Article:

Mukuria, C. orcid.org/0000-0003-4318-1481, Peasgood, T., McDool, E. et al. (3 more authors) (2023) Valuing the EQ Health and Wellbeing Short (EQ-HWB-S) using time trade-off and a discrete choice experiment: a feasibility study. Value in Health, 26 (7). pp. 1073-1084. ISSN 1098-3015

https://doi.org/10.1016/j.jval.2023.02.008

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

ScienceDirect

Preference-Based Assessments

Valuing the EQ Health and Wellbeing Short Using Time Trade-Off and a Discrete Choice Experiment: A Feasibility Study

Clara Mukuria, PhD, Tessa Peasgood, PhD, Emily McDool, PhD, Richard Norman, PhD, Donna Rowen, PhD, John Brazier, PhD

ABSTRACT

Objectives: The EQ Health and Wellbeing Short (EQ-HWB-S) is a new generic measure that covers health and wellbeing developed for use in economic evaluation in health and social care. The aim was to test the feasibility of using composite time trade-off (cTTO) and a discrete choice experiment (DCE) based on an international protocol to derive utilities for the EQ-HWB-S and to generate a pilot value set.

Methods: A representative UK general population was recruited. Online videoconference interviews were undertaken where cTTO and DCE tasks were administered using EuroQol Portable Valuation Technology. Quality control (QC) was used to assess interviewers' performance. Data were modeled using Tobit, probit, and hybrid models. Feasibility was assessed based on the distribution of data, participants, and reports of understanding from the interviewer, QC and modeling results.

Results: cTTO and DCE data were available for 520 participants. Demographic characteristics were broadly representative of the UK general population. Interviewers met QC requirements. cTTO values ranged between −1 to 1 with increasing disutility associated with more severe states. Participants understood the tasks and the EQ-HWB-S states; and the interviewers reported high levels of understanding and engagement. The hybrid Tobit heteroscedastic model was selected for the pilot value set with values ranging from −0.384 to 1. Pain, mobility, daily activities, and sad/depressed had the largest disutilities, followed by loneliness, anxiety, exhaustion, control, and cognition in the selected model.

Conclusions: EQ-HWB-S can be valued using cTTO and DCE. Further methodological work is recommended to develop a valuation protocol specific to the EQ-HWB-S.

Keywords: discrete choice experiment, EQ Health and Wellbeing Short, EQ-HWB-S, EuroQol Valuation Technology, time trade-off.

VALUE HEALTH. 2023; 26(7):1073–1084

Introduction

Quality-adjusted life-years (QALYs) are outcome measures used in economic evaluation of healthcare technologies to support resource allocation. A QALY combines the value of health-related quality of life (HRQoL) with the length of life in a single number. HRQoL measures, which are scored using preferences, are typically used to derive the value element of a QALY in terms of utilities.[1] There are a number of existing generic HRQoL utility measures, such as EQ-5D-5L™.[2] The EQ Health and Wellbeing™ (EQ-HWB™) has been developed to capture a broad range of health and wellbeing outcomes for economic evaluation of interventions in health, public health, and social care including for informal carers.[3] There are 2 versions of the instrument, a longer profile measure with 25-items, and the short version of the measure, the EQ-HWB Short (EQ-HWB-S™), which has 9-items.[3] The measures cover items related to 7 dimensions: activity, relationships, cognition, self-identify, autonomy, feelings, and physical sensations. The measures are experimental with further testing and validation being undertaken.

In order to generate utilities, different preference-elicitation methods can be used, including time trade-off (TTO), standard gamble, and discrete choice experiments (DCEs) with or without duration, which have been successfully applied to the valuation of other measures.[1] The EQ-HWB-S was limited to ≤10 items in the development phase, because it was not considered likely that the 25-item questionnaire could be valued using these methods.[3] We conducted a small (n = 19) mixed-methods pilot and found that it was feasible and practical to apply the EuroQol Valuation Technology (EQ-VT) vs2 protocol[4] that is used to value the EQ-5D-5L to the EQ-HWB-S.[5] The protocol uses composite TTO (cTTO) and DCE and has been successfully applied to EQ-5D-5L.[6] This larger study aimed to test the feasibility of cTTO and DCE for valuing EQ-HWB-S in a larger sample and to generate a pilot value set.

Methods

Study Design

The research design and data collection adapted methods developed by the EuroQol group for valuation of the EQ-5D-5L,[4] which were tested in 2 pilot studies, a previous mixed-methods study[5] and a second pilot (n = 23) undertaken in this study. Ethical approval for the study was obtained from the University of Sheffield School of Health and Related Research Ethics Committee (038012).

Descriptive System

States were described using the EQ-HWB-S (which includes 9 items) and each item had 5 response options with 3 different response categories based on difficulty/ frequency/ severity (Table 1 and see Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008, eg). States combine items and levels ranging from having no problems (111111111) to the worst problems in all dimensions (555555555) with a total of 1 953 125 states (59). Levels in each state can be summed into the level sum score, for example, 9 for no problems.

Preference-Elicitation Methods

We used the EQ-VT vs2[4] approach for cTTO in which states better than dead are valued using a choice of living for a shorter period in full health ($n < 10$ years) versus living in the impaired state for a longer period (n = 10 years) (Appendix Fig. 1A in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008). For states worse than dead (WTD), the choice is living in full health for a shorter period ($n < 10$) versus living for 20 years—10 years in full health followed by 10 years in the impaired state. The "full health" descriptor was changed to "full health and quality of life" to reflect the breadth of the EQ-HWB-S.

In the DCE, participants choose a preferred state from a pair. Each pair had overlaps in 4 of 9 items to reduce the cognitive burden, with color highlighting the differences to help with engagement[7,8] (Appendix Fig. 1B in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008). Piloting in the previous[5] and this study led to minor modifications, including showing all possible levels for each response option on the screen for both cTTO and DCE.

State Selection

Designs were based on the use of a main effects model without interactions. For cTTO, an orthogonal array was used to select states, resulting in 50 states, 1 of which was state 111111111, which was not included. In a similar way to the selection of states for EQ-5D-5L,[9] 10 mild states were added to the design, that is, with only mild problems in 1 or 2 dimension, for example, state 111111112. States were arranged into 10 blocks with 7 states per block, 5 from the orthogonal array plus worst possible state (555555555), and 1 mild state.

