Article:
Rowen, D., Mukuria, C., Bray, N. et al. (2026) A United Kingdom value set for the EQ-5D-5L. Value in Health, 29 (5). pp. 858-869. ISSN: 1098-3015

https://doi.org/10.1016/j.jval.2026.03.008

ScienceDirect

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

Preference-Based Assessments

A United Kingdom Value Set for the EQ-5D-5L

Donna Rowen, PhD, Clara Mukuria, PhD, Nathan Bray, PhD, Jill Carlton, PhD, Louise Longworth, PhD, David Meads, PhD, Yemi Oluboyede, PhD, Ciaran O'Neill, PhD, Yaling Yang, PhD

ABSTRACT

Objectives: A UK EQ-5D-5L value set is urgently required to enable the latest version of EQ-5D to inform policy, including evidence submitted to the National Institute for Health and Care Excellence as part of health technology appraisals. This paper presents the EQ-5D-5L UK value set generated from preference data elicited using the time trade-off (TTO) technique with a representative sample of the UK public.

Methods: In-person and videoconference interviews were undertaken using the composite TTO method for 102 health states with a representative sample of UK adults (age 18 years and older) across England, Wales, Scotland, and Northern Ireland. Data quality was rigorously and independently assessed throughout the study. TTO data was modeled using a range of models, with the value set being selected as the preferred model using predefined criteria.

Results: Data quality standards were achieved. A total of 1200 interviews were conducted, 1102 (91.8%) by means of videoconference and 98 (8.2%) in person. The sample was representative of age, sex, ethnicity, and socioeconomic groups and proportionally representative for the 4 nations: England, Scotland, Wales, and Northern Ireland. TTO values highlighted by participants as ones they would reconsider were excluded. The value set was generated using a random effects Tobit model. The dimensions of anxiety/depression and usual activities have a greater relative impact on utilities than in the UK EQ-5D-3L value set, though the worst state value is comparable.

Conclusion: The study collected good-quality data from a representative sample of the UK public, modeled the data appropriately and transparently, to generate a UK EQ-5D-5L value set suitable for informing policy.

Keywords: EQ-5D-5L, preference elicitation, time trade-off, face-to-face, videoconference interview.

Cite this article as: Rowen D, Mukuria C, Bray N, et al. A United Kingdom value set for the EQ-5D-5L. *Value Health*. 2026; 29(5):858-869.
doi.org/10.1016/j.jval.2026.03.008

Highlights

- The article presents the development of a UK value set for EQ-5D-5L, which can be used to generate utility values for EQ-5D-5L responses.
- The study produced good quality data reflecting preferences of the UK public.
- The study benefited from the following: (1) use of both in-person and videoconference interviews that enhanced inclusivity; (2) study governance and input from a wide range of stakeholders; (3) independent assessment of data quality at 4 time points by a Quality Control team; and (4) final endorsement of data and analyses by the Quality Control team and study Steering Group.
- The dimensions of anxiety/depression and usual activities have a greater relative impact on utilities than in the UK EQ-5D-3L value set, though utility for the worst state is comparable.

Introduction

EQ-5D-5L is a self-reported generic measure of health-related quality of life that is widely used internationally.[1,2] EQ-5D-5L has 5 dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), each with 5 response options (ie, levels of severity) (no problems; slight problems; some problems; severe problems; extreme problems/unable to do) (see Appendix Box A1).[3] EQ-5D-5L is accompanied by a scoring based on preferences, which are used to generate utility values. This scoring is referred to as a value set, which generates utility values for every health state described by EQ-5D-5L. These values lie between 1 and −1, such that 1 is equivalent to full health and 0 is equivalent to dead; values below 0 indicate a health state is considered as worse than being dead, and by convention, the lowest value possible is typically −1. These values can then be combined with the length of life to generate quality-adjusted life-years (QALYs) for use in health technology assessment (HTA). The EQ-5D-5L was developed to improve on EQ-5D-3L, a previous version with 3 response options per dimension.[4] Although both EQ-5D-3L and EQ-5D-5L are commonly used, the use of EQ-5D-5L has been increasing because of its greater sensitivity in capturing health-related quality of life in patient populations.[5,6]

The National Institute for Health and Care Excellence (NICE) in England,[7] the Scottish Medicines Consortium, and the All-Wales Medicines Strategy Group recommend the use of the EQ-5D to generate QALYs for HTA, with utility values generated using a UK-specific value set derived from a representative sample of the public.

A previous EQ-5D-5L value set for England,[8] generated following an international protocol[9] and jointly modeling time trade-off (TTO) and discrete choice experiment tasks, was not recommended for use by NICE,[10] following concerns regarding data quality and modeling,[11-16] which the value set authors addressed in a rebuttal.[17] NICE currently recommends[10] that EQ-5D-5L data should be scored by mapping[18,19] onto the UK EQ-5D-3L value set[20] until a definitive value set is generated.

A UK EQ-5D-5L value set is required since: (1) the UK EQ-5D-3L value set is based on data collected in the 1990s[20] that is unlikely to reflect current societal preferences (particularly for mental health where attitudes may have changed[21]); (2) UK population demographics have changed over time meaning the composition of a representative sample has altered; (3) methods for eliciting health state preferences have advanced including both data collection and analysis; and (4) direct estimation of utility using a value set is preferable to indirect utility estimation based on mapping utilities from another measure with a different health state classification system.[22]

This study aimed to generate a new UK value set for the EQ-5D-5L reflecting the current preferences of the UK public. This study intended to generate a UK EQ-5D-5L value set suitable for informing policy, and hence, aimed to generate good-quality data from a representative sample of the UK public and to model this data appropriately and transparently

Methods

The protocol for the UK valuation of EQ-5D-5L has been detailed previously.[23] The protocol was informed by the international EQ-5D-5L valuation protocol[9] now successfully used in over 30 studies worldwide,[24-28] consideration of recommendations from the independent quality assessments of the previous EQ-5D-5L value set for England,[11-16] a systematic review of all published EQ-5D-5L value sets,[24] a pilot study in England assessing acceptability, feasibility and equivalence of in-person and videoconference TTO interviews[29] and sister study in Australia,[30] and input from patient and public involvement sessions. The study benefited from study governance consisting of regular input and final approval from a study Steering Group (of representatives from NICE, National Health Service England, Department of Health and Social Care, international experts, and EuroQol scientific leaders), an independent Quality Control team, and EuroQol's Executive Committee.[31]

The reporting in this article follows the checklist for reporting valuation studies (CREATE) checklist.[32] Ethical approval was granted from the University of Sheffield Research Ethics Committee (Reference 048042).

Recruitment and Sampling

The study aimed to conduct 1200 interviews with members of the UK public. Participants were recruited and sampled by a market research agency using a blended approach to ensure a mix of people were contacted to be invited to participate, including a large postal mailout targeting postcodes (sent to 84 324 households), Facebook posts, paid Facebook adverts, Twitter/X tweets, word of mouth/snowballing from participants, and posters. People interested in being interviewed completed a short screening survey either online or by phone. Given the recruitment strategy used, a response rate cannot be calculated (see Appendix Table A1 for details of how participants who registered an interest in being interviewed found out about the study).

Quota sampling was undertaken to ensure a representative sample was achieved. Quota groups combined age (18-40 years, 41-64 years, 65 years and older) and sex (male, female). Separately, there were quotas for ethnicity (White, non-White) and socioeconomic group (captured by the index of multiple deprivation using postcode [deciles: 1 or 2; 3 to 8; 9 or 10]). The sample included participants with and without health problems and was selected to be proportionately representative across England, Scotland, Wales, and Northern Ireland.

Preference Elicitation Method: TTO

Preferences were elicited using the composite TTO approach using version 2.7 of the EuroQol Valuation Technology (EQ-VT)[33] (Fig. 1). This approach combines a procedure for states better than dead, in which states have a maximum duration of 10 years, with an additional 10-year lead-time TTO for states worse than dead (making 20 years in total).[9] Participants trade years of life to avoid impaired health, with the number of years in full health varying until the participant is indifferent between the 2 options. Given the maximum duration in the lead-time task, the lowest TTO value possible was −1, and therefore, the data is considered censored at −1.

