0 + 0.023 + 0.055 + 0 + 0 + 0 + 0.023 + 0.051 + 0 + 0 + 0.026) = 0.800.
Eg, for level 2 of mobility, normalized coefficient; Normalized MAPR coefficients = MAPR coefficients/148.385 (division value). = −(3.22/148.385) = −0.022.
CS indicates Château-Santé; MAPR, multiattribute preference response.

attributes. Specifically, the coefficients for level 3 of all the 12 attributes were consistently lower than any of the level 2 coefficients.

The division value (−148.385) which distinguishes states BTD from WTD was used to normalize the MAPR coefficients. The normalized coefficients were equal to the coefficients derived from the MAPR responses divided by −148.385. For example, for level 2 of the mobility attribute, the MAPR coefficient was −3.22, the normalized coefficient was calculated as −0.022 (−[3.22/148.385]). The normalized level 2 coefficients for all the 12 attributes ranged from −0.026 (independence and self-esteem) to −0.021 (anxiety, Table 3); those of level 3 for all 12 attributes

Figure 2. Distribution of coefficients (and their 95%CI) derived from the MAPR responses collected from patients.



MAPR indicates multiattribute preference response.

ranged from −0.060 (mobility) to −0.050 (anxiety); and those of level 4 ranged from −0.104 (mobility) to −0.079 (daily activities). Regarding the coefficients derived from the DCE + Dead data collected from the general population, nonmonotonic coefficients (eg, positive coefficients, higher weight of level 2 than level 3) were observed for more than half of the 36 coefficients (Appendix Table 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.06.013).

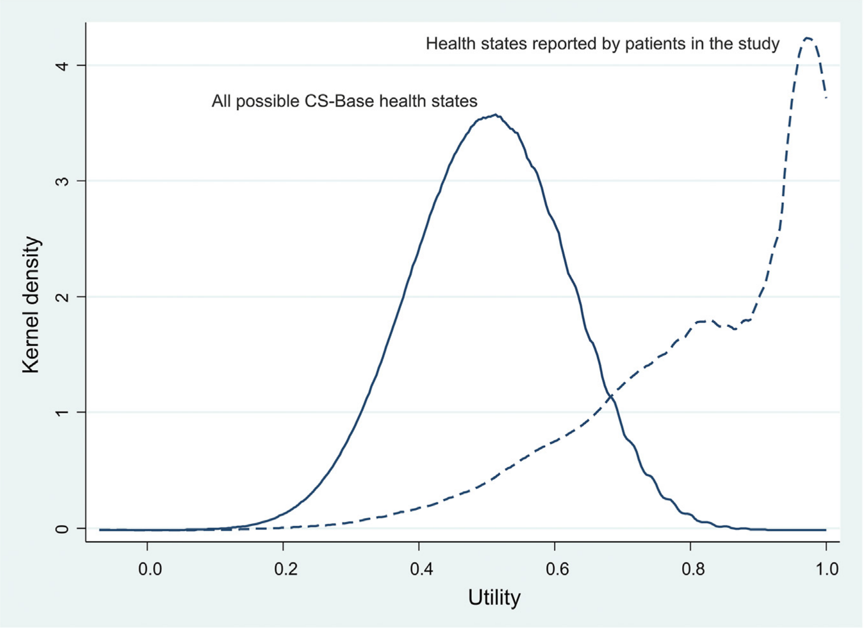
Health States, Values, and Utilities

In the MAPR task 1, 1774 unique health states were reported by patients. There were 681 (21%) patients who reported the full health state (111 111 111 111). One patient reported the worst state (444 444 444 444), which was the only reported state considered WTD by patients in the study. The values of all 16 777 216 CS-Base states ranged from −158.763 (worst state 444 444 444 444) to 0.0 (full state 111 111 111 111). The utilities were estimated solely on the first normalization strategy (division value); the second strategy did not perform well in our study. The utilities (normalized values) of all CS-Base states ranged from −0.071 (state 444 444 444 444) to 1.0 (state 111 111 111 111). Among the 16.777,216 states, 53 (0.0003%) were valued WTD based on utilities (<0.0, Appendix Table 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.06.013). The second lowest utility (0.111) was assigned to state 342 444 443 344. There were 2 states with only level 3 or 4 (334 434 343 334 and 343 333 333 333) reported by patients: their utilities were 0.184 and 0.314. Most patients reported mild impaired health states: 2388 (74%) patients reported states with utilities in the range of 0.5 to 1.0, with only 153 (5%) patients reporting states with utilities lower than 0.5 (Fig. 3). Specifically, 878 (27%) patients reported states in the utility range of 0.9 to 1.0, whereas 547 (17%) reported states in the utility range of 0.8 to 0.9, and 963 (30%) reported states in the utility range of 0.5 to 0.8. The distribution of utilities for all possible CS-Base states and for the 1774 states reported by patients in the study are shown in Figure 3.