For DCE, a D-efficient design was implemented using Ngene 1.2.1.[10] A candidate set (n = 5000) of randomly generated pairs with the required constraints across 4 dimensions (ie, for each pair, 4 of 9 dimensions were identical, and these dimensions and levels varied across pairs) was used to enable overlap in the design. The candidate set was checked for whether levels were equally distributed and that there were no large differences between pairs, that is, dominant pairs. The design utilized small nonzero priors for the dummy variables for each dimension to denote monotonically increasing severity in levels within each dimension.[11] This resulted in 140 pairs, with 20 blocks of 7 pairs.

Sampling and Recruitment

A convenience sample was recruited from the University of Sheffield (n = 23) to pilot the survey in April 2021. Data collection for the main study took place between May and November 2021 in England, with expansion to the rest of the United Kingdom from October to November. Participants were sampled via quota sampling using age and sex combined and ethnicity separately with a target sample size of 600, because this was a feasibility study. A recruitment agency sent targeted postal invites to individuals identified via sources in the public domain and in an online research panel. Advertising was undertaken on social media by the recruitment agency, interviewers, and other researchers, on a research advertising website, and via snowballing from previous participants. Potential participants completed an article (postal) or online screener survey. The inclusion criteria were the following: (1) $\geq$18 years of age, (2) current UK resident, (3) access to a computer, laptop, or large screen tablet with an internet connection and the ability to access videoconferencing, and (4) the ability to complete tasks in English. Participants were offered an incentive of a shopping voucher for study participation. The value of the incentive was £25, which was increased to £40 after 11 interviews to improve recruitment.

Survey Administration

The study was initially designed to be undertaken face-to-face, but video conferencing was used because of ongoing COVID-19 restrictions. Video conferencing has been used successfully in online interviews for other valuation studies,[12-14] and the data produced are similar to cTTO data collected face-to-face.[15] The second pilot was used to test online administration.

All interviews were undertaken via videoconferencing using EQ Portable Valuation Technology (EQ-PVT). EQ-PVT is the same as EQ-VT but the tasks are presented via MS PowerPoint[16] rather than using an online survey. EQ-PVT has been used previously to value EQ-5D-5L[17,18] and allows EQ-HWB-S modifications. The EQ-VT vs2 interview script for EQ-5D-5L was modified to match EQ-HWB-S and videoconference presentation.

All participants were asked to give consent and complete questions about themselves via an online survey before the interview. In the interview, participants were presented with the EQ-HWB-S, including an example of a state that could be considered implausible to familiarize them with the measure and who could complete it (patients, social care users, and informal carers).

Participants then did 4 practice cTTO tasks—a state with mobility problems (wheelchair), a state they considered worse than requiring a wheelchair, and a mild and a severe EQ-HWB-S state. The implausible state was not valued in the practice. For the few who considered the wheelchair state as being WTD, interviewers demonstrated the better than dead part of the cTTO task to ensure familiarity with both the better than and WTD tasks. Seven EQ-HWB-S states were valued in random order. Therefore, the number of practice states (4 vs 5) and actual states (7 vs 10) was less than those for EQ-5D-5L, which was aimed at minimizing the burden. This was followed by 7 paired DCE tasks presented in a random order, with randomization within the pairs to the left/right of the display followed by feedback questions. For all preference tasks, participants were asked to read each state description aloud but were also offered the option to have the interviewer read the state aloud to increase accessibility. After the cTTO and DCE tasks, participants answered questions on understanding, ease of decision making, and how easy it was to think about EQ-HWB-S states.

Finally, interviewers completed questions on their views of participants' engagement and understanding, technical problems, difficulties with the participant, and any additional comments.

Table 1. EQ-HWB-S classification system.

Item	Level	
Mobility	1	**No** difficulty getting around inside and outside
	2	**Slight** difficulty getting around inside and outside
	3	**Some** difficulty getting around inside and outside
	4	**A lot of** difficulty getting around inside and outside
	5	**Unable** to get around inside and outside
Daily activities	1	**No** difficulty doing day to day activities
	2	**Slight** difficulty doing day to day activities
	3	**Some** difficulty doing day to day activities
	4	**A lot of** difficulty doing day to day activities
	5	**Unable** to do day to day activities
Exhaustion	1	**Never** exhausted
	2	**Only occasionally** exhausted
	3	**Sometimes** exhausted
	4	**Often** exhausted
	5	Exhausted **most or all of the time**
Loneliness	1	**Never** lonely
	2	**Only occasionally** lonely
	3	**Sometimes** lonely
	4	**Often** lonely
	5	Lonely **most or all of the time**
Cognition	1	**Never** have trouble concentrating/thinking clearly
	2	**Only occasionally** have trouble concentrating/thinking clearly
	3	**Sometimes** have trouble concentrating/thinking clearly
	4	**Often** have trouble concentrating/thinking clearly
	5	Trouble concentrating/thinking clearly **most or all of the time**
Anxiety	1	**Never** anxious
	2	**Only occasionally** anxious
	3	**Sometimes** anxious
	4	**Often** anxious
	5	Anxious **most or all of the time**
Sadness/	1	**Never** sad/depressed
depression	2	**Only occasionally** sad/depressed
	3	**Sometimes** sad/depressed
	4	**Often** sad/depressed
	5	Sad/depressed **most or all of the time**
Control	1	**Never** feel you have no control over your day to day life
	2	**Only occasionally** feel you have no control over your day to day life
	3	**Sometimes** feel you have no control over your day to day life
	4	**Often** feel you have no control over your day to day life
	5	Feel you have no control over your day to day life **most or all of the time**
Physical pain	1	**No** physical pain
	2	**Slight** physical pain
	3	**Moderate** physical pain
	4	**Severe** physical pain
	5	**Very severe** physical pain

Note. © 2022 EuroQol Research Foundation. EQ-HWB™ is a trade mark of the EuroQol Research Foundation. UK (English) Not to be used without permission. Note: The classification system is not the questionnaire.
EQ-HWB-S indicates EQ Health and Wellbeing Short.

Table 2. Respondent characteristics.