Selection of Health States for Valuation

A total of 102 health states were valued. Of these, 86 were in line with the international protocol[9]; the 16 additional states were selected because they had the best predictive performance in the extended design used in the Indian EQ-5D-5L valuation[25] (see et Rowen et al[23] for further details). Health states were separated into 12 blocks of 10 health states, and each block contained mild, moderate, and severe states as well as the worst state (Appendix Table A2). The block and ordering of health states were randomly selected at the point of interview.

The Interview

Participants chose to be interviewed through videoconference (using Zoom) or in-person (see Rowen et al[23] for further explanation of this decision), and in English or Welsh. Participants completed the EQ-5D-5L and then received an explanation of the TTO techniques for states better and worse than dead during the valuation of 2 practice states. Participants then valued 3 practice EQ-5D-5L states before completing TTO tasks for 1 block of 10 health states.

Participants completed a "feedback module" in which they were shown their implied rankings (including any ties) for the 10 health states, derived from their TTO responses. Participants were asked to indicate any states whose values they would now reconsider, though no further TTO data were collected for any states that were indicated, in line with the EQ-VT protocol (Appendix Fig. A1).

Participants were also asked questions about their socio-demographic characteristics, EQ-HWB-9,[34] and questions about their understanding and what they thought of the interview. Participants were offered £50 through bank transfer or Amazon vouchers as a "thank you" for their participation.

Data Quality

An EQ-VT support team monitored protocol compliance (see [23] for further details) and data weekly alongside the study team. All interviewers received comprehensive training and achieved protocol compliance in pilot interviews prior to interviewing study participants.

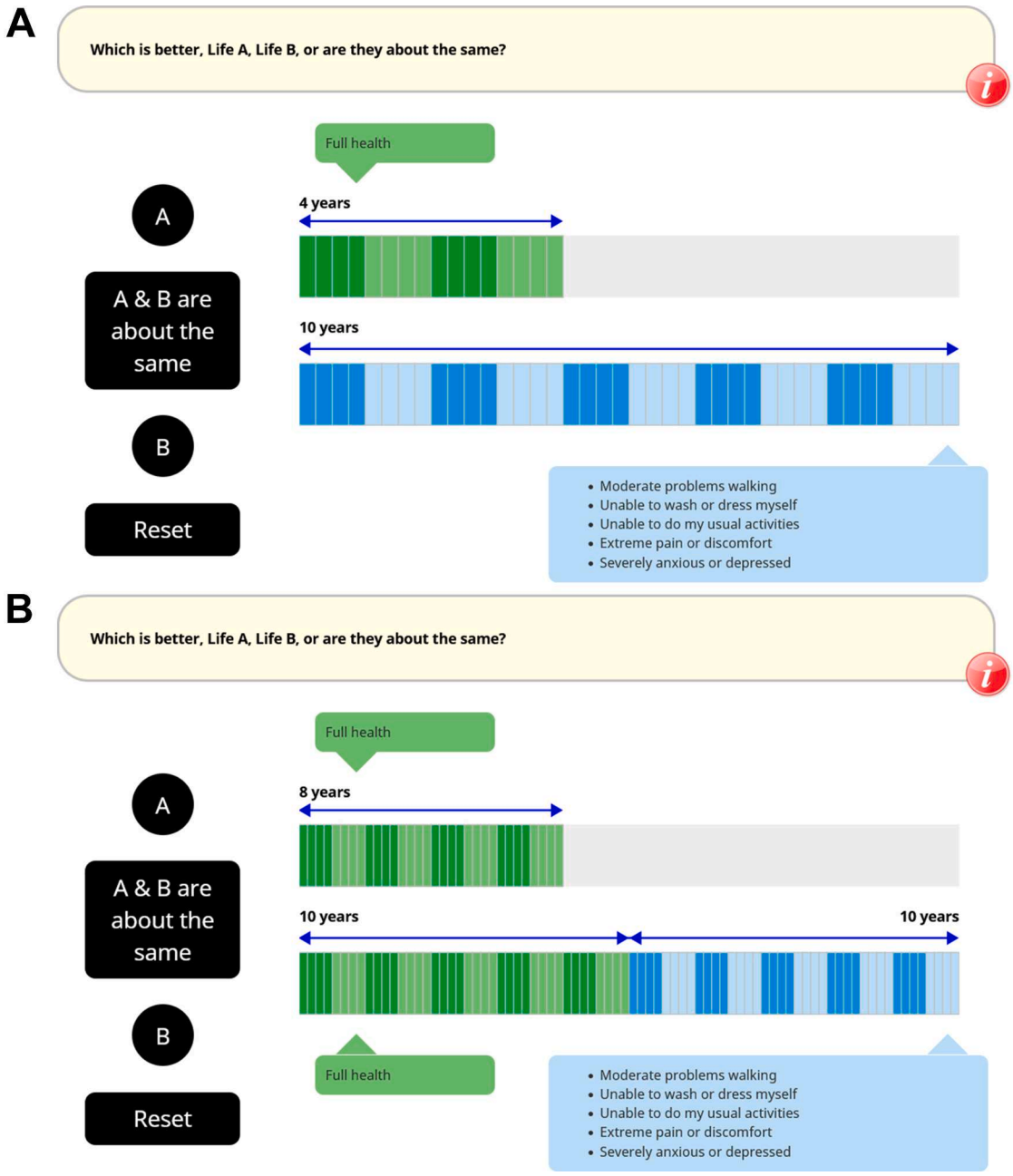
TTO data quality was assessed, including participant understanding and engagement (using both self-report and interviewer-report questions), examination of logical inconsistencies (where utility increases as health worsens), and proportions of the data at certain easy-to-achieve TTO values (−1, −0.5, 0, 0.5, 1).

Modeling

Exclusion criteria

All data were retained unless they met either of the following exclusion criteria: (1) data from interviewers with protocol non-compliance in 40% or more of their interviews; (2) TTO values identified by the participant using the feedback module as values

Figure 1. Example screenshots of EQ-VT version 2.1 system for valuing a state considered better than dead and a state considered worse than dead. (A) TTO screenshot for valuing a state considered better than dead. (B) TTO screenshot for valuing a state considered worse than dead.



TTO, time trade-off.

they would now reconsider (all remaining TTO values were retained).

Model specification

The dependent variable is TTO disvalue (1−TTO value) since the model estimates disutilities. The model estimates utility decrements for each severity level for each dimension using 20 parameter level dummies for levels 2 to 5 of each of the 5 dimensions, leaving level 1 as the reference category.

For a main effects random effects model, this can be specified as:

$$\mu_{ij} = 1 + \alpha + \beta_1 mo2_j + \beta_2 mo3_j + \beta_3 mo4_j + \beta_4 mo5_j + \beta_5 sc2_j + \beta_6 sc3_j + \beta_7 sc4_j + \beta_8 sc5_j + \beta_9 ua2_j + \beta_{10} ua3_j + \beta_{11} ua4_j + \beta_{12} ua5_j + \beta_{13} pd2_j + \beta_{14} pd3_j + \beta_{15} pd4_j + \beta_{16} pd5_j + \beta_{17} ad2_j + \beta_{18} ad3_j + \beta_{19} ad4_j + \beta_{20} ad5_j + u_i + \epsilon_{ij} \quad (1)$$

Where μ_{ij} is TTO disvalue of subject i for health state j, α is a constant term, β are the coefficients of the parameter level dummies *mo2* to *ad5*, wherein *mo* is mobility, *sc* is self-care, *ua* is usual activities, *pd* is pain/discomfort, *ad* is anxiety/depression, u_i

is a subject-level random effect and ϵ_{ij} is the error term that is assumed to be independent and identically distributed.[35] The utility value for each health state is calculated as 1 plus the sum of the utility decrements for the applicable level of each dimension.

Models were explored, removing the constant term and including interaction effects that evaluated whether there is a combined effect on utility from having 1 (or more) dimension(s) at a severe or extreme level (similar to the 'N3' term in the UK EQ-5D-3L value set[20]).

