
Article:
Rowen, D. orcid.org/0000-0003-3018-5109, Mukuria, C., Bray, N. et al. (6 more authors) (2023) UK valuation of EQ-5D-5L, a generic measure of health-related quality of life: a study protocol. Value in Health, 26 (11). pp. 1625-1635. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2023.08.005

ScienceDirect

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

Methodology

UK Valuation of EQ-5D-5L, a Generic Measure of Health-Related Quality of Life: A Study Protocol

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

ABSTRACT

Objectives: A high-quality and widely accepted UK EQ-5D-5L value set is urgently required to enable the latest version of EQ-5D scored using recent UK public preferences to inform policy including health technology assessments submitted to the National Institute for Health and Care Excellence. This article outlines the study protocol for the generation of a new EQ-5D-5L UK value set.

Methods: Twelve hundred interviews will be undertaken using the composite time trade-off elicitation technique for 102 health states (86 from the international EQ-5D-5L valuation protocol, plus 16 with best predictive performance in an extended design used in the Native American EQ-5D-5L valuation). The sample will be UK adults (age ≥18 years) proportionately representative across England, Wales, Scotland, and Northern Ireland, representative for age, sex, ethnicity, and socioeconomic group, with inclusion of participants with/without health problems. Participants will choose to be interviewed via videoconference (by Zoom) or in-person in a central venue. Data quality will be rigorously assessed.

Results: The value set will be generated using tobit random effects and heteroscedastic tobit models (with censoring at −1) using all data, excluding time trade-off values highlighted by participants as ones they would reconsider and data from interviewers failing protocol compliance. Quality and acceptance will be achieved by public involvement, regular Steering Group meetings, independent assessment of data quality at 4 time points, and final endorsement of data and analyses.

Conclusion: This study will produce a UK value set for the EQ-5D-5L for use in prospective and retrospective data sets containing EQ-5D-5L data.

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

VALUE HEALTH. 2023; 26(11):1625–1635

Introduction

EQ-5D-5L is a widely used generic measure of health-related quality of life[1,2] developed to improve on a previous version—the EQ-5D-3L.[3] EQ-5D-5L is usually self-completed and consists of 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).[4] EQ-5D-5L can be scored using preferences that generate utility values, using a value set published in the scientific literature. The value set generates utility values for every health state described by EQ-5D-5L, where values lie between 1 and −1 such that 1 equals full health, 0 is equivalent to being dead, and values below 0 indicate that a health state is worse than being dead. Utility values are combined with length of life to generate quality-adjusted life-years that are used in health technology assessment (HTA).

Previous and current National Institute for Health and Care Excellence (NICE) guides for undertaking HTA in England recommend that EQ-5D is used to generate quality-adjusted life-years, scored using a UK-specific value set from a representative sample of the public.[5,6] The Scottish Medicines Consortium and the All Wales Medicines Strategy Group also stipulate EQ-5D data for HTA. Therefore, EQ-5D-3L and EQ-5D-5L are widely used in the UK, and due to greater sensitivity of EQ-5D-5L than EQ-5D-3L in patient populations (eg, Devlin et al[7]), it is expected the use of EQ-5D-5L will increase. Without a UK value set for EQ-5D-5L, an interim approach to generating utility values was developed that uses the current UK EQ-5D-3L value set[8] and mapping/cross-walk algorithms.[9-11]

A new international valuation protocol, EuroQol Valuation Technology (EQ-VT) version 1, was developed to standardize the methods for EQ-5D-5L valuation studies and inform country-specific value sets to generate utility values.[12] This was applied in a previous valuation of EQ-5D-5L in England[13] that elicited preferences from a representative sample of 996 members of the

public using time trade-off (TTO) and discrete choice experiment (DCE) tasks. The data were modeled using a hybrid model that jointly analyzed the TTO and DCE data to generate a value set. Nevertheless, the data quality and modeling were criticized in independent quality assessments.[14-19] The English value set authors responded to, and in some cases refuted, these criticisms in a rebuttal paper.[20] Table 1[13-19] summarizes the published critiques of the English EQ-5D-5L value set without judgment on their validity nor summary of the response of the English value set authors. Table 1[13-19] summarizes the concerns raised around health state selection and coverage, sample size and representativeness, quality assurance and peer review, and participant understanding and engagement, including that the TTO data distribution had a large proportion of responses at easy-to-obtain values (1, 0.5, 0, −0.5, −1), with inconsistencies in values (worse health states have higher values), which casted doubt on the precision and accuracy of the values. The recommended model supplemented the TTO data with DCE data, and concerns were raised around this modeling (see Table 1[13-19] for a full summary). Therefore NICE released a position statement[25] that EQ-5D-5L data should be scored by mapping onto the UK EQ-5D-3L value set.[10,11] Based on learnings from the earliest EQ-5D-5L valuation studies, including the one in England, a new version of EQ-VT was developed, with more quality control over the interviews.[26] The UK EQ-5D-3L value set is old—data were collected in the early 1990s.[8] Therefore, it is unlikely the UK EQ-5D-3L value set reflects current societal preferences (eg, regarding mental health where attitudes have changed over time). In addition, UK population demographics have changed over time, and methods for eliciting health state preferences have advanced including both data collection and econometric methods for analyzing data. There are also concerns about the high proportion of EQ-5D-3L health states valued worse than dead[13,27-29] that is not replicated in other studies. Finally, the use of a measure with its own value set is preferable to mapping.[30] Therefore, a high-quality and widely accepted UK EQ-5D-5L value set is urgently required to enable the latest version of EQ-5D scored using recent UK public preferences to inform policy including HTAs submitted to NICE.

This article outlines the study protocol for the generation of a new UK value set for the EQ-5D-5L using general public preferences.

Methods and Analysis

The protocol has been informed by the international EQ-5D-5L valuation protocol,[12] which has been successfully used in 27 studies worldwide,[22,24] but deviates from this for health state selection and choice of elicitation technique in response to concerns raised by experts on behalf of NICE.[14-19] Furthermore, the international protocol does not make recommendations around sampling and recruitment of study participants, modeling of data, or selection of the value set. The protocol and statistical analysis plan has been developed through:

- Consideration of recommendations from the independent quality assessments of the previous EQ-5D-5L value set for England
- Systematic review of all published EQ-5D-5L value sets[22]
- Pilot study assessing acceptability, feasibility, and equivalence of in-person and videoconference TTO interviews[23],
- Input from public involvement sessions
- Regular input and final approval from a study Steering Group, independent Quality Control team, and EuroQol Executive Committee

A summary of the criticisms[14-19] of the English value set[13] in relation to the TTO data is presented in Table 1,[13-19] with a response showing how these have been addressed in the current protocol. The protocol is summarized in Table 2.