Discussion

In this study, we implemented a 2-step procedure to generate utilities for the CS-Base, relying on perspectives from both patients and the general population. The innovative MAPR measurement framework was used to establish values for CS-Base health states through patient responses. These values were subsequently rescaled into utilities by anchoring (reference point) the position of “dead” on the CS-Base scale. This reference point was

Figure 3. Distribution of utilities for all the possible CS-Base health states and for the states reported by patients in this study.



CS indicates Château-Santé.

determined through a DCE that included "dead" and was based on responses from the general population.

The first step in our attempt to derive values for the CS-Base health states was based on the MAPR measurement framework, which encompasses a new and statistically powerful preference-based method (DD).[24,25] The DD method shares some similarities with the recently introduced Kaizen task.[27] The major difference between these 2 methods is that the DD method is based on patient responses, as opposed to responses from the general population. The Kaizen method necessitates its own study design, in contrast to the DD method, which relies directly on patient-assessed health states, eliminating the need for constructing a separate study design. The coefficients produced with the DD method were all statistically significant and showed a logical order. In addition, they demonstrated good face validity. In the DD method, patients are not asked to assess hypothetical health states, which can be difficult for them to imagine. They are only asked to assess their own health state and to indicate which attributes (levels) disturbed them the most. Therefore, the DD method could be simpler to execute compared with conventional preference-based methods (eg, TTO and SG). Furthermore, the DD method is based on ranking health states, presenting a more straightforward and logical approach compared with methods that requires numerical, absolute expressions or involve complex assessments as seen in the SG and TTO. However, it cannot directly generate utilities, because of the absence of "dead" or time duration attributes in the tasks.

Grounded in a robust measurement framework, the MAPR tasks are integrated into a mobile application named HealthSnApp. This app facilitates a seamless transition between tasks and is designed for easy self-completion. The application is useful in several ways. First, it facilitates data collection from patients by allowing them to perform the tasks remotely. Another advantage is that, combined with an integrated database server, the HealthSnApp also allows for rapid storage and retrieval of data. Furthermore, each study using the HealthSnApp can produce its unique value set because task 1 (description) and task 2 (preference) are integrated. It stands apart from commonly used preference-based health-outcome measures in its consistent operation. Specifically, in the case of EQ-5D studies conducted in various countries, 3 significant distinctions are observed in its construction and analysis: the use of various preference-based methods, the implementation of different designs, and the use of a range of analyses on the data.[28,29]

Anchoring health-state values to "dead = 0" has been a convention in health-state valuation and QALY computation for years.[30-32] Nevertheless, the use of the concept of "dead" in health measurement methods is a controversial issue[33,34]; some find it astonishing that health economists make "dead" a central element of their valuation methods.[35] More recently, this practice has been questioned, arguing that it is not essential to the theoretical foundations of health-status measurement suitable for cost-utility analysis.[36] Moreover, a critical issue is the lack of a reliable method to determine the position of "dead" on such utility scales. In this study, we applied 2 strategies to determine the position of "dead" on the utility scale. One method was detecting the value on the scale (division value), which divided states that were considered BTD from those considered WTD. The alternative method included "dead" as a separate element in a DCE task and as a distinct regression coefficient in the analysis. The first method performed well, however, the second method yielded counterintuitive results.