Model estimation

TTO data is characterized by the following:

- Censoring at −1, since participants cannot express a lower TTO value than −1, though they may wish to do so;
- Repeated observations (up to 10) per participant, since each participant values 10 TTO health states and the views and characteristics of the participant can affect their preferences and hence TTO values (note this is up to 10 observations per participant since participants can identify states they would now reconsider in the feedback module, and these are removed under exclusion criteria [2]);
- Larger variance for more severe health states, since TTO values for milder health are similar across participants and differ much more as states become more severe, suggesting that modeling should account for heteroscedasticity;
- Different preferences across participants, namely, preference heterogeneity.

We explored the following: (1) the Tobit model with censoring at –1; (2) generalized least squares (GLS) random intercept model; (3) heteroscedastic Tobit model with censoring at −1 (see Appendix Figure A2 for further details); and (4) random effects Tobit model with censoring at −1. Although these do not account for all characteristics of the TTO data outlined above, it is not, to our knowledge, possible to account for these characteristics simultaneously. The TTO data is not considered censored at 1 since values above 1 are not feasible or sensible. GLS does not account for censoring at −1, but is one of the more commonly estimated models in EQ-5D-5L value set studies.[24]

Estimation was undertaken in Stata version 18 (StataCorp, College Station, TX)[36] (Appendix Box A2). The data is available on request from the EuroQol Research Foundation.

Model performance

Model performance was assessed using the following: (1) logical consistency of coefficients; (2) number and proportion of significant coefficients (also using a model estimated with incremental dummies capturing disutility associated when moving between adjacent levels [eg, from level 2 to level 3]); (3) Akaike Information Criterion (AIC) and Schwarz (Bayesian) Information Criterion (BIC); and (4) mean absolute error (MAE).

Logical consistency of coefficients is a crucial requirement for a value set, since it means that, for example, level 4 of a dimension should have a utility decrement no larger in absolute terms than the level 5 decrement of the same dimension. If worse levels of health had smaller (or the same) disutility values than preceding levels, this would imply that health improvements result in a loss (or no change) in utility, which is counterintuitive.

Predictive ability of the model, assessed using MAE, is important given that we have TTO data for 102 of the 3125 possible EQ-5D-5L health states, and the model is required to predict values for all states. Generating MAEs for Tobit models can be criticized on the basis that the errors reflect the difference between predicted utility and observed TTO, but the 2 are not directly comparable because of censoring. This is because the observed TTO values are censored at −1, but the predicted utility from the Tobit model adjusts for the censoring at −1. To reflect this, separate MAEs have been generated comparing predicted utility to an "uncensored TTO mean," generated using the predicted value from a Tobit model estimated separately for each state.

The use of *P* values for informing policy decisions has been criticized by the American Statistical Association[37] and here, *P* values were only used alongside other criteria.

Robustness

Robustness of candidate models was assessed by comparing their results to identical models estimated on a subset of the dataset. Comparisons focused on the robustness of coefficients and their significance between models using the estimation sample and data subsamples. The models were compared for the following: (1) TTO data with and without TTO data excluded by the feedback module; (2) TTO data with and without participants who were reported by the interviewer as either not understanding and/or engaging with the TTO tasks; and (3) TTO data for the full 12-block design used here (102 states) and the smaller 10 block international protocol design (86 states).[9] Robustness concerns were raised if the sign, significance, logical consistency, and relative importance of the dimensions differed across subsamples.

Independent assessment and patient and public involvement

An independent Quality Control team accessed raw data, assessed the regression models, and checked reproducibility. Sessions with members from the NICE Public and Patient Involvement Program were undertaken during (1) protocol development, (2) the study comparing in-person and videoconference interviews, (3) confirming protocol development following the results of the comparison study, and (4) following selection of the value set. Each stage involved 2 sessions, and each session involved 2 to 4 members, and for 2 stages, an additional participant provided written feedback. Discussion within these sessions encompassed study design, sampling characteristics, participant recruitment strategies, interview settings, participant "thank you" payments, formulation of study materials, and the interpretation, acceptability, and dissemination of the study findings, including the value set. The perspectives elicited were subsequently incorporated into the project's design and implementation.

Selection of UK value set

The value set should have face validity in terms of the coefficients being theoretically sensible and appropriate with logical consistency, should account for data characteristics, should fit the data, and be fit for purpose for discriminating against different health states. The model selected as the value set was the best-performing model overall.

Results

The Sample

A total of 1200 interviews were conducted from January 2023 to September 2023, of which 1102 were conducted by videoconference and 98 were conducted in person, and 7 interviews were conducted in the Welsh language, with the remainder conducted in English. Interviews were conducted by 13

Table 1. Sociodemographic characteristics of the sample.

Characteristic	Categories	Final sample (n=1200), %	Videoconf interviews (n=1102), %	In-person interviews (n=98), %	UK general population, %*	*P* value (2-sample test of proportions) comparing videoconf/in-person†
Sex‡	Male	48.2	48.1	49.0	48.9	.867
	Female	51.5	51.5	51.0	51.1	.921
	Other	0.3	0.4	0		
Age	18 to 24 years	12.3	12.5	9.2	10.6	.334
	25 to 34 years	18.3	19.5	4.1	17.0	<.001
	35 to 44 years	15.6	16.2	9.2	16.1	.068
	45 to 54 years	16.5	16.7	14.3	16.9	.538
	55 to 64 years	14.2	14.0	16.3	15.8	.522
	65 to 74 years	19.8	18.5	34.7	12.7	<.001
	75 years and above	3.4	2.6	12.2	10.9	<.001
	Mean age (SD)	47.0 (17.3)				
Marital status	Single	32.5	31.9	39.8		.108
	Married/partner	53.7	55.4	33.7		<.001
	Separated	1.6	1.5	2.0		.705
	Divorced	8.8	8.1	16.3		.006
	Widowed	2.8	2.4	8.2		<.001
	Prefer not to say	0.7	0.7	0.0		
Main activity§	In employment or self-employment	56.7	59.0	30.6	57.2	<.001
	Retired	20.8	19.0	41.8	21.6	<.001
	Student	10.1	10.1	10.2	5.6	.967
	Unemployed	3.0	3.0	3.1	3.4	.970
	Long-term sick	3.3	2.9	7.1	4.2	.023
	Carer or volunteer	1.8	1.6	4.1		.083
	Not seeking work	1.7	1.6	2.0		.763
	Other	2.3	2.4	1.0		
	Prefer not to say	0.4	0.5	0		
Index of multiple deprivation (indicative index of socioeconomic group)	1 (lowest)	8.2	7.9	11.2	10.0	.249
	2	9.8	9.2	16.3	10.0	.022
	3	7.4	6.9	13.3	10.0	.021
	4	9.4	9.5	8.2	10.0	.658
	5	9.1	9.3	7.1	10.0	.486
	6	10.9	11.0	10.2	10.0	.813
	7	9.5	9.3	12.2	10.0	.334
	8	9.3	9.6	5.1	10.0	.139
	9	9.2	9.3	7.1	10.0	.469
	10 (highest)	8.8	9.0	7.1	10.0	.538
	Missing	8.5	9.1	2.0	0	
Ethnic group	White	84.4	84.0	88.8	84.8	.214
	Asian/Asian British	7.7	7.9	5.1	8.0	.319
	Black/African/Caribbean/ Black British	2.7	2.7	2.0	3.5	.688
	Mixed or multiple ethnic groups	2.7	2.8	1.0	1.8	.291
	Other	1.8	1.8	1.0	1.9	
	Prefer not to say	0.8	0.7	2.0		
Nation	England	84.1	86.1	61.2	84.4	<.001
	Scotland	8.3	7.9	13.3	8.2	.065
	Wales	4.8	3.6	17.3	4.6	<.001
	Northern Ireland	2.8	2.4	8.2	2.8	<.001
Education after the minimum school-leaving age		92.5	93.1	85.7		.008
Degree or equivalent professional qualification		73.8	75.0	59.2		<.001