Characteristic	(1) Full sample (%) n = 521*	(2) General population UK (%)
Age	48.45	
18-30	20.15	20.53
31-50	31.67	32.61
51-65	29.94	24.33
66+	18.23	22.53
Sex		
Male	45.30	49.4
Female	54.13	50.6
Other	0.58	
Ethnicity		
White	81.96	86
Black	5.95	4
Asian	9.40	8
Mixed/other	2.69	2
Employment/activity status		
Employed	59.42	
Unemployed	2.88	
Caring for family	3.08	
Looking after home	4.62	
Student	4.62	
Retired	22.50	
Long-term sick	2.69	
Other inc. volunteer	0.19	
Working status		
Usual place	47.87	
Working from home	50.49	
Furloughed	0.66	
Leave of absence	0.98	
Degree	66.35	
Social care services used	3.27	
Caring responsibilities (elderly/disabled)	14.04	
1-19 hours	61.64	
20-49 hours	19.18	
≥50 hours	13.70	
Don't know–caring hours	5.48	
Additional caring responsibilities eg, children	11.73	
Long-term health condition	31.54	
Experience of serious illness		
Serious illness – self	21.92	
Serious illness – family	52.69	
Serious illness – others	15.77	
Health satisfaction	7.01	
Life satisfaction	7.18	
EQ-VAS (n = 520)	76.29	
EQ-5D-5L index value (n = 429)†	0.80	
General health		
Excellent/very good/good	84.42	
Fair/poor	15.58	

Continued in the next column

Note. The age group percents is calculated as the percent of the adult UK population based on population projection data from the ONS.[29] Ethnicity figures are sourced from the 2011 census.
ONS indicates Office for National Statistics; UK, United Kingdom; VAS, visual analog scale.
*Sample consists of 521 individuals, but we only have the full characteristics of 520 individual—for 1 respondent we only have age sex and ethnicity.
†All participants were asked to complete the EQ-5D-5L, but these questions were not compulsory to complete.

Interviewers and Quality Control Process

Interviewers were PhD students (n = 6) recruited from the University of Sheffield or members of staff (n = 1). Interviewers received full training from the trained study team leader and EuroQol scientific staff, with adapted materials supplied by the EuroQol Group.

Interviews received a "flag" for suspected quality concerns if they did not meet 4 criteria as set out in EQ-VT vs2 protocol (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008) with a flag rate of >40% requiring retraining of the interviewer and their data up to that point would be dropped.[4] Distribution of cTTO data by interviewer was also assessed. Feedback was provided to all interviewers both collectively and individually.

Data Analysis

Descriptive analysis of the respondents and the cTTO and DCE data was undertaken. Feasibility was assessed based on how tasks were undertaken and face validity of the data. Time taken to complete a cTTO task was assessed, with longer times assumed to indicate more difficulty. In the EQ-VT vs2 protocol, it is assumed that a minimum of 5 minutes is needed to complete 10 cTTO tasks. No time is recorded for DCE tasks; therefore, this was not assessed.

The distribution of cTTO (with higher values for mild states) and minimal clustering at specific values, for example, 0 and 0.5 were used as indicators of feasibility along with few participants displaying inconsistencies in their values or not giving up time (non-traders). DCE data were assessed for evidence of particular patterns, for example, ABABABA/AAAAAAA etc., in which fewer participants using these approaches indicated feasibility. We also reviewed interviewer and participants reported understanding and engagement with the tasks and EQ-HWB-S states to assess feasibility.

Feasibility was also based on whether modeling the preference data resulted in logically ordered and statistically significant coefficients. Less weight was given to this assessment because of the smaller sample size in this feasibility study. Different approaches can be used to model cTTO and DCE[19] data, and we focus on selected models (additional analysis are reported in the Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008).

The cTTO data are censored because the WTD task only allows participants to go as low as −1 when they may be willing to go lower; therefore, a generalized Tobit[20] model was used for cTTO data. For the cTTO data, the dependent variable was the utility value from cTTO tasks, which was transformed by subtracting this value from 1 to generate disutility. Dummy variables representing levels 2 to 5 were included for each dimension, for example, mo2, mo3, mo4, and mo5 for mobility; with additional analysis based on change between levels, that is, incremental dummies. No constant term was included, because given the broader aspects of health and wellbeing, it was assumed that there is no gap between full health and quality of life and having no problems. Furthermore, the constant term in a linear prediction was small and not statistically significant.

DCE data are binary data; therefore, a generalized probit[21] model was used for DCE data with the dependent variable being the choice options. The dummy variables were the difference between the paired choice to facilitate estimation using probit and hybrid models. DCE models were estimated on a latent scale and were not anchored on the 0 to dead utility scale. To allow comparison, modeled DCE values were anchored using the cTTO value for the worst state from the Tobit heteroscedastic model.[22]

A hybrid Tobit model that combined cTTO and DCE data and accounts for censored data was also estimated (see Ramos-Goñi et al[23] for details), which has the advantage of using all of the data. The dependent and independent variables were equivalent to those used for cTTO and DCE data. Combining the data relied on an assumption that the 2 preference-elicitation methods are measuring the same thing, i.e. preferences for states described by EQ-HWB-S, with a constant proportional relationship between them, which can be modeled jointly.[23] The cTTO and DCE results were compared using Lin's concordance correlation coefficient[24] and a plot of the predictions to assess how they compared.

Across all models, only main effects were estimated. In models which there were logically disordered coefficients, adjacent severity levels were merged to constrain them to the same value. Observed variance of cTTO and DCE values is known to increase with severity of the state indicating heteroscedasticity.[19] All models took into account multiplicative heteroscedasticity[25] using a maximum likelihood estimation. Each participant completed 7 cTTO and DCE tasks; therefore, clustering of standard errors was used in all models. Models were estimated in Stata 17.0 MP.[26]

Study Sample

The pilot data (n = 23) were not included. Interviewers completed between 48 and 94 interviews with some indication that 1 interviewer (number 3) was faster than the other interviewers in completing cTTO (Appendix Fig. 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008). No interviewer reached the 40% flag rate; therefore, no cTTO data were dropped (Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008). No respondents were excluded based on potentially suspicious patterns in DCE data.

Because this was a feasibility study, sensitivity analysis was not reported, but we examined the impact of age, sex, and interviewer effects by including these as variables in the estimation. Interviewer effects were assessed by inclusion of a dummy variable for each interviewer and separately running the models while sequentially excluding 1 interviewer. We also examined whether there were differences in participants that interviewers saw based on age (analysis of variance) and sex (chi-square test), because this may explain the effects of interviewer.

Pilot Value Set

We recommended a model as the pilot value set based on logical ordering and statistical significance with the caveat that this is a feasibility study. Akaike and Bayesian information criterion and mean absolute errors could not be compared across different model types (eg, Tobit and the hybrid Tobit models). Statistical analyses were performed using Stata MP 17.[26]

Results

Recruitment and Sample

The final sample consisted of 520 participants with cTTO data and 521 with DCE data (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008for response rates). The target of 600 participants was not met due to challenges of recruiting participants, and the decision was made that the smaller sample size was sufficient for a feasibility study. Three participants who had data were not included (2 did not consent and 1 participant had 2 entries in July and November; the latter data were excluded). One participant did not want to complete cTTO and only had DCE data.