Ethical approval was granted from the Research Ethics Committee administered by the Sheffield Centre for Health and Related Research at the University of Sheffield.

Study Governance

Study governance is reported in Figure 1. The study benefits from regular Steering Group meetings that consists of representatives from NICE, National Health Service England, Department of Health and Social Care (DHSC), international experts, and EuroQol scientific leaders. EuroQol EQ-VT support team will provide support throughout data collection to ensure data quality is being achieved with weekly monitoring sessions. The Quality Control team will be independently assessing the data at set time points and the data analyses (further details below).

Valuation Method and System

There were numerous criticisms of the England EQ-5D-5L valuation on the DCE design, data, and modeling,[14-19] in particular the quality and interpretability of a hybrid model combining TTO and DCE data, which informed a decision by the study's Steering Group to base a UK value set on TTO data alone. Therefore, additional DCE data are not required and so will not be collected.

All interviews will be conducted using version 2.1 of the EQ-VT interviewing system,[26] an updated version to that used in the English valuation study. Version 2 made several improvements to ensure greater participant understanding and better interviewer performance and to reduce the proportion of inconsistent utility values, clustering of values (at 1, 0.5, 0, −0.5, and −1), and interviewer effects.

The TTO variant, often referred to as composite TTO, combines a procedure for states better than dead with no lead time where states have a maximum duration of 10 years, with lead-time TTO for states worse than dead with a 10-year lead time in addition to the maximum duration of 10 years, making 20 years in total (see Fig. 2[12]).

Sample Size

The sample will consist of 1200 members of the UK general public. This is larger than the 1000 participants typically used for EQ-5D-5L valuation studies,[22] following recommendations from the independent quality assessment of the previous English EQ-5D-5L value set[14,15] and subsequent advice from NICE, the Scottish Medicines Consortium, and the All Wales Medicines Strategy Group.

Sample Representativeness

The sample will be nationally representative for age, sex, ethnicity, and socioeconomic group. We will use a multistage stratified quota approach. This means we will have quota groups for age (18-40 years, 41-64 years, 65 years and older) and sex (male, female) combined, and across the entire sample, there will be quotas for ethnicity (White, non-White) and socioeconomic group (captured by the index of multiple deprivation using postcode [deciles: 1 or 2, 3-8, 9 or 10]) but there will not be quotas within each age and sex quota group. The sample will include participants with and without health problems, from urban and rural areas. The sample will be proportionately representative across England (84%), Wales (5%), Scotland (8%), and Northern Ireland (3%), with multiple geographical locations used for each nation.

Table 1. Checklist responding to concerns of previous English EQ-5D-5L valuation study raised in quality assessments.[14-19]

Summary of concern in English EQ-5D-5L value set	Protocol for new UK study with rationale	Rationale where changes have not been made in response to previous criticism
Selection of health states and health state coverage The Devlin et al[13] study used 86 health states, in line with the international protocol. Summary of criticisms: • very small number relative to the total number of possible health states (86/3125 = 3% coverage of all possible health states) • not representative of health states normally identified in cost-effectiveness studies • was not designed to take into account interaction effects between dimensions in the model, ie, is based on an additive model (although this is only important if interaction effects are also deemed important to include) • was not designed to enable assessment of inconsistent responses (although some assessments are possible) • not equally distributed coverage and gaps in coverage according to the misery index (sum of the levels of each dimension)	This study will use the standard 86 states from the international protocol plus 16 additional states selected statistically. The rationale for this: • For health state valuation studies, it is crucial that the design is able to estimate the specified model, and it is this that is important rather than coverage per se or inclusion of states that commonly occur. The design does precisely this, but only for an additive model. • Selecting health states based on prevalence rather than on experimental design theory does not improve estimations for prevalent states[21] and results in biased value sets.[13] • It would not necessarily be expected that a statistical design would have equal coverage across the misery index or coverage at every possible misery index value, given that, eg, only 5 health states have a misery score of 24, whereas 381 health states have a misery score of 15. The 86 states have been successfully used in many countries and are here supplemented with 16 additional health states to improve the predictive performance of the regression models, while maximizing comparability across countries.	• The current design was not selected for the purposes of estimating a model with interaction effects among different severity levels of different dimensions given that a main effects model has been deemed sufficient for the purpose of generating a value set. There is no evidence that the authors are aware of that suggests that interactions among different severity levels are warranted using existing elicited preference data, and allowing for all possible interactions to be estimated in a model would require an unmanageable number of health states to be valued. • The design has small health state coverage relative to the potential number of health states but is necessitated to enable an achievable sample size with sufficient observations per health state. • The study was not purposefully designed to be able to examine inconsistent responses, but it is possible to assess all of the inconsistencies identified in the previous quality assessments, as well as additional potential indicators of data quality (see Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.08.005).
Data quality The Devlin et al study[13] used the first version of the international protocol for the valuation of EQ-5D-5L: EuroQol Valuation Technology (EQ-VT) v1.0, with 48 interviewers. The quality of the data quality was criticized on the following grounds: • Participants did not understand or engage with the task. • Participants found the tasks difficult. • There was a large proportion of "problematic" responses, although the proportion differs widely depending on the definition of "problematic." • The "problematic" responses correlate negatively with participants reporting they found the tasks difficult. • There was a lack of inbuilt detection in EQ-VTv1.0 of problematic responses and methods of how to deal with these responses. • There were interviewer effects in the data and low protocol compliance by interviewers. • There were spikes in observed TTO responses at −1, 0.5, 0, 0.5, and 1. • There was low use of the worse than dead task.	This study will use an updated version of the valuation protocol, EQ-VTv2.1, which has been revised in particular to address concerns raised with EQ-VTv1.0 around data quality. EQ-VTv2 includes inbuilt weekly checks on data quality and highlights interviewer nonprotocol compliance so that this can be quickly addressed and resolved. If an interviewer has 40% or more of their interviews identified as not complying with the protocol ("red flagged"), their data are dropped and they are retrained. Central training will be provided to interviewers. Interviewers will conduct interviews in more than 1 geographical location wherever possible and will each conduct approximately 100 interviews, with approximately 13 interviewers. If there are matters arising during the course of the interviews, we will feed this back to all interviewers and guidance will be provided to all interviewers on how to handle the matter to ensure consistency. All participants are shown both versions of the TTO task, where the state is regarded as better than dead and where the state is regarded as worse than dead, during the warm-up tasks. Participants are also shown the implied ranking of the health states given their TTO values and asked to indicate those they would now reconsider (the "feedback module"). The TTO values for those health states are then flagged for exclusion in data analyses. This study has an independent quality control (QC) team who will assess the quality of the raw data, at several points throughout the study, and final data and modeling analyses.	• There may still be an issue that self-reported difficulty does not indicate a lack of engagement or understanding. Additional analyses of understanding will assess inconsistent responses, proportions of the data at certain easy-to-achieve TTO values, and interviewer-reported understanding and engagement and self-reported difficulty. • We are not excluding responses on the basis of inconsistencies, but will assess the extent of this in the data. We will explore the inconsistencies and the size of utility differences where these occur, given that small inconsistencies and large inconsistencies should not be equally regarded. • We will fully explore the data including peaks at 1, 0.5, 0, −0.5, and −1 and will also assess this during data collection.
Sample size The international protocol for valuation of EQ-5D-5L specifies a sample size of 1000 participants, and this was used in the Devlin et al[13] study. The sample size was criticized on the following grounds: • no formal justification of sample size • lack of analysis of sampling error and specification robustness • small in comparison with health and social surveys	The new UK value set will sample 1200 participants, an increase of 200 on the previous study. This is a 20% increase on the rate that has been successfully used in a large number of EQ-5D-5L valuation studies conducted worldwide.	• The higher sample size has not been calculated using formal analysis of sampling error and specification robustness but is based upon experience of previous EQ-5D-5L value sets across 27 countries worldwide.[22,24] • The sample will be nationally representative, meaning that there are small samples for Scotland, Wales, and Northern Ireland, but the proportion of participants across the 4 nations is nationally representative.