continued on next page

Table 1. Continued

Characteristic	Categories	Final sample (n=1200), %	Videoconf interviews (n=1102), %	In-person interviews (n=98), %	UK general population, %*	*P* value (2-sample test of proportions) comparing videoconf/in-person†
Household living	Own your own home outright, or with a mortgage	62.1	63.2	49.0		.005
	Rent from the government/ housing commission/social housing	11.8	9.9	32.7		<.001
	Rent from the private sector	23.1	23.6	17.3		.160
	Other	2.8	2.9	1.0		
	Prefer not to say	0.3	0.4	0		
Annual household income before tax (including benefits)	Up to £5199	1.3	1.2	2.0		.462
	£5200 and up to £10 399	3.6	3.5	5.1		.399
	£10 400 and up to £15 599	5.5	4.7	14.3		<.001
	£15 600 and up to £20 799	6.1	6.1	6.1		.987
	£20 800 and up to £25 999	8.0	8.2	6.1		.475
	£26 000 and up to £31 199	7.8	7.3	14.3		.013
	£31 200 and up to £36 399	7.3	7.4	5.1		.392
	£36 400 and up to £51 999	16.4	17.1	9.2		.044
	£52 000 and above	30.5	32.1	12.2		<.001
	Prefer not to say	13.6	12.5	25.5		
Parent/guardian for child/children aged under 18 years		22.8	24.2	6.1		<.001
Day-to-day activities limited because of a health problem or disability‖	Yes, limited a lot	7.4	7.4	7.1		.914
	Yes, limited a little	26.3	26.0	28.6		.586
	No	65.9	66.2	63.3		.563
	Prefer not to say	0.4	0.4	1.0		
Experienced serious illness	In yourself	44.1	43.6	49.0		.308
	In family	75.7	75.9	73.5		.597
	In caring for others	34.9	34.1	43.9		.052
Location of interview	At home	87.0	94.7	0		
	Hired university room	4.9	0	60.2		
	Hired interview room	3.3	0	39.8		
	Other	4.8	5.3	0		

ONS indicates Office for National Statistics.
*Statistics for UK for age, sex, and nation are from the ONS' Mid-Year Population Estimates June 2020. Statistics for ethnicity are from the ONS' Mid-Year Population Estimates 2019 and corrected based on England and Wales, as the ONS had not published ethnicity for Scotland and Northern Ireland. The statistics on the main activity are for England and Wales only in the Census 2021. The census includes persons aged 16 and above, whereas this study only surveys persons aged 18 and above.
†*P* values have been generated using a two-sample test of proportions by mode and have not been calculated for "prefer not to say" or "other" categories.
‡A "prefer not to say" response option was also included but not selected by any participant.
§The main activity data from the census does not include the same categories as the main activity question included in this study, and hence it was not possible to categorize the general population into all response options to our main activity question.
‖Day-to-day activities limited because of a health problem/disability which has lasted, or is expected to last, at least 12 months (include problems related to old age).

interviewers, each undertaking between 50 and 120 interviews. Quotas for the study were achieved, meaning the sample is representative for age, sex, ethnicity, and socioeconomic group as set according to the quota groups, and is proportionally representative for the 4 nations (Table 1). The sample has mixed health, with some participants in poor health (see also Appendix Table A3 for self-reported EQ-5D-5L dimensions and EQ-VAS). The sociodemographic characteristics of the sample differ by whether the interview is conducted via videoconference or in person, as anticipated.[29]

The quota bands for age were wide, and within the age band of 65 years and above, the proportion of participants aged 75 years and older is below the proportion in the UK general population. The sample is broadly representative of non-quota characteristics, though education at degree level, proportion of students, and income are higher than is representative.

TTO Data

Each of the 102 selected health states was valued by at least 90 participants. All interviewers achieved protocol compliance, meaning no TTO values were excluded under exclusion criteria (1) (Appendix Table A4). Under exclusion criteria (2), 1320 (of 12 000) TTO responses were indicated by participants as responses

they would now reconsider, and hence, these are removed from the data used to generate the value set. The data, therefore, has 1200 participants with 10 680 TTO values.

The distribution of TTO values has peaks in the distribution observed at 1, 0.5, and −1, and the proportion of values representing whole years in the TTO task is higher (eg, 0.1, 0.2 rather than 0.15, 0.25) than responses at 6-month intervals (Appendix Fig. A3). The peaks at 1 and −1 are expected since every participant values at least 1 mild state and the worst state (Appendix Tables A2, A5, and A6). The distribution of TTO values per interviewer is comparable (Appendix Fig. A4).

The TTO value distribution for the 102 health states shows that the proportion of values at 1 is highest for the milder states, and the proportion of values at −1 is highest for the most severe states and the worst state (Appendix Tables A5 and A6). Standard deviation (SD) increases as severity increases (Table 2). The ordering of the states by TTO value is affected by which dimensions have higher relative importance across the severity levels and hence is not expected to perfectly correspond with the summed score of each level of each dimension (level sum score (LSS), ie, for state 12 111, LSS = 6). Overall, the mean TTO value reduces as LSS increases, with 1 exception between LSS of 13 and 14 (Fig. 2).

Data Quality

For the majority of participants, understanding was confirmed both using self-report and interviewer-report questions (Appendix Tables A7-8). Participants' views on the interview were positive overall (Appendix Table A9).

Data quality assessments indicate good data quality (Appendix Table A10). No mild states were valued lower than the worst state; the majority of participants (98.8%) did not value any state below the worst state; the proportion of participants (16.3%) with any inconsistent values was considered acceptable (though the literature has no accepted guidance on this); few participants (0.8%) had more than 2 inconsistencies following the feedback module. The proportion of participants giving values at easy-to-reach values (−1, −0.5, 0, 0.5, or 1) for all states is low (1.8%).

Modeling

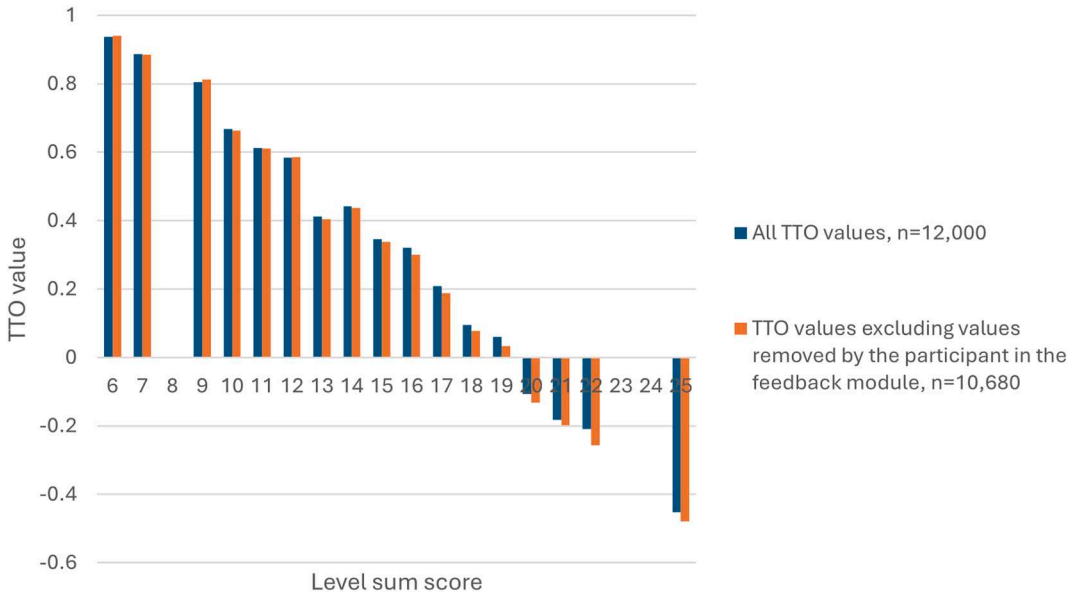
The models generate logically consistent coefficients, except for the heteroscedastic Tobit estimated with no constant (Table 3). The constant is never significant, but most other coefficients are significant (see Appendix Table A11 for the models estimated using incremental dummies). The MAE across all

Table 2. TTO value distribution by health state, excluding values removed by participants in the feedback module (ordered by observed mean TTO).