Descriptive Statistics

The mean age was 48.5 years and 45% were male with some differences compared with the general population for employment[27] and education, with 66% of 21-64-year-olds in the sample having a degree compared with 42% in the same age-group in the general population[28] (Table 2). There were 32% of respondents who reported having a longstanding (12 months) physical or mental health impairment, illness, or disability diagnosed by a doctor. A small proportion (10%) reported no problems (111111111) using the EQ-HWB-S.

Assessing Feasibility and Practicality

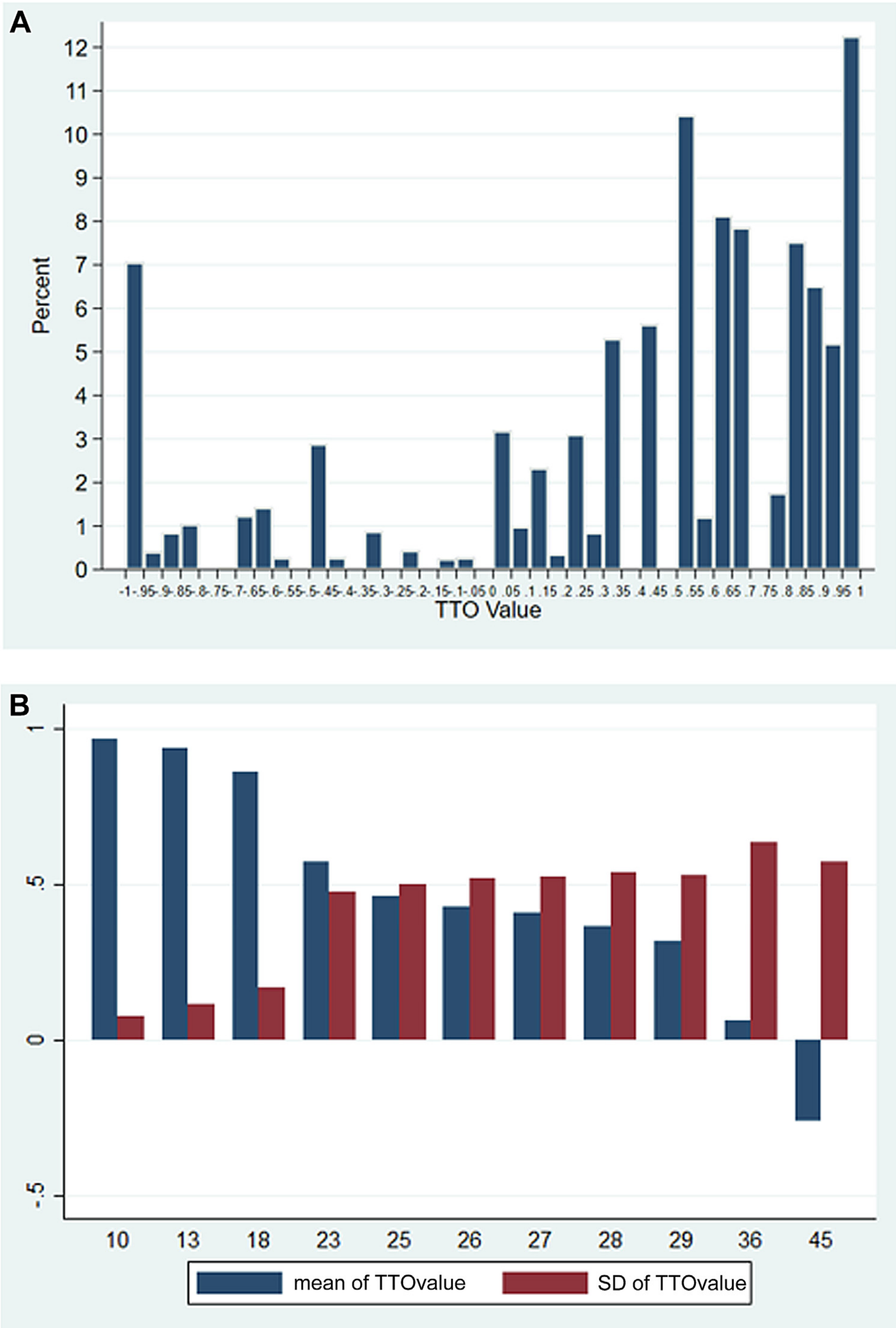
The mean (SD) time taken to complete a single cTTO task was 102.5 (SD = 69.2) seconds with a median of 86 seconds, which was longer than the minimum time estimated for EQ-5D-5L. For 10% of the cTTO tasks, participants took >3 minutes.

The distribution of cTTO values ranged from −1 to 1 (Fig. 1A) with lower mean cTTO values and larger SD as the level sum score increased (Fig. 1B and Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008). The coverage across the level sum scores for the selected TTO states was low. The observed mean cTTO values for the states ranged from 0.982 (state 111111121) to −0.264 (state 555555555). The proportion of values clustered at −1, −0.5, 0, 0.5, and 1 was 7%, 3%, 3%, 10%, and 12%, respectively, with 17.4% cTTO responses with a value below 0. All the states apart from the mild states had 1 or more negative value, but these were mixed states with majority including at least 1 dimension with a level 4 or 5 (Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008). For most of these states (74%), a negative value was given 10 or less times. The worst state that was seen by all participants was given a negative value by 278 participants (53%).

Forty participants had inconsistencies involving the worst state, and out of these, 10 participants had inconsistencies with the worst possible state valued at least 0.5 higher than another state. There were 2 nontraders. Few respondents exhibited any response pattern in the DCE indicative of poor engagement (n = 11).

Participants reported that they found it easy to understand both TTO and DCE tasks, although they found it difficult to decide in both the cases despite reporting that they were able to think about and tell the difference between the EQ-HWB-S states (Fig. 2A). Interviewers reported only a small proportion of participants (≤3%) who did not understand or engage (Fig. 2B and Appendix Table 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008).

Figure 1. (A) Distribution of TTO values. (B) Distribution of TTO values by level sum score.