continued on next page

Table 1. Continued

Summary of concern in English EQ-5D-5L value set	Protocol for new UK study with rationale	Rationale where changes have not been made in response to previous criticism
Sample and nonresponders In the Devlin et al study,[13] participants were sampled on the basis of postcode, and it was not specified whether they were sampled for representativeness based on other characteristics. The sample did not appear representative for age, sex, ethnicity, education, and marital status. The response rate was 47.7% of the 2220 addresses that were sampled. The sample and response rate was criticized on the following grounds: • Response rate is low, eg, in comparison with Health Survey for England. • No information was maintained on non-responders, but bias from nonresponse is due to the pattern of nonresponders rather than the proportion of nonresponders. • background characteristics data missing for 12 participants • numerical base unclear when generating sample percentages • Sample weighting used in the modeling to improve sample representativeness is unsuccessful. • Experience/nonexperience of illness may be an important characteristic on which to assess preferences.	This study is not aiming for geographical representation, other than for the 4 nations where the sample will be proportionally representative. We will use a multistage stratified quota approach, with quota groups for age and sex in each nation, and across the whole sample, there will be quotas for ethnicity and socioeconomic group (thus not within each quota group for age and sex) using index of multiple deprivation (IMD) using postcode. In addition, sampling will ensure the inclusion of participants with and without health problems and participants from urban and rural areas. A blended recruitment approach will be used to ensure a mix of people are contacted to be invited to participate in the study, including a postal mailout, social media, and snowballing. We will: • obtain responses for all sociodemographic questions wherever possible • ensure that in summaries of sample characteristics the numerical base on which each sample percentage is calculated is clear • obtain a representative sample, and hence, we do not propose the use of sample weights in our analyses • use multiple waves of recruitment to ensure the sample is representative	Due to the proposed recruitment approach, we will not be able to calculate response rates, but this approach has been selected to make the study more accessible and allow a wider range of participants to be recruited, ultimately maximizing the sample's representativeness of and generalizability to the UK population.
Modeling of TTO data The Devlin et al[13] study selected a hybrid model that jointly modeled the DCE and TTO data. Many criticisms of the modeling that was undertaken are specific to this hybrid model. Criticisms that are generic or that relate to the modeling of the TTO data in particular are: • TTO values were incorrectly assumed to be censored at 1. • Modeling of heteroscedasticity was inappropriate or potentially not the best alternative. • no justification for accounting for preference heterogeneity • lack of interaction effects at worse levels of health • lack of interaction effects between the dimensions (especially at the worst levels) • did not assess model performance using a measure of absolute fit • Alternative models and factors affecting their results or interpretation of results were not reported. • Treatment of multiple observations per participant in the modeling was insufficient. • lack of consideration of the assumptions made in the modeling, including normality of distributions, linearity of distributional means in covariates, and homoscedasticity, assumption that the TTO responses are accurate • (the converse) A common value set may not be appropriate since preferences may differ for different people.	This study is collecting only TTO data, which make the modeling much simpler and more transparent, and avoids the problems and criticisms related to combining DCE and TTO data using a hybrid model. We reviewed the methods used in other published EQ-5D-5L studies estimating models using only the TTO data to inform the statistical analysis plan.[22] We will: • not interpret data at 1 as being censored • include models that are appropriate for the data including models that account for multiple observations per participant (eg, heteroscedastic tobit model with censoring at −1, random effects tobit model with censoring at −1) • not use a model accounting for preference heterogeneity as the recommended value set • estimate models including interaction effects that are equivalent to the N3 term in reflecting a combined effect on utility from having more than 1 dimension at a severe or extreme level • include a measure of absolute fit to assess model performance (Akaike information criteria [AIC] and Schwarz information criteria [BIC]) • ensure all analyses that are referred to are reported and interpreted (eg, assessing interviewer effects) • ensure correct use of predictions for all models • produce a fully consistent model for the value set	• We are using an additive model that does not allow for interactions between levels of the different dimensions given that a main effects model has been deemed sufficient for the purpose of generating a value set (see above in section on selection of health states). Note we are including interaction effects to reflect a combined effect on utility from having more than one dimension at a severe or extreme level. • The sample size has not been selected to detect differences across different characteristics given that this is not the aim when generating a value set.