Based on the estimated division value from the DCE + Dead tasks, the lowest health-state utility for the worst CS-Base health state (444 444 444 444) was −0.071. Consistently, the utilities for the worst state presented in some of the widely used generic health-outcome measures are usually negative. For example, it is −0.57 in the US version of EQ-5D-5L,[37] in the original EQ-5D-3L it was −0.59.[38] However, there are also generic health-outcome measures that have had positive values for the worst state, for example, 0.32 and 0.20 in the quality of well-being index and the SF-6D, respectively.[39] The proportion of WTD states in the CS-Base (0.0003%, 53/16 777 216) was considerably smaller compared with those observed in EQ-5D-5L utility sets from various countries, which ranged from 0% to 36%.[10] Many reasons might lead to this difference. The valuation approach used by EQ-5D-5L studies are different from each other and from ours. In our study, preferences were collected from both patients and the general population, whereas in EQ-5D studies, they are usually collected from the general population. Consequently, CS-Base utilities may not be suitable for comparison with EQ-5D utilities. Nevertheless, the current valuation approach (division value) we used has room for improvement. For instance, integrating a computer adaptive system to refine the valuation process and enhance the precision in identifying where "dead" lies on the scale of health states.

The data collected through the DCE + Dead method yielded counterintuitive results for half of the coefficients. This might be associated with the high cognitive demands of the tasks. Respondents may encounter difficulty in making trade-offs between 2 hypothetical health states, or between health states and "dead."[40] Another possible reason might be that the number of parameters (36) in our case was relatively large, especially compared with that of other health-outcome measures (eg, 20 for EQ-5D-5L). An additional explanation may be attributed to the limited variability in attributes between the compared states. We included pairs that differed on 4 attributes and overlapped on 8. However, the CS-Base comprises 12 attributes, suggesting that the proportion of varied attributes might not be sufficient for estimating robust coefficients. Furthermore, upon examining other studies that used ranking data to estimate health coefficients, we speculate that the nonmonotonic coefficients observed in our DCE + Dead data could be due to the restricted number of ranked states (3 in our study). In contrast, other studies applied at least 5 to 8 ranks of health states,[41] even up to 15 ranks.[42] It is noteworthy that when the ranking task includes a substantial number of ranks (16), it seems fair results can be obtained even with a small sample size.[34] In future studies, improvements can be made by addressing these aspects explained above.

In the literature, there appears to be a lack of consensus regarding the most suitable perspective, whether it should be societal or patient centered, from which to derive utilities. Both vantage points present their unique strengths and drawbacks.[15,43] Leveraging our 2-step strategy, the CS-Base utilities were generated by integrating the patient's view and that of the broader society. The health-state values stem directly from patients acquainted with those health states firsthand. Concurrently, the localization of "dead"—essential for normalizing these values onto a utility scale—originates from a broader societal sample. As such, the patients determine the weights of the attributes and the location of the health states on the scale, the general population determines the interpretation of the scale.

Conclusion

This study presented a first utility set for the CS-Base and provided a preliminary exploration, showing how the utility generation process can be carried out using our 2-step approach. This approach offers the advantage of combining both patient and societal perspectives in utility generation for health-outcome measures. It also overcomes some drawbacks of conventional preference-based methods and produces more robust coefficients for generating utilities. However, improvement of the normalization procedure is expected. Estimating CS-Base utilities is an ongoing process that gains precision over time. By improving the valuation protocol, we anticipate achieving more consistent outcomes in the future.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2024.06.013.

Article and Author Information

Accepted for Publication: June 18, 2024

Published Online: August 8, 2024

doi: https://doi.org/10.1016/j.jval.2024.06.013

Author Affiliations: Department of Epidemiology, University Medical Center Groningen, Groningen, The Netherlands (Zhang, Vermeulen, Veeger, Krabbe); Metyis, Amsterdam, The Netherlands (Jabrayilov); Château Santé, Zeist, The Netherlands (Krabbe).