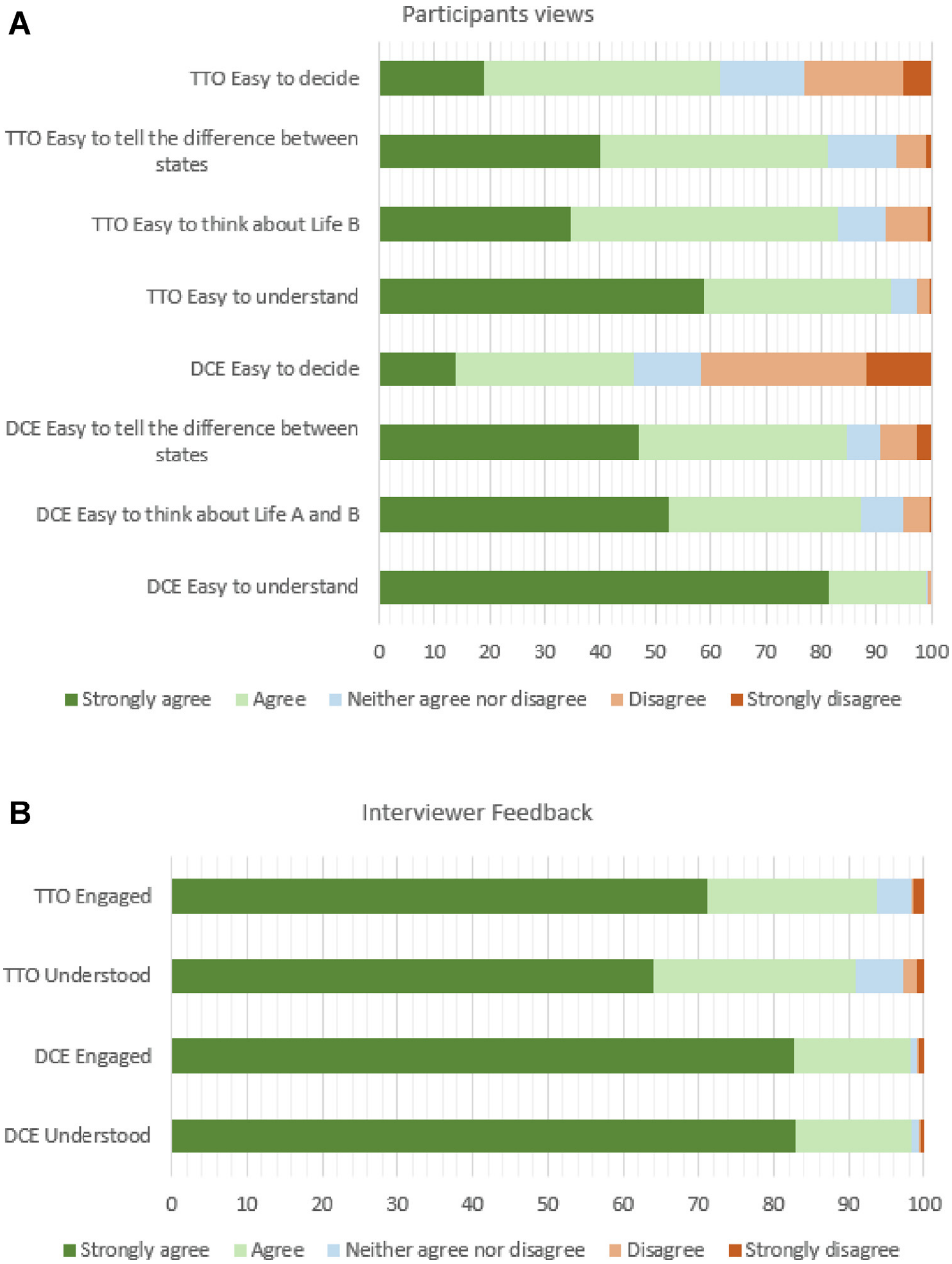
TTO indicates time trade-off.

Modeling Results

In the 3 selected models (models 1-3; see Appendix Tables 5 and 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008 for additional results), most of the dimensions demonstrated logical ordering in the response levels, with more severe levels having larger disutility (Table 3). Nevertheless, there were logical inconsistencies in the cognition (level 5) and anxiety (level 3) dimensions across models 1-3 and in exhaustion, loneliness, sadness/depression, and control (Table 3). Merging levels in models in which there was disordering led to new inconsistencies, sometimes in different dimensions, apart from in the hybrid model (model 4).

Ranking dimensions based on level 5 was similar across the 3 models for pain, mobility, daily activities, sadness/depression, and loneliness (Table 3). The remaining dimensions varied in order depending on the model. Predictions from the Tobit

Figure 2. Participant and interviewer feedback. (A) Participant views on TTO and DCE tasks and states. (B) Interviewer feedback on participant engagement and understanding.



DCE indicates discrete choice experiment; TTO, time trade-off.

heteroscedastic model were correlated with those from the probit heteroscedastic model (Lin's concordant correlation coefficient = 0.98; see Appendix Fig. 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008). Despite being correlated in terms of predictions, there were differences in most dimensions when coefficients were compared across the 3 models (Appendix Fig. 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008), but most of the differences were small (≤0.01), apart from the pain dimension.

Coefficients for levels 2 and 3 were <0.1 in all the models for all dimensions (Table 2). Some of the wellbeing and mental health-related dimensions (anxiety, control, and cognition) also had coefficients that were <0.1 for levels 4 and 5 in all models. Moving between levels was associated with incremental changes that were below 0.1 apart from in the pain dimension (Appendix Table 7 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008).

Most of the coefficients were statistically significant at the 5% level with variations by response level and dimension (Table 3), although for Tobit and Probit, not as many coefficients were statistically significant when incremental variables were used (Appendix Table 7 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008). Hybrid models had fewer coefficients that were not statistically significant compared with the other models.

Overall mean absolute error was 0.039, 0.066, and 0.056 for the Tobit, probit, and hybrid models, respectively (Table 2).

Table 3. Parameter estimates for main effects models estimated using cTTO data, DCE data, and both cTTO and DCE data.