continued on next page

Table 1. Continued

Summary of concern in English EQ-5D-5L value set	Protocol for new UK study with rationale	Rationale where changes have not been made in response to previous criticism
Quality assurance and peer review The Devlin et al[13] study was publicly funded. The data and coding were made available for quality assurance purposes, and placed in a EuroQol data archive following journal publication, in line with the original study proposal, and can be made available to individual members of the public by request. However, they were not publicly available prior to journal publication. Data quality was monitored throughout the data collection process, and concerns were presented and discussed in working papers, conferences and seminars. However, no independent quality control took place during the study conduct. The study was criticized on the following grounds: • Data, coding, and analyses, including any sensitivity analyses, should be made available without restriction to the research community. • Data quality should be independently and transparently assessed. • Peer review via journals is insufficient given that reviewers cannot access data and coding and are unlikely to review to the required level of detail. • The selected journal should have been a statistical or econometric journal. • Future research should assess country differences in data generation and modeling.	The new study has a QC team who will independently assess and model the raw data. They will make recommendations to the Steering Group after interim analyses (pilot data, ~20% of the data, ~40% of the data, ~60% of the data) with Steering Group approval required for data collection to continue, and assess the raw final data and modeling analyses. The data and analyses, including any sensitivity analyses, will be made available to those who may be separately appointed by DHSC or NICE to assess the study. It is recommended that this is undertaken before journal publication. Once the expected articles have been accepted for publication, the data, coding (annotated with explanations), and analyses will be made available, on request via the EuroQol Research Foundation.	• The target journal has not been determined, but it is doubtful that the value set article would be best suited to a statistical or econometric journal. Nevertheless, given the scrutiny this study will face, it can be recommended that reviewers include econometricians and/or statisticians. • If the data are available from the EuroQol Research Foundation, rather than a body such as UK Data Archive, it is less publicly accessible. • Nevertheless, this project is funded by the EuroQol Research Foundation whereas the English study was publicly funded.
Classification system The Devlin et al[13] study used the EQ-5D-5L classification. Concerns were raised around the ambiguity in wording for levels 4 (severe problems) and 5 (extreme problems/unable to do).	The new study will use the EQ-5D-5L classification system. There may not be differences in the impact on TTO values for levels 4 and 5 of each dimension. This will be explored in the data analyses, and the value set will be a fully consistent model.	The new study will be using the same classification system, but this is a necessity of valuing a standardized international measure.

DHSC indicates Department of Health and Social Care; NICE, National Institute of Health and Care Excellence; TTO, time trade-off.

Selection of Health States

The international EQ-5D-5L valuation protocol includes 86 states for valuation using TTO[12] in 10 blocks of 10 states used in the TTO tasks, where each block comprises 8 states selected statistically, 1 mild state and the worst state. Given the larger sample size and in response to quality assessment criticisms of the England EQ-5D-5L value set (see Table 1[13-19]), these 86 states and 10 blocks will be supplemented with 16 additional states in 2 new blocks. This decision was based on evidence showing that increasing the number of observations per state was not beneficial in terms of predictive performance of the regression model, but adding more blocks/states was beneficial.[24] This strategy enables approximately 100 observations per block in line with the international protocol and expands the number of states valued while including already well-performing and widely used states from the international protocol.

Additional states were selected of 64 additional states included in the EQ-5D-5L valuation study in India[24] by the study authors using lowest mean squared error across models estimated for the standard 86 states plus 16 states, with 100 000 repeats and replicated in an external data set. Additional 16 states have been allocated to blocks using the blocking algorithm used in the international protocol.[12] Two mild states, added 1 each to these blocks, were selected from the mild states in the international protocol with level 2 problems with the fewest occurrences.

Participant Recruitment

Participants will be recruited using a blended approach to ensure a mix of people are contacted to be invited to participate in the study, including postal mailouts targeting postcodes, Facebook posts, Twitter tweets, paid Facebook adverts, local websites, and word of mouth/snowballing from participants. Difficult-to-reach quotas may require additional recruitment methods, including posters, flyers, and adverts in local settings. We will use existing panels of willing participants if required to obtain hard-to-reach participants to ensure a representative sample. Participants interested in being interviewed will complete a short screening survey to obtain contact details and sociodemographic characteristics. Survey responses will be used to sample willing participants and arrange the interview.

Mode of Administration

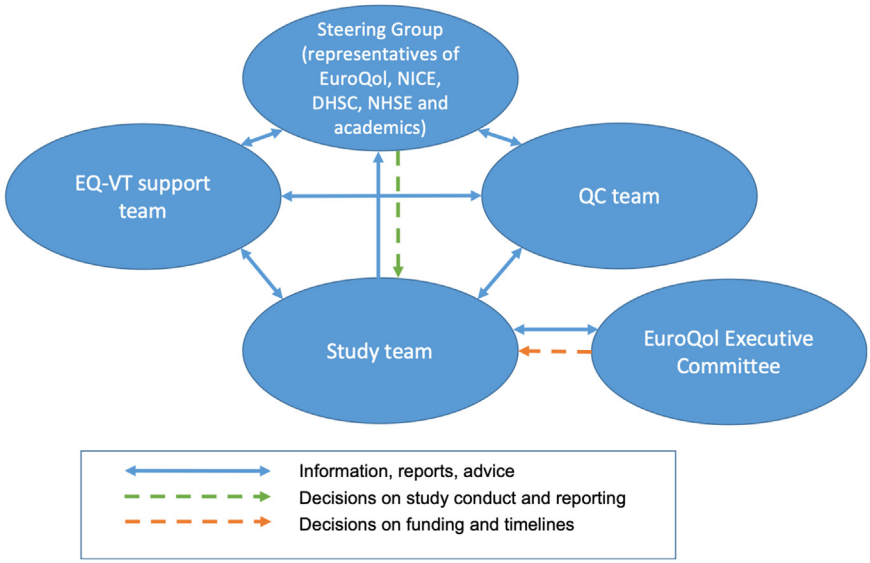
To directly inform this study protocol, we assessed the acceptability, feasibility, and equivalence of in-person and video-conference TTO interviews.[22] We found mode did not affect TTO values per se, and although video interviews have lower data quality for some aspects, the quality of video interviews was still good. Video TTO interviews have also been found to be feasible in other studies.[31-34] Nevertheless, different participants are likely to be willing to participate if only 1 mode of interview is offered; the sociodemographic characteristics of people who preferred video interviews differed significantly from those who preferred face-to-face interviews,[22] including characteristics that are protected under the Equalities Act 2010 such as age, sex, and ethnicity. Therefore, in this study participants will be offered a choice to be interviewed via video or in-person, making the study more inclusive and improving the sample's representativeness. These qualities are important for evidence considered by policy makers such as NICE, in particular national value sets.[6,35] There are no quotas on the number of interviews per mode. In-person interviews will take into account applicable guidelines due to COVID-19. Participants are required to switch on their camera to enable the interviewer to monitor understanding and

Table 2. Overview of the UK EQ-5D-5L valuation protocol.