Correspondence: Paul Krabbe, PhD, Château Santé B.V., Prinses Irenelaan 16, 3708 EL, Zeist, The Netherlands. Email: p.f.m.krabbe@umcg.nl or p.krabbe@chateau-sante.com

Author Contributions: *Concept and design*: Krabbe, Vermeulen, Zhang, Jabrayilov
Acquisition of data: Research market companies (Survey Sampling International, Dynata)
Analysis and interpretation of data: Zhang, Krabbe, Vermeulen
Drafting of the manuscript: Zhang, Krabbe, Vermeulen
Critical revision of the paper for important intellectual content: Krabbe, Vermeulen, Zhang
Statistical Analysis: Zhang, Krabbe
Provision of study materials or patients: Research market companies (Survey Sampling International, Dynata)
Obtaining funding: Zhang, Krabbe
Administrative, technical, or logistic support: Krabbe, Veeger, Jabrayilov
Supervision: Krabbe, Vermeulen

Funding/Support: Financial support for this study was provided by a PhD grant sponsored by the Chinese Scholarship Council (CSC) and the University Medical Center Groningen (UMCG). Xin Zhang is employed by the sponsors.

Role of the Funders/Sponsors: The funding serves as the budget for Xin Zhang's PhD study. The funding agreement guaranteed the independence of the authors in designing the study, interpreting the data, and writing and publishing the manuscript.

REFERENCES

1. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford, England: Oxford University Press; 2015.
2. Krabbe PFM. *The Measurement of Health and Health Status: Concepts, Methods, and Applications From a Multidisciplinary Perspective*. London, UK: Academic Press; 2017.
3. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. *Cost-Effectiveness in Health and Medicine*. Oxford, England: University Press; 2016.
4. von Neumann J, Morgenstern O. *Theory of Games and Economic Behavior*. 16th ed. Princeton, NJ: Princeton University Press; 2004.
5. Torrance GW. Health states worse than death. In: van Eimeren W, Engelbrecht R, Flagle CD, eds. *Third International Conference on System Science in Health Care*. Berlin, Germany: Springer Verlag; 1984:1085–1089.
6. Lugnér AK, Krabbe PFM. An overview of the time trade-off method: concept, foundation, and the evaluation of distorting factors in putting a value on health. *Expert Rev Pharmacoecon Outcomes Res*. 2020;20(4):331–342.
7. McFadden D. Economic choices. *Am Econ Rev*. 2001;91(3):351.e378.
8. Bansback N, Hole AR, Mulhern B, Tsuchiya A. Testing a discrete choice experiment including duration to value health states for large descriptive systems: addressing design and sampling issues. *Soc Sci Med*. 2014;114(100):38–48.