Variables	cTTO data		DCE data			cTTO and DCE data			
	Tobit het		Probit het		Rescaled	Hybrid Tobit with control for het			
	1		2			3		4 (no disordered coefficients)	
Mobility2	0.049*	(0.012)	0.257†	(0.105)	0.042	0.050*	(0.009)	0.053*	(0.009)
Mobility3	0.100*	(0.027)	0.504*	(0.097)	0.081	0.071*	(0.009)	0.070*	(0.010)
Mobility4	0.115*	(0.028)	0.841*	(0.121)	0.136	0.132*	(0.013)	0.136*	(0.013)
Mobility5	0.195*	(0.025)	1.271*	(0.122)	0.205	0.200*	(0.012)	0.207*	(0.013)
Activity2	0.046*	(0.010)	0.092	(0.124)	0.015	0.038*	(0.008)	0.041*	(0.008)
Activity3	0.089*	(0.029)	0.314*	(0.105)	0.051	0.059*	(0.010)	0.063*	(0.010)
Activity4	0.131*	(0.022)	0.912*	(0.102)	0.147	0.141*	(0.011)	0.150*	(0.011)
Activity5	0.183*	(0.030)	1.192*	(0.108)	0.192	0.195*	(0.014)	0.199*	(0.014)
Exhaustion2	0.029*	(0.009)	0.116	(0.091)	0.019	0.023*	(0.007)	0.019†	(0.007)
Exhaustion3	0.059†	(0.024)	0.157†	(0.080)	0.025	0.029*	(0.009)	0.027*	(0.009)
Exhaustion4	0.071†	(0.030)	0.382*	(0.083)	0.062	0.071*	(0.011)	0.066*	(0.011)
Exhaustion5	0.059†	(0.028)	0.574*	(0.095)	0.093	0.084*	(0.010)	0.082*	(0.011)
Loneliness2	0.022*	(0.007)	0.098	(0.092)	0.016	0.019*	(0.006)	0.021*	(0.006)
Loneliness3	0.070†	(0.029)	0.338*	(0.097)	0.055	0.044*	(0.010)	0.052*	(0.011)
Loneliness4	0.051‡	(0.029)	0.567*	(0.102)	0.091	0.092*	(0.011)	0.101*	(0.011)
Loneliness5	0.099*	(0.026)	0.775*	(0.120)	0.125	0.116*	(0.012)	0.120*	(0.012)
Cognition2	0.020*	(0.006)	0.178‡	(0.092)	0.029	0.022*	(0.006)	0.003	(0.007)
Cognition3	0.061*	(0.024)	0.181†	(0.091)	0.029	0.039*	(0.011)	0.016	(0.011)
Cognition4	0.077*	(0.029)	0.502*	(0.094)	0.081	0.093*	(0.011)	0.057*	(0.009)
Cognition5	0.043‡	(0.023)	0.370*	(0.091)	0.060	0.077*	(0.011)	0.057*	(0.009)
Anxiety2	0.027*	(0.010)	−0.074	(0.096)	−0.012	0.013‡	(0.007)	0.022*	(0.006)
Anxiety3	0.034‡	(0.020)	0.048	(0.087)	0.008	0.0080	(0.009)	0.022*	(0.006)
Anxiety4	0.030	(0.026)	0.514*	(0.094)	0.083	0.063*	(0.011)	0.069*	(0.011)
Anxiety5	0.062†	(0.027)	0.567*	(0.109)	0.091	0.088*	(0.011)	0.092*	(0.011)
Sad/depress2	0.033*	(0.009)	0.086	(0.098)	0.014	0.031*	(0.007)	0.031*	(0.007)
Sad/depress3	0.027	(0.019)	0.118	(0.097)	0.019	0.046*	(0.010)	0.034*	(0.009)
Sad/depress4	0.052	(0.033)	0.546*	(0.105)	0.088	0.113*	(0.012)	0.113*	(0.012)
Sad/depress5	0.155*	(0.027)	0.937*	(0.112)	0.151	0.172*	(0.011)	0.173*	(0.011)
Control2	0.010‡	(0.006)	−0.130	(0.099)	−0.02	0.004	(0.006)	0.004	(0.006)
Control3	0.054‡	(0.033)	0.128	(0.103)	0.021	0.048*	(0.011)	0.045*	(0.011)
Control4	0.070†	(0.032)	0.211	(0.147)	0.034	0.065*	(0.011)	0.065*	(0.012)
Control5	0.083*	(0.024)	0.422*	(0.107)	0.068	0.080*	(0.010)	0.082*	(0.011)
Pain2	0.040*	(0.014)	0.277*	(0.094)	0.045	0.037*	(0.010)	0.038*	(0.010)
Pain3	0.082*	(0.027)	0.585*	(0.102)	0.094	0.075*	(0.011)	0.080*	(0.011)
Pain4	0.325*	(0.031)	1.479*	(0.127)	0.239	0.248*	(0.014)	0.258*	(0.014)
Pain5	0.457*	(0.035)	2.169*	(0.172)	0.350	0.357*	(0.018)	0.372*	(0.018)
Observations	3640		3647			7287		7287	
AIC	4280		2977			7248		7301	
BIC	4732		3424			7758		7784	
Disordered	EX5,CG5,LN4,AN4,SD3		CG5,AN3,CL2			CG5,AN3,CL2		-	
MAE	0.039				0.066	0.056		0.056	
MAE mild states	0.006				0.022	0.010		0.010	
Ranking	PN>MO>AC> SD>LN>CL> CG>EX>AN		PN>MO>AC> SD>LN>EX> AN>CL>CG			PN>MO>AC> SD>LN>CG> EX>AN>CL		PN>MO>AC> SD>LN>AN> EX>CL>CG	

continued on next page

Table 3. Continued

Variables	cTTO data	DCE data		cTTO and DCE data	
	Tobit het	Probit het	Rescaled	Hybrid Tobit with control for het	
	1	2		3	4 (no disordered coefficients)
Range of values	−0.335 to 0.990	−0.335 to 1.021		−0.368 to 0.996	−0.384 to 0.997
Estimated utility by selected states					
111111112	0.960		0.955	0.963	0.962
222222222	0.724		0.852	0.763	0.768
333333333	0.424		0.617	0.581	0.591
444444444	0.078		0.039	−0.018	−0.015
555555555	−0.335		−0.335	−0.368	−0.384

Note. Heteroscedastic models estimated with clustered standard errors to account for repeated observations. Ranking based on largest decrement for each dimension. Rescaled—the DCE values were anchored using the cTTO value for the worst state from the Tobit heteroscedastic model. Standard errors in parenthesis.
AC indicates activity; AIC, Akaike information criterion; AN, anxiety; BIC, Bayesian information criterion; CG, cognition; CL, control; cTTO, composite time trade-off; DCE, discrete choice experiment; EX, exhaustion; het, heteroscedasticity; LN, loneliness; MAE, mean absolute error by state; mild states, states with only 1 or 2 dimensions at level 2; MO, mobility; PN, pain; SD, sadness/depression.
* $P < .01$.
† $P < .05$.
‡ $P < .1$.