Health state	n	Mean	SD	Health state	n	Mean	SD	Health state	n	Mean	SD	Health state	n	Mean	SD
11211	187	0.954	0.088	53221	85	0.554	0.503	31525	84	0.332	0.487	53243	84	0.136	0.588
11121	295	0.947	0.102	11235	90	0.499	0.491	15151	89	0.326	0.576	52335	84	0.132	0.586
11112	197	0.945	0.112	32314	93	0.491	0.505	55233	80	0.324	0.51	23255	76	0.085	0.615
21111	199	0.942	0.164	45133	73	0.488	0.413	12543	83	0.319	0.513	24443	79	0.044	0.616
12111	289	0.922	0.193	52431	85	0.486	0.451	24342	84	0.318	0.557	21444	82	0.034	0.569
12121	86	0.912	0.169	13224	91	0.485	0.483	34342	84	0.311	0.571	55225	84	0.027	0.645
11122	92	0.895	0.148	11414	87	0.484	0.442	23152	104	0.307	0.592	51451	86	0.023	0.627
21112	101	0.887	0.212	23514	82	0.463	0.449	51124	77	0.305	0.632	35245	89	0.02	0.601
12112	106	0.884	0.287	12334	89	0.448	0.501	51152	90	0.294	0.516	34244	102	0.015	0.661
11212	101	0.88	0.295	25513	91	0.446	0.53	12244	86	0.244	0.554	34155	79	0.011	0.626
13122	94	0.865	0.202	31514	76	0.441	0.471	43514	101	0.237	0.532	54342	88	0.005	0.62
11221	99	0.863	0.303	53412	83	0.441	0.494	43315	84	0.235	0.588	53244	81	−0.016	0.637
11421	81	0.754	0.19	21334	93	0.432	0.54	21345	103	0.231	0.577	45144	89	−0.028	0.612
13313	87	0.738	0.337	54231	94	0.431	0.546	35143	84	0.228	0.527	54144	92	−0.051	0.634
25113	90	0.704	0.306	22434	92	0.423	0.498	55333	88	0.228	0.612	55424	94	−0.066	0.644
11532	99	0.684	0.336	21315	89	0.404	0.571	32443	81	0.227	0.542	54153	83	−0.069	0.632
25122	83	0.666	0.415	45233	79	0.395	0.507	22144	94	0.223	0.608	14554	91	−0.089	0.617
14113	91	0.663	0.411	45413	94	0.378	0.555	44342	79	0.217	0.599	24445	94	−0.12	0.627
25222	92	0.658	0.405	23242	95	0.376	0.553	43542	74	0.199	0.528	52454	95	−0.159	0.653
35311	89	0.63	0.419	12514	94	0.372	0.544	31443	95	0.198	0.598	44345	89	−0.172	0.623
42321	90	0.626	0.426	31524	86	0.372	0.508	24543	87	0.193	0.611	52455	86	−0.177	0.569
12513	89	0.614	0.369	11425	88	0.368	0.556	44125	91	0.185	0.579	44553	100	−0.216	0.606
25331	84	0.599	0.378	43324	78	0.356	0.545	12344	91	0.181	0.567	43555	85	−0.256	0.626
42431	77	0.587	0.376	23415	82	0.355	0.551	34515	92	0.176	0.608	55555	1056	−0.48	0.534
35332	97	0.579	0.431	52215	70	0.337	0.512	33253	85	0.174	0.626				
34232	91	0.554	0.438	42115	88	0.334	0.609	24553	89	0.144	0.556				

TTO indicates time trade-off; SD, standard deviation.

Figure 2. Mean TTO value by level sum score (summed value of each level of each dimension).



TTO indicates time trade-off.

health states is relatively low, always less than 0.05, and the MAE for the "uncensored mean" varies between 0.042 and 0.046. A third of the states each have MAE greater than 0.05, though there are a few health states with MAE greater than 0.1. The ordering of the dimensions with the largest utility decrement at level 5 is robust across models, with the order of impact from highest to lowest being: pain/discomfort, anxiety/depression, mobility, usual activities, and self-care.

Models estimated without a constant term have a lower AIC and BIC, meaning they are preferred. The MAE at the health state level is improved for the random effects Tobit, where no constant term is included, but not the MAE for the uncensored mean. MAE does not indicate an improvement in the heteroscedastic Tobit model with no constant term, and there are logical inconsistencies (self-care levels 2/3 and pain/discomfort levels 2/3).

Models estimated including interaction effects indicated that the interaction terms were not robust; their sign and size were not theoretically sensible, and they did not systematically improve model performance (Appendix Table A12).

Table 3 provides utility values for each model for some indicative states across the severity range, showing that while these values differ by model, they are of a similar magnitude.

Robustness

The results (Appendix Table A13) are largely robust across the subsamples assessed. The main models estimated on all TTO data are logically consistent and surprisingly have lower MAE than equivalent models excluding values removed by participants in the feedback module. The main models estimated, including participants reported as understanding and engaging by the interviewer, are better performing models, with lower MAE. Comparison of results for the block design used here and the smaller block design typically used in EQ-5D-5L valuation studies indicates that while MAE is lower for the standard design, logical inconsistencies are observed for the heteroscedastic Tobit models. This indicates that the use of the larger design has been beneficial. The relative importance of usual activities and self-care is very similar across models for the level 5 coefficients.

Selected Value Set

The UK EQ-5D-5L value set is generated using the random effects Tobit model with no constant term (Table 4).

The best performing models, considering logical consistency, proportion of significant coefficients, and MAE, are the GLS model, random effects Tobit, and heteroscedastic Tobit (Table A14). All of these models perform well with no clear better performing model across these performance assessments, and therefore, the purpose of the value set and the ability to account for data characteristics is also considered (Appendix Table A14). The GLS model performs best in terms of MAE but does not account for censoring or heteroscedasticity. The Tobit model has lower MAE at the health state level than the random effects Tobit and heteroscedastic Tobit models, but does not account for heteroscedasticity or repeated observations per participant. The random effects Tobit and the heteroscedastic Tobit both perform well in terms of proportion of significant coefficients, MAE, and accounting for TTO data characteristics, where the heteroscedastic Tobit model accounts for heteroscedasticity and the random effects Tobit accounts for repeated observations per participant (Appendix Fig. A5-6).

The random effects Tobit model was selected over the heteroscedastic Tobit model, since it produces estimates with lower MAEs and is robust to the removal of the constant term. The estimates provide higher discriminatory power between milder levels of severity, as there is a larger difference between utility decrements for the milder levels within a dimension. The random effects Tobit model is also a simpler model to estimate than the heteroscedastic Tobit, with no researcher choice around which variables and the functional form when accounting for heteroscedasticity. Models removing the constant term perform better according to BIC and MAE at the individual level, but do not perform better for MAE at the health state level. The value set model excludes the constant term, as doing so has the advantage that a consistent decrement is given for a change in a dimension regardless of whether that change is from full health or an impaired health state (see Appendix Table A15 for the covariance matrix of the value set).

Table 3. Modeling results.