Issue	Protocol
Elicitation technique and protocol	Composite time trade-off (cTTO) elicitation technique administered using the EQ-VTv2.1 computer-assisted personal interview (CAPI) system in English, with the option of Welsh where requested.
Health state selection	102 health states, 86 of these in line with the international EQ-5D-5L valuation protocol, with 16 additional states added that were identified as having best predictive performance in the extended design used in the Indian EQ-5D-5L valuation. A health state is a 5-digit identifier of the levels of the 5 dimensions in the order of mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. For example, the worst health state is state 55 555, and state 21 111 has mobility level 2 and all other dimensions at level 1.
Sample size	1200 members of the UK general public, with representative proportions across England (84%), Wales (5%), Scotland (8%), and Northern Ireland (3%)
Sampling and recruitment of participants	Multistage stratified quota approach, with quota groups for age and sex in each nation, and across the whole sample, there will be quotas for ethnicity and socioeconomic group (thus not within each quota group for age and sex). The definition of socioeconomic group will be the index of multiple deprivation (IMD) using postcode. In addition, sampling will ensure the inclusion of participants with and without health problems and participants from urban and rural areas. The sample will be proportionately representative across England (84%, n = 1010), Wales (5%, n = 57), Scotland (8%, n = 100), and Northern Ireland (3%, n = 33), with at least 2 geographical locations used for each nation, given that HTA agencies for the 4 nations did not express a preference for overrepresentation of the smaller nations. A blended recruitment approach will be used to ensure a mix of people are contacted to be invited to participate in the study, including a postal mailout, social media, and snowballing.
Mode of administration	Videoconference interview (by Zoom) and in-person interview. Interested participants will be able to indicate their preferred mode and wherever feasible the preferred mode will be offered (note that for some geographical locations, eg, rural areas, the nearest in-person interview location may mean that interested participants expressing a preference for an in-person interview later decline to be interviewed in-person).
Location of face-to-face interviews	Central location, using rooms that can be hired. The locations need to be as accessible as possible. We will consider home interviews if it is determined that the use of central locations excludes some participants.
COVID-19	If social distancing is at a level where face-to-face research is allowed, in-person interviews will be conducted in a central location, with appropriate safety precautions that will be in accordance with national and local guidelines, although may be stricter than these guidelines.
Interviewers	Each interviewer will conduct interviews in more than one area. Each interviewer will conduct approximately 100 interviews. Interviewers will be recruited from available PhD students and researchers where possible.
Financial incentive to participants	£50 offered via bank transfer or Amazon vouchers.
Training of interviewers	Interviews will receive in-person EQ-VT training that builds upon training materials provided by the EuroQol EQ-VT support team. Interviewer training will be led by the study team with some sessions led by the EQ-VT support team.
Pilot	Each interviewer will undertake a minimum of 10 interviews as a pilot. A sample of these interviews from each interviewer will be observed by the study team. All interviewers will receive feedback. These pilot interviews will not be retained in the data set used to generate the value set. We will follow the recommendations of the EQ support team and quality control (QC) team and follow the EQ-VT procedure to determine whether the findings of the pilot are favorable to enable us to continue with the remaining data collection and seek approval of the Steering Group before starting the main study. Any interviewer not meeting protocol compliance will continue with another 10 pilot interviews that will be closely monitored. Once full data collection has started, in line with the EQ-VTv2.1 system, if interviewers are flagged around issues of data quality/protocol noncompliance for 40% or more of their interviews, their data are dropped and they are retrained. In this instance, the interviewer will be monitored for the next 10 interviews and the process repeated. If the interviewer fails again, they will be removed from the study. Monitoring of protocol compliance and the quality of interviews will continue during the whole study.
Public involvement	We are using the NICE Public Involvement Programme and have public involvement throughout the project. Public involvement sessions have been undertaken at 3 time points during protocol development and the pilot study, and public involvement sessions will be held once data has been collected and analyzed.
Ethical approval	Obtained via the Research Ethics Committee administered by the Sheffield Centre for Health and Related Research at the University of Sheffield
QC	The project will adhere to the QC approach in the EQ-VT protocol and will include regular team debriefs involving interviewers. Independent QC has been built in at every stage, with the QC team undertaking a review of: study protocol including statistical analysis plan, interim analyses (pilot data, ~20% of the data, ~40% of the data, ~60% of the data) with a recommendation made to the Steering Group at each stage and Steering Group approval required for data collection to continue, the final data set, and the final modeled value set.
Statistical analysis plan	A statistical analysis plan has been approved by the Steering Group and the QC team before the project start. The modeled value set will be nationally representative.

EQ-VT indicates EuroQol Valuation Technology; HTA, health technology assessment; NICE, National Institute of Health and Care Excellence.

Figure 1. Study governance.



DHSC indicates Department of Health and Social Care; EQ-VT, EuroQol Valuation Technology; NICE, National Institute of Health and Care Excellence; NHSE, National Health Service in England; QC, quality control.

engagement. Each interviewer will conduct interviews with participants from more than 1 geographical area, although may be restricted to a single geographical area for in-person interviews, and will conduct approximately 100 interviews.

The Interview

Informed consent will be taken before the start of the interview. At the start of the interview, the participant will complete brief sociodemographic questions and EQ-5D-5L, familiarizing themselves with the severity levels of each dimension. In the first practice TTO task, the interviewer will explain in detail the TTO technique, using the state "mobility problems which mean you need to use a wheelchair" followed by a state that is either better than or worse than this, selected to ensure TTO procedures for states better than dead and states worse than dead are explained. This is followed by another 3 TTO practice tasks using EQ-5D-5L health states (mild, severe, and difficult to imagine states). After this, TTO data collection will start and participants complete 10 TTO tasks. Participants are then shown their implied rankings for the 10 health states, derived from their TTO responses. Any states with the same value appear side by side, and states where the value was worse than dead (ie, below 0) are shaded in a different color (called the "feedback module"). Participants are asked to indicate any states whose values they would now reconsider, although no further TTO data are collected for any states that are indicated. Participants are asked questions about their understanding, what they thought of the interview, and further sociodemographic and health characteristics, including EQ Health and Wellbeing Short (EQ-HWB-S).[36] Interviewers report their interpretation of participant understanding and effort and concentration after the interview has been completed. Participants will receive £50 as a thank you for their participation in the study, offered as a choice of cash (via bank transfer) or Amazon vouchers. If a participant is unable to complete the interview due to cognitive ability or distress, the interview is stopped and is not included in the analysis (although the participant receives their full incentive for participating).

Interviewer Training

Interviewers will be trained in person by the study team and receive training from a member of the EQ-VT support team. Each interviewer will undertake a minimum of 10 pilot interviews, some of which will be observed by the study team for training purposes. Data quality and interviewer protocol compliance will be assessed (see below), and feedback provided to each interviewer. Pilot data will not be included in the final data set.