Predictions on selected states were similar for mild states (0.955 to 0.963 for state 111111112 mild pain), but there were larger differences for other states with results based on only cTTO data, resulting in lower utilities for states 222222222 and 333333333 compared with the model based on DCE data (Table 3). For the worst state, predictions ranged from −0.335 to −0.384.

Selected Model

The hybrid Tobit model, which controls for heteroscedasticity with no disordered coefficients, was selected as the pilot value set, with values ranging from −0.384 for the worst state to 0.997 for the mildest state (Table 2 [model 4] and Fig. 3). This model combines data from cTTO and DCE, which maximizes the use of the data, and it also had a low number of coefficients that were not statistically significant.

There was some evidence of interviewer effects based on coefficient size and statistical significance. Consequently, the estimated range varied when individual interviewers were excluded sequentially (Appendix Table 8 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2023.02.008). This may have been because of differences in the participants whom the interviewers saw, for example, some interviewers saw more females (60%) compared with others (48%), although these were not statistically significant differences (chi-square = 3.0, P =.81), or for age (F [6513] = 1.04, P =.40).

Discussion

The primary aim of this study was to assess the feasibility of valuing a new measure, the EQ-HWB-S, using a modified international protocol (EQ-VT vs2) developed for EQ-5D-5L, which employs cTTO and DCE. The results indicate that applying the modified protocol to the new measure using EQ-PVT was feasible.

The time taken for a single cTTO task (mean 103 seconds) was longer than the minimum time set in EQ-VT vs2 protocol, which may indicate difficulties with engaging with the tasks. Nevertheless, minimum protocol times does not necessarily reflect what participants do, for example, in the Italian EQ-5D-5L valuation study,[12] the time taken to complete a single cTTO task ranged from 69 to 98 seconds. Given the longer EQ-HWB-S descriptive system and the use of an orthogonal array to generate states, which results in mixed states that may be more difficult to value, it is expected that individuals will take longer.

There were a small proportion of participants whom the interviewers thought did not engage or understand the preference-elicitation task, which is similar to what is observed in other studies.[12,30] Participants did not report problems understanding the task or the EQ-HWB-S states. Videoconferencing added other challenges related to who was recruited and how they engaged, which may have had an impact. Nevertheless, this mode of administration has successfully been applied in a full valuation study for the EQ-5D-5L[12] and shown to generate cTTO data similar to face-to-face interviews.[15,31]

There were no major concerns raised with QC, including in the number of inconsistencies or nontraders, which would have been indicators of poor feasibility of using cTTO for EQ-HWB-S. Nevertheless, 40 participants were identified for having any inconsistency for the worst state. It is recommended that future studies that test the use of the EQ-VT protocol use the "feedback module" (in which participants view the rank ordering of the states according to their TTO responses and highlight any they would reconsider) to check if participants are able to identify these inconsistencies when ranked against other states.[32]

The cTTO data covered the full TTO range from −1 to 1 with a proportion of values below zero with lower mean and larger SDs observed for states as the level sum score increased. The proportion of states valued at selected points (−1, 0, 0.5, and 1) was 7%, 3%, 10%, and 12%, respectively, which was reasonable compared with other EQ-VT vs2 studies, for example, United States[30] had 15%, 5%, 7%, and 21%, respectively, whereas for Italy[12] (conducted online via videoconferencing), this was 8%, 2%, 7%, and 11%, respectively.

Different models were estimated, and all models had disordering that occurred in most dimensions apart from mobility, daily activities, and pain, and this was most likely to occur for level 4 or 5. The dimensions with disordered levels all had frequency response options with level 4 as "often" and level 5 as "most or all of the time," which may have driven some of these results. It may also be related to the dimensions, for example, there were disordered level for cognition and anxiety across all models. EQ-HWB-S dimensions and response levels have been shown to work well from a measurement perspective,[33,34] but it is recommended that

Figure 3. Selected model—hybrid Tobit with heteroscedasticity.

	Mobility	Activity	Exhaustion	Loneliness	Cognition	Anxiety	Sad/depressed	Control	Pain
Level 2	−0.053	−0.041	−0.019	−0.021	−0.003	−0.022	−0.031	−0.004	−0.038
Level 3	−0.070	−0.063	−0.027	−0.052	−0.016	−0.022	−0.034	−0.045	−0.080
Level 4	−0.136	−0.150	−0.066	−0.101	−0.057	−0.069	−0.113	−0.065	−0.258
Level 5	−0.207	−0.199	−0.082	−0.120	−0.057	−0.092	−0.173	−0.082	−0.372

future studies consider using qualitative methods to assess how participants engage with these in a valuation context to ensure they are appropriate.

Although there were some differences in the estimated models, there were similarities between cTTO and DCE results with strong associations based on Lin's concordance correlation coefficient. Pain, daily activities, mobility, sadness/depression, and loneliness were ranked in the same way across the selected models, which may indicate consistency across preference-elicitation methods. EQ-5D-5L studies tend to have pain, mobility, and anxiety/depression as the worst dimensions,[6] for example, ranking in the EQ-5D-3L UK value set was pain, mobility, and anxiety/depression,[35] whereas this was pain, anxiety/depression, then mobility in the English EQ-5D-5L value set.[36] Anxiety is a separate item in the EQ-HWB-S and had relatively low weights, but it is based on frequency, not severity, that is used in EQ-5D. Depression alone has been found to have a greater utility decrement than the composite anxiety/depression or than anxiety on its own in a different study.[37] The similarities in ranking with EQ-5D-5L for similar dimensions provides some evidence of the feasibility of using cTTO and DCE for EQ-HWB-S. Nevertheless, more evidence is required, because the low ranking for the additional dimensions may reflect a lack of engagement with the whole EQ-HWB-S. It could also be an ordering effect, which has been found to have an impact for EQ-5D-5L,[38] because mobility and daily activities are the first 2 dimensions, and pain is the last dimension. Methodological work on the impact of ordering on valuation is recommended for the EQ-HWB-S.