Variables	Model 1 Tobit			Model 2 GLS			Model 3 Random effects Tobit			Model 4 Heteroscedastic Tobit			Model 5 Random effects Tobit no constant (UK value set)			Model 6 Heteroscedastic Tobit no constant		
	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals
Mobility level 2	0.024	0.018	.180	0.033	0.012	.005	0.028	0.013	.027	0.036	0.012	.003	0.032	0.012	.010	0.045	0.010	<.001
Mobility level 3	0.056	0.020	.005	0.063	0.013	<.001	0.056	0.014	<.001	0.055	0.018	.003	0.058	0.014	<.001	0.059	0.018	.001
Mobility level 4	0.171	0.021	<.001	0.177	0.014	<.001	0.177	0.015	<.001	0.169	0.018	<.001	0.179	0.015	<.001	0.171	0.018	<.001
Mobility level 5	0.279	0.019	<.001	0.265	0.013	<.001	0.278	0.014	<.001	0.288	0.017	<.001	0.279	0.014	<.001	0.290	0.017	<.001
Self-care level 2	0.022	0.017	.203	0.033	0.012	.006	0.033	0.013	.012	0.049	0.011	<.001	0.038	0.012	.002	0.060	0.009	<.001
Self-care level 3	0.056	0.020	.006	0.061	0.014	<.001	0.059	0.015	<.001	0.054	0.016	.001	0.060	0.015	<.001	0.052	0.016	.001
Self-care level 4	0.167	0.020	<.001	0.157	0.014	<.001	0.161	0.015	<.001	0.183	0.019	<.001	0.162	0.015	<.001	0.185	0.018	<.001
Self-care level 5	0.197	0.019	<.001	0.186	0.012	<.001	0.205	0.013	<.001	0.201	0.016	<.001	0.206	0.013	<.001	0.201	0.016	<.001
Usual activities level 2	0.046	0.018	.014	0.043	0.013	0.001	0.045	0.014	.001	0.050	0.012	<.001	0.049	0.013	<.001	0.059	0.010	<.001
Usual activities level 3	0.093	0.019	<.001	0.083	0.013	<.001	0.083	0.014	<.001	0.094	0.016	<.001	0.086	0.014	<.001	0.099	0.016	<.001
Usual activities level 4	0.176	0.019	<.001	0.176	0.013	<.001	0.181	0.014	<.001	0.180	0.016	<.001	0.184	0.013	<.001	0.184	0.015	<.001
Usual activities level 5	0.198	0.018	<.001	0.193	0.012	<.001	0.210	0.013	<.001	0.202	0.015	<.001	0.212	0.013	<.001	0.205	0.015	<.001
Pain/discomfort level 2	0.043	0.017	.012	0.052	0.012	<.001	0.050	0.013	<.001	0.040	0.010	<.001	0.056	0.012	<.001	0.050	0.007	<.001
Pain/discomfort level 3	0.051	0.020	.011	0.067	0.014	<.001	0.065	0.015	<.001	0.042	0.016	.007	0.066	0.015	<.001	0.043	0.016	.006
Pain/discomfort level 4	0.358	0.018	<.001	0.355	0.012	<.001	0.370	0.013	<.001	0.355	0.017	<.001	0.371	0.013	<.001	0.353	0.017	<.001
Pain/discomfort level 5	0.469	0.019	<.001	0.448	0.013	<.001	0.477	0.014	<.001	0.495	0.020	<.001	0.479	0.014	<.001	0.502	0.019	<.001
Anxiety/depression level 2	0.039	0.019	.038	0.038	0.013	.004	0.035	0.014	.015	0.044	0.010	<.001	0.041	0.013	.002	0.052	0.009	<.001
Anxiety/depression level 3	0.121	0.021	<.001	0.128	0.015	<.001	0.122	0.016	<.001	0.114	0.017	<.001	0.126	0.015	<.001	0.121	0.016	<.001
Anxiety/depression level 4	0.320	0.019	<.001	0.304	0.013	<.001	0.309	0.014	<.001	0.330	0.016	<.001	0.313	0.014	<.001	0.334	0.016	<.001
Anxiety/depression level 5	0.386	0.019	<.001	0.370	0.013	<.001	0.388	0.014	<.001	0.408	0.016	<.001	0.391	0.014	<.001	0.411	0.016	<.001
Constant	0.029	0.019	.120	0.020	0.016	.218	0.019	0.018	.290	0.016	0.010	.118						
Observations	10 680			10 680			10 680			10 680			10 680			10 680		
Number of participants	1200			1200			1200			1200			1200			1200		
Log likelihood	−9062			−5533			−6831			−7706			−6832			−7707		
Logically consistent	Yes			Yes			Yes			Yes			Yes			No (SC2, SC3, PD2, PD3)		
Number of significant coefficients at 1% level	14			20			17			20			19			20		
Proportion of significant coefficients at 1% level	66.7			95.2			81.0			95.2			95.0			100.0		
Number of significant coefficients at 5% level	18			20			20			20			20			20		
Proportion of significant coefficients at 5% level	85.7			95.2			95.2			95.2			100.0			100.0		
AIC	18 168			11 113			13 709			15 497			13 708			15 497		
BIC	18 328			11 280			13 876			15 802			13 868			15 795		
MAE at the health state level	0.041			0.038			0.044			0.047			0.043			0.047		
Number of health states with MAE>0.1	7			2			9			13			5			13		
Number of health states with MAE>0.05	33			35			41			32			39			32		
MAE at individual level	0.369			0.366			0.368			0.370			0.366			0.370		
Number of observations with MAE>0.1	7935			7963			7929			7974			7848			8001		
Number of observations with MAE>0.05	9444			9436			9359			9466			9063			9403		
Dimension with largest weight at level 5	PD			PD			PD			PD			PD			PD		
Dimension with second largest weight at level 5	AD			AD			AD			AD			AD			AD		
Dimension with third largest weight at level 5	MO			MO			MO			MO			MO			MO		
Dimension with fourth largest weight at level 5	UA			UA			UA			UA			UA			UA		
Dimension with smallest weight at level 5	SC			SC			SC			SC			SC			SC		
Value of state 22222	0.797			0.781			0.790			0.765			0.784			0.734		
Value of state 33333	0.594			0.578			0.596			0.625			0.604			0.626		

continued on next page

Table 3. Continued

Variables	Model 1 Tobit			Model 2 GLS			Model 3 Random effects Tobit			Model 4 Heteroscedastic Tobit			Model 5 Random effects Tobit no constant (UK value set)			Model 6 Heteroscedastic Tobit no constant		
	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals	Coeffs	SE	P-Vals
Value of state 44444	−0.221			−0.189			−0.217			−0.233			−0.209			−0.227		
Value of state 55555	−0.558			−0.482			−0.577			−0.610			−0.567			−0.609		
Value of 21111	0.947			0.947			0.953			0.948			0.968			0.955		
Value of 12111	0.949			0.947			0.948			0.935			0.962			0.940		
Value of 11211	0.925			0.937			0.936			0.934			0.951			0.941		
Value of 11121	0.928			0.928			0.931			0.944			0.944			0.950		
Value of 11112	0.932			0.942			0.946			0.940			0.959			0.948		
Number of health states with utility<0[a]	466			393			474			520			461					
Proportion of health states with utility<0[a]	14.91			12.58			15.17			16.64			14.75					
MAE at the health state level using uncensored mean	0.042			0.046			0.043			0.042			0.045			0.042		
Number of health states with MAE>0.1 using uncensored mean	5			12			5			10			5			10		
Number of health states with MAE>0.05 using uncensored mean	39			40			43			28			42			33		
MAE at health state level, leaving one state out of estimation[a]	0.050			0.046			0.050			0.054			0.050					
MAE at health state level, leaving one block out of estimation[a]	0.053			0.045			0.054			0.060			0.054					

Notes: Model 1 = Tobit model with censoring at −1; Model 2 = Generalized least squares random intercept model; Model 3 = Random effects Tobit model with censoring at −1; Model 4 = Heteroscedastic Tobit model with censoring at −1; Model 5 = Random effects Tobit model with censoring at −1 and no constant; Model 6 = Heteroscedastic Tobit model with censoring at −1 and no constant. Mobility level 2 equals 1 if mobility is at level 2, 0 otherwise; Mobility level 3 equals 1 if mobility is at level 3, 0 otherwise; Mobility level 4 equals 1 if mobility is at level 4, 0 otherwise; Mobility level 5 equals 1 if mobility is at level 5, 0 otherwise; with equivalent definitions for the dummy variables for the other dimensions of self-care, usual activities, pain/discomfort, and anxiety depression.
AD indicates anxiety/depression; AIC, Akaike information criterion; Coeffs, coefficients; BIC, Bayesian information criterion; MAE, mean absolute error; MO, mobility; PD, pain/discomfort; SC, self-care; SE, standard errors; P-Vals, *P* values; UA, usual activities.
[a]These have not been reported for model 6 as this model is not logically consistent.