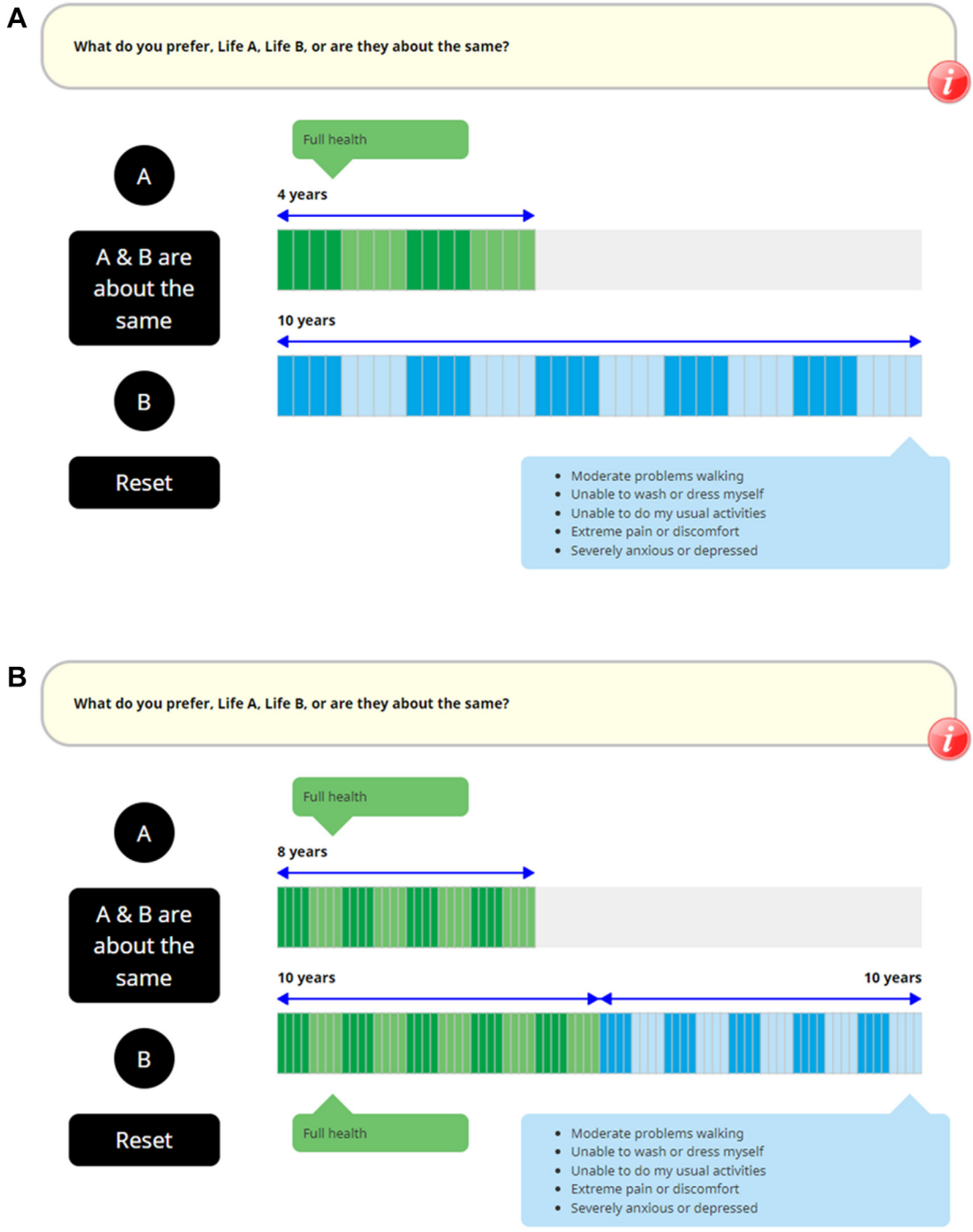
Data Quality and Data Quality Control

Preferences differ across individuals and cultures, meaning assessments of data quality should focus on objective measures. One avoidable cause of poor data quality is poor interviewer performance; therefore, data quality will be closely monitored for each interviewer with retraining and dialogue between the interviewer and study team to resolve any data quality issues. Version 2 of the EQ-VT protocol includes a weekly quality control report[37] during data collection that focuses upon data quality and protocol compliance at the aggregate level and for each interviewer. An interview is red flagged to indicate protocol noncompliance (as a proxy for poor data quality) where there is one or more of the following[37]:

1) no explanation given of the task for states considered worse than dead in the first practice question
2) less than the predefined minimum time (3 minutes) spent on the initial practice question
3) inconsistent TTO ratings where the most severe state (55555) is valued at least 0.5 higher than the lowest valued state
4) less than the predefined minimum time (5 minutes) spent on the 10 TTO tasks

If an interviewer has 40% or more of their interviews identified as poor quality (red flagged) in weekly assessments (as recommended in the international EQ-5D-5L valuation protocol[37]), their data will be dropped and the interviewer will be retrained. If the interviewer fails again, they will be removed from the study.

Figure 2. Example screenshots of EQ-VTv2.1 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.



EQ-VT indicates EuroQol Valuation Technology; TTO, time trade-off.

This weekly report does not assess data quality at the individual participant level or indicate potential misunderstanding or lack of engagement for participants. Nevertheless, there is no consensus in the literature around markers nor threshold levels of TTO data quality at the individual participant level, making it difficult to judge what constitutes permissible preferences or poor data quality. Data assessments at the individual participant level to better understand the TTO data and to indicate data quality will be assessments of logical inconsistencies, proportions of the data at certain easy-to-achieve TTO values (−1, −0.5, 0, 0.5, 1), and understanding and engagement (see Table 3 and Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.08.005).

The study research team, EuroQol EQ-VT support team, and the Quality Control team will assess the data after the pilot, after ~20%, 40% and 60% of the data has been collected, and the final

Table 3. Individual participant level assessments of data and data quality.

What is assessed	Interpretation of results and what this indicates
Logical inconsistencies	The definition of logical consistency, when used to assess data at the individual level, means that a logical inconsistency occurs when individuals provide a higher utility value to a health state that is worse (eg, 31 111 is better than 21 111). Although logical inconsistencies can indicate a lack of understanding and engagement, some inconsistencies are expected given that participants value health states in a random order, and there are both learning effects at the start of TTO tasks and fatigue effects for final TTO tasks. It is important to distinguish between ties in TTO values between health states, indicating the participant does not distinguish between the 2 in terms of the number of life-years they are willing to sacrifice to avoid each health state, and an inconsistency, where the worse state is given a higher TTO value. It is also important to understand whether these inconsistencies are for similar health states, eg, 44 554 and 44 344, or for very different health states, eg, 21 111 and 55 555, and furthermore the size of the differences in the TTO values for the states, eg, 0.05 or 1.5.
Proportions of the data at certain easy-to-achieve TTO values (−1, −0.5, 0, 0.5, 1)	This can indicate misunderstanding or lack of engagement given that these TTO values require fewer moves in the TTO and are a way of completing the task quickly or can show that the individual is unwilling to think about precision in their values for the case of −1, which is the lowest value that can be provided.
Understanding and engagement	This is indicated by participant self-reported understanding and views on the TTO tasks and interviewer-reported perceptions of the participant's understanding and engagement. This can indicate that the participants' values may not accurately reflect their preferences.