9. Shah KK, Ramos-Goñi JM, Kreimeier S, Devlin NJ. An exploration of methods for obtaining 0 = dead anchors for latent scale EQ-5D-Y values. *Eur J Health Econ*. 2020;21(7):1091–1103.
10. Sullivan T, Hansen P, Ombler F, Derrett S, Devlin N. A new tool for creating personal and social EQ-5D-5L value sets, including valuing "dead". *Soc Sci Med*. 2020;246:112707.
11. Xie S, Wu J, He X, Chen G, Brazier JE. Do discrete choice experiments approaches perform better than time trade-off in eliciting health state utilities? Evidence from SF-6Dv2 in China. *Value Health*. 2020;23(10):1391–1399.
12. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health dead QALY scale. *Med Decis Mak*. 2015;35(3):328–340.
13. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal. https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781; Published 2013. Accessed May 20, 2024.
14. Ludwig K, Ramos-Goñi JM, Oppe M, Kreimeier S, Greiner W. To What extent do patient preferences differ from general population preferences? *Value Health*. 2021;24(9):1343–1349.
15. Rowen D, Mulhern B, Banerjee S, et al. Comparison of general population, patient, and carer utility values for dementia health states. *Med Decis Mak*. 2015;35(1):68–80.
16. Elliott J, Tsuchiya A. Do they just know more, or do they also have different preferences? An exploratory analysis of the effects of self-reporting serious health problems on health state valuation. *Soc Sci Med*. 2022;315:115474.
17. Krabbe PFM. A generalized measurement model to quantify health: the multi-attribute preference response model. *PLoS One*. 2013;8(11):1–12.
18. Groothuis-Oudshoorn CGM, van den Heuvel ER, Krabbe PFM. A preference-based item response theory model to measure health: concept and mathematics of the multi-attribute preference response model. *BMC Med Res Methodol*. 2018;18(1):62.
19. Gibson AEJ, Longworth L, Bennett B, Pickard AS, Shaw JW. Assessing the content validity of preference-based measures in cancer. *Value Health*. 2024;27(1):70–78.
20. Krabbe PFM, van Asselt ADI, Selivanova A, Jabrayilov R, Vermeulen KM. Patient-centered item selection for a new preference-based generic health status instrument: CS-Base. *Value Health*. 2019;22(4):467–473.
21. Porter ME, Larsson S, Lee TH. Standardizing patient outcomes measurement. *N Engl J Med*. 2016;374(6):504–506.
22. Food and Drug Administration. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims; availability. https://www.regulations.gov/document/FDA-2006-D-0362-0005; Published 2009. Accessed June 23, 2023.
23. Zorginstituut Nederland. Guideline for economic evaluations in healthcare. https://english.zorginstituutnederland.nl/about-us/working-methods-and-procedures/guideline-for-economic-evaluations-in-healthcare; Published June 16, 2016. Accessed May 20, 2024.
24. Zhang X, Krabbe PFM. From simple to even simpler, but not too simple: A head-to-head comparison of the Better-Worse and Drop-Down methods for measuring patient health status. *BMC Med Res Methodol*. 2023;23(1):299.
25. Zhang X, Vermeulen KM, Krabbe PFM. Different frameworks, similar results? Head-to-head comparison of the generic preference-based health-outcome measures CS-Base and EQ-5D-5L. *Appl Health Econ Health Policy*. 2024;22(2):227–242.
26. Krabbe PFM, Jabrayilov R, Detzel P, Dainelli L, Vermeulen KM, van Asselt ADI. A two-step procedure to generate utilities for the infant health-related quality of life instrument (IQI). *PLoS One*. 2020;15(4):e0230852.
27. Craig BM, Rand K, Hartman JD. Preference paths and their kaizen tasks for small samples. *Patient*. 2022;15(2):187–196.
28. Norman R, Cronin P, Viney R, King M, Street D, Ratcliffe J. International comparisons in valuing EQ-5D health states: a review and analysis. *Value Health*. 2009;12(8):1194–2000.
29. 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):675–697.
30. Weinstein MC, Fineberg HV. *Clinical Decision Analysis*. Philadelphia, PA: WB Saunders; 1980.
31. Torrance GW, Thomas WH, Sackett DL. A utility maximization model for evaluation of health care programs. *Health Serv Res*. 1972;7(2):118–133.
32. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. *Value Health*. 2009;12(suppl 1):S5–S9.
33. Norman R, Mulhern B, Viney R. The impact of different DCE-based approaches when anchoring utility scores. *Pharmacoeconomics*. 2016;34(8):805–814.
34. Menzies NA, Salomon JA. Non-monotonicity in the episodic random utility model. *Health Econ*. 2011;20(12):1523–1531.
35. Kamm FM. *Morality, mortality: volume I: death and whom to save from it*. New York, NY: Oxford University Press; 1993.
36. Sampson C, Parkin D, Devlin N. Is anchoring at "dead" a theoretical requirement for health state valuation? *Health Econ*. 2024. https://doi.org/10.1002/hec.4863.
37. Pickard AS, Law EH, Jiang R, et al. United States valuation of EQ-5D-5L health states using an international protocol. *Value Health*. 2019;22(8):931–941.
38. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
39. Richardson J, Mckie J, Bariola E. Multiattribute utility instruments and their use. In: Culyer AJ, ed. *Encyclopedia of Health Economics* San Diego, CA: Elsevier; 2014.
40. Mattmann M, Logar I, Brouwer R. Choice certainty, consistency and monotonicity in discrete choice experiments. *J Environ Econ Policy*. 2018;8(2):109–127.
41. McCabe C, Brazier J, Gilks P, et al. Using rank data to estimate health state utility models. *J Health Econ*. 2006;25(3):418–431.
42. Craig BM, Busschbach JJV, Salomon JA. Keep it simple: ranking health states yields values similar to cardinal measurement approaches. *J Clin Epidemiol*. 2009;62(3):296–305.
43. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: a call to reconsider current guidelines. *Soc Sci Med*. 2016;165:66–74.