Discussion

This study reported the development of a UK EQ-5D-5L value set based on the preferences of the UK public that is suitable for use to inform policy. The study achieved good-quality preference data using the TTO elicitation technique from an overall representative sample of the UK public for age, gender, ethnicity, and socioeconomic group. The preference data were modeled appropriately and transparently to produce utility values for all EQ-5D-5L health states.

The choice to offer both in-person and videoconference interviews to enable greater inclusivity was pioneering at its time and has since been used by other studies (eg, Al Sayah et al).[27] Substantial effort was made to use a range of approaches to advertise the study to recruit participants, with no recruitment via an existing panel of participants. Our large postal campaign was not as successful as social media at attracting people to register an interest in participating in the study (approximately 12% vs 70% of people registering to participate) (Appendix Table A1). As expected, the characteristics of participants being interviewed in-person or online vary, and while offering both is more challenging logistically, it is advantageous for obtaining a sample that is representative in characteristics and reflecting diverse views. While our sampled quotas were achieved, because of wide age bands, the final sample has a lower proportion of people aged 75 years and above than the UK population. The sample has mixed health, including some participants in poor health, to reflect a public sample rather than a "healthy" sample. Participants with experience of ill health can have a better understanding of the impact of ill health, though their preferences may reflect adaptation to ill health (see Versteegh et al[38] for a discussion of these issues). The choice for proportional representation across the 4 nations means the results are informative of UK preferences and all 4 nations. The sample sizes for each nation are not large enough to enable preference comparisons or generate value sets for each nation.

This study has benefited from study governance and input from a diverse group[31] including public involvement, independent assessment of data quality at 4 time points by a Quality Control team, and final endorsement of data and analyses by the Quality Control team and Steering Group. Comprehensive interviewer training, retraining, and pilot interviews prior to the main study, alongside careful monitoring of data quality throughout data collection, and use of the feedback module, have ensured good-quality data. Participant understanding is high, and the proportions of logically inconsistent preferences are overall low (Appendix Table A10).

There is a balance between the purpose of a model for generating a value set, what is accepted in the relevant literature, what is a fully appropriate model given the data, and what combination of characteristics of the data any model can realistically deal with. We aimed for an overall modeling approach that is interpretable and defensible, recognizing the need for transparency, reproducibility, and simplicity for value set users. The Tobit model can be criticized on the grounds that some of its assumptions are violated, which can lead to biased and unstable results. One issue raised with the Tobit is fragility to the censoring point of −1, and in wider economics literature, censored least

Table 4. UK EQ-5D-5L value set.

Dimension and severity levels	Utility decrement	Standard error
Mobility level 1	0	
Mobility level 2	0.032	0.012
Mobility level 3	0.058	0.014
Mobility level 4	0.179	0.015
Mobility level 5	0.279	0.014
Self-care level 1	0	
Self-care level 2	0.038	0.012
Self-care level 3	0.060	0.015
Self-care level 4	0.162	0.015
Self-care level 5	0.206	0.013
Usual activities level 1	0	
Usual activities level 2	0.049	0.013
Usual activities level 3	0.086	0.014
Usual activities level 4	0.184	0.013
Usual activities level 5	0.212	0.013
Pain/discomfort level 1	0	
Pain/discomfort level 2	0.056	0.012
Pain/discomfort level 3	0.066	0.015
Pain/discomfort level 4	0.371	0.013
Pain/discomfort level 5	0.479	0.014
Anxiety/depression level 1	0	
Anxiety/depression level 2	0.041	0.013
Anxiety/depression level 3	0.126	0.015
Anxiety/depression level 4	0.313	0.014
Anxiety/depression level 5	0.391	0.014

Notes: The utility for a health state is generated by subtracting from 1 the utility decrements for each severity level of each dimension. A health state is referred to using a 5-digit number made up of the severity levels of each of the five EQ-5D-5L dimensions. State 12345 would consist of mobility level 1, self-care level 2, usual activities level 3, pain/discomfort level 4 and anxiety/depression level 5. For example, the utility for health state 21111 is 1-0.032-0-0-0-0 = 0.968. The utility for health state 32432 is 1-0.058-0.038-0.184-0.066-0.041 = 0.613. The utility for health state 55555 is 1-0.279-0.206-0.212-0.479-0.391 = −0.567.

absolute deviations is a common alternative. Alternative models were explored, including censored least absolute deviations, symmetrically censored least squares, and Bayesian models. The study team and a majority of the Steering Group concluded that none of these models were unambiguously better than the models reported in this article. Although alternative methods offer considerable promise, there is insufficient evidence available to confidently recommend these models for generating the UK EQ-5D-5L value set, and further research exploring alternative models would be worthwhile to inform future value set studies.

The composite TTO approach used here was determined following a body of research undertaken to inform the selection of a suitable TTO protocol for the international valuation of EQ-5D-5L.[9,33] The use of this protocol has meant that participants could not express a TTO value below −1, though some participants may have regarded some of the health states as being below −1. The composite TTO approach uses a 10-year time frame for valuing states considered to be better than dead, and a 20-year time frame for states considered to be worse than dead. For younger participants, these time frames are lower than their own anticipated life expectancy, and this may impact their responses.[39,40] Furthermore, there is some evidence that responses to TTO tasks may be impacted by time preferences (see, for example, Attema et al[41]), but there is no accepted method for taking this into account.[42]

This study uses a state-of-the-art TTO interface, a consistent method for eliciting and scoring states better than and worse than dead, with data modeled accounting for well-established TTO data characteristics, all of which provide methodological advances to the UK EQ-5D-3L value set. Our results may suggest that preferences have changed since the EQ-5D-3L study, given that the dimensions of anxiety/depression and usual activities have a greater relative impact on the utility values than the 30-year-old EQ-5D-3L UK value set.[20] It must also be noted, however, that sample composition and study methods also differ across the 2 studies. Unlike the UK EQ-5D-3L value set, the EQ-5D-5L value set does not include interaction terms (ie, the "N3" term included in UK EQ-5D-3L) or a constant, and the distribution of utilities for any population would not be expected to follow the same distribution. The utility for the worst state is comparable (−0.567 EQ-5D-5L vs −0.594 EQ-5D-3L), though a lower proportion of health states have a utility value below zero (14.8% vs 34.6%), and the utility for the least severe impaired state differs (0.968 EQ-5D-5L vs 0.883 EQ-5D-3L).

Conclusion

The study satisfied the intention of generating good-quality data modeled appropriately and transparently from a representative sample of the UK public, to be able to generate a UK EQ-5D-5L value set suitable for informing policy. The UK value set can be used for a range of purposes, including generating utilities to calculate QALYs for use in HTA.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section. Dr Rowen is an editor for *Value in Health* and had no role in the peer-review process of this article.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2026.03.008.

Article and Author Information

Accepted for Publication: March 6, 2026

Published Online: April 28, 2026

doi: https://doi.org/10.1016/j.jval.2026.03.008

Author Affiliations: Sheffield Centre for Health and Related Research, University of Sheffield, UK (Rowen, Mukuria, Carlton); Centre for Health Economics and Medicines Evaluation, Bangor University, Wales, UK (Bray); Arrow Health Economics, London, UK (Longworth); Academic Unit of Health Economics, University of Leeds, Leeds, England, UK (Meads); Putnam Associates, Inizio Ignite, Putnam, UK (Oluboyede); Centre for Public Health, Queen's University, Belfast, Northern Ireland (O'Neill); Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, England, UK (Yang).

Correspondence: Donna Rowen, PhD, Sheffield Centre for Health and Related Research (SCHARR), University of Sheffield, Regent Court, 30 Regent Street, Sheffield, South Yorkshire, S1 4DA, United Kingdom. Email: d.rowen@sheffield.ac.uk

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: This work was supported by the EuroQol Research Foundation.

Role of the Funder/Sponsor: The funder contributed to defining certain study parameters (eg, inclusion criteria and initial sample size) and participated in the Steering Group. The academic study team retained final authority over the study design, conduct, data analysis, interpretation, and the decision to submit the manuscript for publication.