TTO indicates time trade-off.

data. The Steering Group will receive a Quality Control report as outlined earlier for all interviews to date, recommendations from the EQ-VT support team, recommendations from the independent quality control team, and a study team report including the data assessment described in Table 3 and Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.08.005. The Steering Group approval will be required for data collection to continue. If at any point during the study, there is an issue arising around data quality the study team and the Steering Group will act quickly to ensure that the issue is resolved.

Data collection will continue until 1200 completed interviews excluding interviews from interviewers flagged at 40% for non-protocol compliance.

Public Involvement

We are using the NICE Public Involvement Programme and held sessions to obtain public input into the design of both the main valuation study and the pilot study assessing the acceptability of combining the data collected by different modes, including sample, financial incentives to participants, recruitment strategies particularly for participants with lower education and from lower socioeconomic groups, interview scheduling, questions asked about sociodemographic status and participant understanding, information provided to participants about the interview and COVID-19 protocols, and information used to inform the choice of mode of administration for the main study.

Public involvement around study results and recommendations will be undertaken before selection of the recommended value set by the study team.

Data Analysis and Modeling

Exclusion criteria, modeling, assessments of model performance, robustness analyses, and selection of the final value set have been informed by a systematic review of all published EQ-5D-5L value sets,[22] consideration of theoretically appropriate analyses for the data, and the purpose of generating a value set.

All data will be retained for modeling purposes unless they meet the following exclusion criteria: TTO values from interviews with interviewers red flagged for protocol noncompliance for 40% or more of their interviews and TTO values identified by the participant during the interview (via the feedback module) as values they would now reconsider (remaining TTO values for the participant are retained).

The model specification will be an additive model with utility decrements for each severity level and TTO value as the dependent variable. This will involve 20 parameter level dummies, which are 20 parameters with dummies for levels 2 to 5 of each of the 5 dimensions, leaving level 1 as the reference category. The coefficient for each severity level of each dimension represents the utility decrement from having no problems to having problems at that severity level for the given dimension. To generate the value set for all health states from this model, the utility value for each health state is calculated as 1 plus the sum of the utility decrements for each dimension.

If logical inconsistencies are observed, for example, between level 4 and 5 coefficients in any dimensions (ie, utility increases as health worsens), we will combine logically inconsistent adjacent dummies and re-estimate the model, as also undertaken for the US and The Netherlands value sets.[38,39] We will combine adjacent dummies only if their coefficients are logically inconsistent, not due to a lack of a statistically significant difference between them.

There may be a combined effect on utility from having more than one dimension at a severe or extreme level, equivalent in interpretation to the "N3" term used in the UK EQ-5D-3L value set.[8] We will explore the inclusion of interaction terms to reflect this, which will only be retained if they are statistically significant, robust across different models and subsets of the data, interpretable, and sensible (in terms of their sign and relative size) and improve model performance.

TTO data are characterized by censoring at −1, given that participants cannot express a lower TTO value than −1 for any health state although they may wish to do so; repeated observations per participant given that each participant values 10 TTO health states; larger variance for more severe health states; and

different preferences across participants (namely preference heterogeneity). To account for these characteristics, we will estimate the heteroscedastic tobit model with censoring at −1, random effects tobit model with censoring at −1, tobit model with censoring at –1, and a generalized least squares random intercept model. We will explore the heteroscedasticity of the error term and note approaches used in the US value set.[38]

Model performance will be assessed, in order of priority, using (1) logical consistency of coefficients (including sign and relative size of coefficients within a dimension), (2) number and proportion of significant coefficients (using a model estimated with incremental dummies) at the 0.05 and 0.01 levels, (3) Akaike information criterion and Schwarz information criterion, and (4) mean absolute error. In the case of disagreement across the different assessments over which is the better performing model for the value set, the ordering of the assessments will be applied.

Robustness analyses will assess the impact on the model coefficients and their statistical significance for the selected value set model in comparison with models estimated on a subset or expansion of the TTO data using (1) TTO data with and without TTO data identified in the feedback module (ie, values participants would reconsider), (2) TTO data with and without participants reported by the interviewer as not understanding or engaging with the TTO tasks, and (3) TTO data for the 86 states (from international protocol) and full design of 102 states. Concerns are raised about the robustness of the results if the sign, significance, logical consistency, and relative importance of dimensions differ across analyses. If there is a lack of robustness, reasons will be explored to understand the cause. If there are data quality concerns, consultation with the EuroQol EQ-VT support team, Quality Control team, and Steering Group will be sought around an acceptable solution for the generation of the value set.

The analyses and recommended value set will be independently checked by the Quality Control team, who will access the raw data and syntax files used to analyze the data.

Selection of the UK Value Set for the EQ-5D-5L

The primary output of the project will be a new UK value set for the EQ-5D-5L, available for use on all prospective and retrospective EQ-5D-5L data. The value set will meet the following criteria:

- The sample for model estimation will be nationally representative and proportionally representative across England, Wales, Scotland, and Northern Ireland and inclusive of all participants and responses meeting the inclusion criteria.
- The modeled value set will be logically consistent (utility does not increase as health worsens).
- The modeling analysis will be transparent and involve only widely accepted regression analyses for TTO data.

The value set will be the best performing model overall and will be subject to scrutiny and feedback from the Steering Group, Quality Control team, public involvement group, and the EuroQol Executive Committee. Study results will be submitted to NICE, DHSC, and National Health Service England where it is anticipated quality assurance will be undertaken to determine whether the value set will be recommended for use by NICE and DHSC. Selection of the value set does not imply immediate implementation by NICE or DHSC to inform policy.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2023.08.005.