Acknowledgment: The authors thank the project Steering Group (Jan van Busschbach, Sophie Cooper, Bilal Evans, Rosie Lovett, Charles Manski, Danny Palnoch, Mark Sculpher, Kristina Secnik Boye, Bernhard Slaap (Chair), Anne Spencer), project Quality Control team (Michał Jakubczyk, John Mullahy, Mark Oppe (Chair), Juan Manuel Ramos-Goñi), the EuroQol EQ-VT support team (for this study this involved Aureliano Finch, Fanni Rencz, Bram Roudijk, Elly Stolk), EuroQol Valuation Working Group (Michał Jakubczyk (Chair), Marcel Jonker, David Mott, Richard Norman, Mark Oppe, Eleanor Pullenayegum, Juan Manuel Ramos-Goñi, Bram Roudijk) and representatives from the NICE Public and Patient Involvement Program for their contributions and feedback into the design and conduct of this project. The authors also thank the interviewers for their hard work and dedication (Amala Anyika, Marek Atter, Anisa Hajizadeh, Genevieve Hopkins, Jiyeon Kang, Will King, Mosco Michalopoulou, Jessica McCreery, Emily McDool, Aileen Neilson, Aliah Sagar, Aziza Sallam, Alaw Thomas-Davies); Juan Manuel Ramos-Goñi and Bram Roudijk for sharing analysis code; and Nancy Devlin, Brendan Mulhern, and Koonal Shah for their advice both on the protocol and during the project.

REFERENCES

1. Richardson J, McKie J, Bariola E. Multiattribute utility instruments and their use. In: Culyer AJ, ed. *Encyclopedia of Health Economics*. San Diego: Elsevier; 2014:341–357.
2. Devlin N, Brooks R. Past, present, future. *Appl Health Econ Health Policy*. 2017;15(2):127–137.
3. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–1736.
4. EuroQol BR. the current state of play. *Health Policy*. 1996;37(1):53–72.
5. Devlin N, Brazier J, Pickard AS, Stolk E. 3L, 5L, what the L? A NICE conundrum. *Pharmacoeconomics*. 2018;36(6):637–640.
6. Buchholz I, Janssen MF, Kohlmann T, Feng YS. A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *Pharmacoeconomics*. 2018;36(6):645–661.
7. National Institute for Health and Care Excellence (NICE). *NICE technology appraisal and highly specialised technologies guidance: the manual*; 2022. http://www.nice.org.uk/process/pmg36. Accessed October 15, 2025.
8. Devlin N, Shah K, 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.
9. 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):445–453.
10. National Institute for Health and Care Excellence (NICE). Use of the EQ-5D-5L value set for England: NICE position statement. https://www.nice.org.uk/position-statements/position-statement-on-use-of-the-eq-5d-5l-value-set-for-england-updated-october-2019. Accessed September 29, 2025.
11. Hernández M, Pudney S, Wailoo A. Quality review of a proposed EQ-5D-5L value set for England. Policy Research Unit in Economic Evaluation of Health and Care Interventions (EEPRU) Report. https://eepru.sites.sheffield.ac.uk/eepru2-projects/validation-of-the-eq-5d-5l-valuation-set; 2018. Accessed October 15, 2025.
12. Hernández M, Pudney S, Wailoo A. The EQ-5D-5L value set for England: findings of a quality assurance program. *Value Health*. 2020;23(5):642–648.
13. National Institute for Health and Care Excellence (NICE). Use of the EQ-5D-5L value set for England: NICE position statement. https://www.nice.org.uk/position-statements/position-statement-on-use-of-the-eq-5d-5l-value-set-for-england-updated-october-2019; 2019. Accessed September 29, 2025.
14. Brouwer W, Attema A, Versteegh M. The EQ-5D-5L valuation set for England: a view from the side. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/position-statements/position-statement-on-use-of-the-eq-5d-5l-value-set-for-england-updated-october-2019; 2019. Accessed September 29, 2025.
15. Fiebig DG. Expert evaluation of EQ-5D-5L. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/position-statements/position-statement-on-use-of-the-eq-5d-5l-value-set-for-england-updated-october-2019; 2019. Accessed September 29, 2025.
16. Manski CL. NICE Expert advice on the EQ-5D-5L valuation set for England. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/position-statements/position-statement-on-use-of-the-eq-5d-5l-value-set-for-england-updated-october-2019; 2019. Accessed September 29, 2025.
17. 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):649–655.
18. National Institute for Health and Care Excellence (NICE) Decision Support Unit. *Mapping between EQ-5D-5L and EQ-5D-3L*. University of Sheffield; 2017. https://www.sheffield.ac.uk/nice-dsu/methods-development/mapping-eq-5d-5l-3l. Accessed September 29, 2025.
19. Hernández Alava M, Pudney S, Wailoo A. Estimating the relationship between EQ-5D-5L and EQ-5D-3L: results from a UK population study. *Pharmacoeconomics*. 2023;41(2):199–207.
20. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
21. Mind. Attitudes to Mental Illness 2023 Summary Report. https://www.mind.org.uk/media/j0hpgbr0/attitudes-to-mental-illness-2023.pdf. Accessed December 3, 2025.
22. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *Eur J Health Econ*. 2010;11(2):215–225.
23. Rowen D, Mukuria C, Bray N, et al. UK valuation of EQ-5D-5L, a generic measure of health-related quality of life: a study protocol. *Value Health*. 2023;26(11):1625–1635.
24. Rowen D, Mukuria C, McDool E. A systematic review of the methodologies and modeling approaches used to generate international EQ-5D-5L value sets. *Pharmacoeconomics*. 2022;40(9):863–882.
25. Jyani G, Sharma A, Prinja S, et al. Development of an EQ-5D value set for India using an extended design (DEVINE) study: the Indian 5-level version EQ-5D value set. *Value Health*. 2022;25(7):1218–1226.
26. Azizi A, Boutib A, Achak D, et al. Valuing health-related quality of life: an value set for Morocco. *Qual Life Res*. 2025;34(6):1783–1796.
27. Al Sayah F, Roudijk B, El Sadig M, et al. A value set for EQ-5D-5L in the United Arab Emirates. *Value Health*. 2025;28(4):611–621.
28. Bailey H, Jonker MF, Pullenayegum E, Rencz F, Roudijk B. The EQ-5D-5L valuation study for Trinidad and Tobago. *Health Qual Life Outcomes*. 2024;22(1):51.
29. Rowen D, Mukuria C, Bray N, et al. Assessing the comparative feasibility, acceptability, and equivalence of videoconference interviews and face-to-face interviews using the time trade-off technique. *Soc Sci Med*. 2022;309:115–227.
30. Peasgood T, Bourke M, Devlin N, Rowen D, Yang Y, Dalziel K. Randomised comparison of online interviews versus face-to-face interviews to value health states. *Soc Sci Med*. 2023;323:115818.
31. Cooper S, Spencer A, Boye K et al. UK EQ-5D-5L value set: insights into process and governance. Unpublished manuscript.
32. Xie F, Pickard AS, Krabbe PFM, et al. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). *Pharmacoeconomics*. 2015;33(8):867–877.
33. 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.
34. Brazier J, Peasgood T, Mukuria C, et al. The EQ-HWB: overview of the development of a measure of health and wellbeing and key results. *Value Health*. 2022;25(4):482–491.
35. 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.
36. StataCorp. *Stata Statistical Software: Release 18*. College Station, TX: StataCorp LLC; 2023.
37. Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. *Am Stat*. 2016;70(2):129–133.
38. 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.
39. van Nooten FE, Koolman X, Brouwer WBF. The influence of subjective life expectancy on health state valuations using a 10 year TTO. *Health Econ*. 2009;18(5):549–558.
40. Lipman SA, Brouwer WBF, Attema AE. Living up to expectations: experimental tests of subjective life expectancy as reference point in time trade-off and standard gamble. *J Health Econ*. 2020;71:102318.
41. Attema AE, Brouwer WB. On the (not so) constant proportional trade-off in TTO. *Qual Life Res*. 2010;19(4):489–497.
42. Attema AE, Edelaar-Peeters Y, Versteegh MM, Stolk EA. Time trade-off: one methodology, different methods. *Eur J Health Econ*. 2013;14(Suppl 1):53–64.