Article and Author Information

Accepted for Publication: August 9, 2023

doi: https://doi.org/10.1016/j.jval.2023.08.005

Author Affiliations: Sheffield Centre for Health and Related Research, University of Sheffield, Sheffield, England, UK (Rowen, Mukuria, Carlton); Academy for Health Equity, Prevention and Wellbeing, Bangor University, Bangor, Wales, UK (Bray); National Institute for Health and Care Excellence, London, England, UK (Cooper); PHMR Consulting, London, England, UK (Longworth); Academic Unit of Health Economics, University of Leeds, Leeds, England, UK (Meads); Centre for Public Health, Queen's University, Belfast, Northern Ireland (O'Neill); National University of Ireland, Galway, 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, University of Sheffield, Regent Court, 30 Regent St, Sheffield, S1 4DA, England, United Kingdom. Email: d.rowen@sheffield.ac.uk

Author Contributions: *Concept and design:* Rowen, Mukuria, Bray, Carlton, Cooper, Longworth, Meads, O'Neill, Yang
Acquisition of data: Longworth, Meads, Yang
Analysis and interpretation of data: Longworth, Yang
Drafting of the article: Rowen, Mukuria, Bray, Carlton, Cooper, O'Neill
Critical revision of article for important intellectual content: Rowen, Mukuria, Bray, Carlton, Longworth, Meads, Yang
Obtaining funding: Rowen, Mukuria, Bray, Carlton, Longworth, Meads, O'Neill, Yang
Administrative, technical, or logistic support: Yang;
Supervision: Cooper, Yang;
Other: Patient and Public Involvement guidance: Carlton

Conflict of Interest Disclosures: Dr Rowen reported receiving grants from EuroQol Research Foundation, during the conduct of the study; other from SF-6Dv2; and grants from EuroQol Research Foundation, outside the submitted work, and is a member of the EuroQol Group. Dr Mukuria reported receiving grants from EuroQol Research Foundation, during the conduct of the study, and grants from EuroQol Research Foundation, outside the submitted work, and is a member of the EuroQol Group. Drs Bray and Carlton reported receiving grants from EuroQol Research Foundation, during the conduct of the study. Dr Cooper reported receiving other from NICE during the conduct of the study. Dr Longworth reported receiving grants from EuroQol Foundation, during the conduct of the study; grants from EuroQol Foundation, outside the submitted work; and is a member of the EuroQol Group, who developed the EQ-5D instruments and funded the research. Drs Meads and O'Neill reported receiving grants from EuroQol Research Foundation, during the conduct of the study. Dr Yang reported receiving grants from EuroQol Research Foundation during the conduct of the study, personal fees from EuroQol Research Foundation, and other from EuroQol Research Foundation outside the submitted work and is a member of the EuroQol member and a member of the Valuation Working Group member of the EuroQol group. No other disclosures were reported.

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

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the article; and decision to submit the article for publication.

Acknowledgment: The authors thank the project Steering Group (Jan Busschbach, Sophie Cooper, Bilal Evans, Rosie Lovett, Charles Manski, Danny Palnoch, Mark Sculpher, Kristina Secnik Boye, Bernhard Slaap [Chair], Anne Spencer, Elly Stolk), project Quality Control team (Michal Jakubszyk, John Mullahy, Mark Oppe [Chair], Juan Manuel Ramos-Goñi), and public involvement representatives from the NICE Public and Patient Involvement Programme for their ongoing contributions and feedback into the design and conduct of this 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, CA: Elsevier; 2014:341–357.

2. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: past, present, future. *Appl Health Econ Health Policy*. 2017;15(2):127–137.
3. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37(1):53–72.
4. 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.
5. Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/process/pmg9/resources/guide-to-the-methods-of-technology-appraisal-2013-pdf-2007975843781. Accessed October 22, 2022.
6. NICE health technology evaluations: the manual. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741. Accessed October 22, 2022.
7. Devlin N, Brazier J, Pickard AS, Stolk E3L. 5L, what the L? A NICE conundrum. *Pharmacoeconomics*. 2018;36(6):637–640.
8. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
9. Van Hout B, Janssen M, Feng YS, et al. Interim scoring for the EQ 5D 5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15(5):708–715.
10. Hernández Alava M, Wailoo A, Pudney S. Report by the Decision Support Unit: methods for mapping between the EQ-5D-5L and the 3L for technology appraisal. NICE DSU. https://www.sheffield.ac.uk/nice-dsu/methods-development/mapping-eq-5d-5l-3l. Accessed October 22, 2022.
11. 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.
12. 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.
13. 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.
14. 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. https://rees-france.com/wp-content/uploads/2015/11/eepru-report-eq-5d-5l-27-11-18-final-1.pdf. Accessed October 22, 2022.
15. Hernández Alava 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.
16. Abrams K. Report on EQ5D-5L UK valuation study. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l. Accessed October 22, 2022.
17. 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/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l. Accessed October 22, 2022.
18. Fiebig DG. Expert evaluation of EQ-5D-5L. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l. Accessed October 22, 2022.
19. 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/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l. Accessed October 22, 2022.
20. 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.
21. Yang Z, Luo N, Bonsel G, Busschbach J, Stolk E. Selecting health states for EQ-5D-3L valuation studies: statistical considerations matter. *Value Health*. 2018;21(4):456–461.
22. Rowen D, Mukuria C, McDool E. A systematic review of the methodologies and modelling approaches used to generate international EQ-5D-5L value sets. *PharmacoEconomics*. 2022 Sep;40(9):863–882.
23. 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. *Social Science and Medicine*. 2022 Sep;309:115227.
24. 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.
25. Position statement on use of the EQ-5D-5L value set for England (updated October 2019). National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/technology-appraisal-guidance/eq-5d-5l. Accessed October 22, 2022.
26. 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.
27. Brazier JE, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and the SF-6D across seven patient groups. *Health Econ*. 2004;13(9):873–884.
28. Hernández Alava M, Wailoo AJ, Ara R. Tails from the peak district: adjusted limited dependent variable mixture models of EQ-5D questionnaire health state utility values. *Value Health*. 2012;15(3):550–561.
29. Parkin D, Devlin N, Feng Y. What determines the shape of an EQ-5D index distribution? *Med Decis Making*. 2016;36(8):941–951.
30. 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.
31. Lipman SA. Time for tele-TTO? Lessons learned from digital interviewer-assisted time trade-off data collection. *Patient*. 2021;14(5):459–469.
32. Finch AP, Meregaglia M, Ciani O, Roudijk B, Jommi C. An EQ-5D-5L value set for Italy using videoconferencing interviews and feasibility of a new mode of administration. *Soc Sci Med*. 2022;292:114519.
33. Estévez-Carrillo A, Dewilde S, Oppe M, Ramos-Goñi JM. Exploring the comparability of face-to-face versus video conference-based composite time trade-off interviews: insights from EQ-5D-Y-3L valuation studies in Belgium and Spain. *Patient*. 2022;15(5):521–535.
34. 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.
35. Our principles. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/about/who-we-are/our-principles. Accessed September 5, 2022.
36. 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.
37. Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJ, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value Health*. 2017;20(3):466–473.
38. 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.
39. M Versteegh M, M Vermeulen K, M A A Evers S, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value Health*. 2016;19(4):343–352.