Although direct comparison of disutilities associated with other measures is difficult because of differences in questions, response options, and states, it is useful to contextualize the EQ-HWB-S disutilities. The disutility associated with being at level 5 were generally smaller in EQ-HWB-S compared with EQ-5D-5L in overlapping dimensions (mobility: 0.200 vs. 0.22 to 0.613; usual activities: 0.195 vs. 0.153 to 0.385; pain: 0.355 vs 0.246 to 0.612; anxiety/depression: 0.088/0.171 vs. 0.19 to 0.646) based on results from 25 EQ-5D-5L value sets.[39]

Loneliness was important, but exhaustion, control, and cognition EQ-HWB-S items had relatively low weight even at the most severe levels. Fatigue can have low utility decrements relative to physical, emotional and social functioning, and pain, for example, in the EORTC QLU-C10D, disutility from all levels of fatigue range between −0.036 and −0.058 for United Kingdom[40] and −0.023 to −0.037 for Australia.[41] Pairwise studies show that EQ-5D-5L dimensions are ranked higher than social care related dimensions, such as control,[42] whereas cognition ranks high relative to other potential bolt-on dimensions, such as relationships, energy (tiredness), and sleep.[43] Nevertheless, these pairwise studies only show ranking—highly ranked dimensions may not attract large utility decrements relative to other dimensions.

Small utility decrements for some dimensions and levels, disordering, and lack of statistical significance may raise the question about the feasibility of valuing the EQ-HWB-S items using EQ-VT vs2. Although other studies have also found disordering, for example, United States,[30] and lack of statistical significance in some models, for example, Italy,[12] in this feasibility study, these issues may have arisen because of the relatively small sample and size of the design relative to the size of the classification system. Separately, small utility decrements may link to the choice of item within each dimension, for example, Finch et al[43] used "remembering" as the cognition test bolt-on dimension, whereas we used "concentrating and thinking clearly." The small utility decrement for feeling exhausted and control is in contrast to the importance of these aspects identified during the development of the EQ-HWB instrument, including qualitative[44] and psychometric evidence,[34] and stakeholders' views.[3] Crocker et al[45] found that individuals who had a disability placed relatively higher importance on broader quality of life dimensions, such as control relative to health status focused dimensions, based on a ranking exercise. In the Adult Social Care Outcomes Toolkit, measures for social care users[46] and for informal carers[47] having control over daily life was weighted highly using best-worst scaling, and not having control was considered to be the worst aspect relative to other social care and informal care dimensions. Understanding this incongruence is an important area of future research.

A hybrid Tobit model, which takes into account heteroscedasticity, was selected as a pilot value set based on maximizing the use of data, the number of inconsistencies, and statistical significance. The values for this model ranges from −0.384 for the worst state to

0.997 for the mildest state. This provides an opportunity to assess the validity of the EQ-HWB-S based on utility values while acknowledging the limitations associated with this feasibility study.

We used an established international protocol developed and refined to ensure high data quality.[4] We modified this protocol to minimize burden by reducing the number of states, using overlap with the DCE and showing potential response options, which has been shown to help especially in the context of DCE.[48] It is recommended that future EQ-HWB-S and EQ-HWB valuation studies take this into account and undertake further development and testing of approaches to support participant engagement to inform a future EQ-HWB-S valuation protocol. This includes selecting states, presentation of TTO and DCE states, and, potentially separation of the TTO and DCE into different samples/modes of administration as has been done for the EQ-5D-Y[49] or in "Lite"[18] protocol studies that offer a cost-effective valuation option.

Limitations

This study was a feasibility study that was conducted online with a relatively small sample. Online participants may not be representative of the wider population, for example our sample was highly educated—and this may mean they were more able to engage with EQ-WB-S states and the tasks. We also encountered problems with online recruitment. Although the QC criteria were met by all the interviewers, interviewers did varying numbers of interviews in varying time, which may have influenced the results. The mixed states (with 9 dimensions) from the TTO design may have been more difficult for participants to engage with than when valuing EQ-5D (with 5 dimensions) for which the EQ-VT protocol was developed. This was mitigated by reducing the number of states participants valued, but this reduced the number of times each state was valued. In addition, the designs for both cTTO and DCE represent only a small number of the possible states that could be valued. Coverage across different level sum scores was low in the TTO state selection. Future studies should consider using the existing study data to optimize the design and selection of the states for valuation of the EQ-HWB-S using a Bayesian approach.

Conclusions

This feasibility study demonstrated that EQ-VT vs2 could be applied to the EQ-HWB-S, and an initial pilot value set has been generated. Future work to inform the development of an international valuation protocol for EQ-HWB-S and the importance of the items in the EQ-HWB-S is recommended.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2023.02.008.

Article and Author Information

Accepted for Publication: February 9, 2023

Published Online: March 22, 2023

doi: https://doi.org/10.1016/j.jval.2023.02.008

Author Affiliations: School of Health and Related Research, University of Sheffield, Sheffield, England, UK (Mukuria, McDool, Rowen, Brazier); Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia (Peasgood); Curtin School of Population Health, Curtin University, Perth, Western Australia, Australia (Norman).

Correspondence: Clara Mukuria, PhD, School of Health and Related Research, University of Sheffield, Regent Ct, 30 Regent Str, Sheffield S14DA, England, United Kingdom. Email: c.mukuria@sheffield.ac.uk

Author Contributions: *Concept and design:* Mukuria, Peasgood, Rowen, Brazier
Acquisition of data: Mukuria, McDool
Analysis and interpretation of data: Mukuria, Norman, Brazier
Drafting of the article: Mukuria, McDool, Norman, Rowen, Brazier
Critical revision of the article for important intellectual content: Mukuria, Peasgood, Norman, Rowen, Brazier
Provision of study materials or patients: Mukuria
Obtaining funding: Mukuria, Peasgood, Norman, Rowen, Brazier
Administrative, technical, or logistic support: Mukuria, McDool
Other: Brazier

Conflicts of Interest Disclosures: Drs Mukuria, McDool, Peasgood, Rowen, and Brazier reported receiving grants from EuroQol Group and the Medical Research Council during the conduct of the study. Dr Mukuria reported receiving grants from EuroQol Group outside the submitted work. Dr Norman reported receiving grants from EuroQol Group during the conduct of the study and personal fees from EuroQol Group outside the submitted work. Dr Brazier reported receiving nonfinancial support from UK MRC and the EuroQol Research Foundation during the conduct of the study. Drs Mukuria, Peasgood, Norman, and Brazier reported Membership of the EuroQol Group Association. Dr Norman is an editor for *Value in Health* and had no role in the peer review-process of this article. No other disclosures were reported. The views expressed in this article are those of the authors and not necessarily of our funders, the National Institute of Health and Care Excellence, the Department of Health and Social Care, other departments, or the UK government or those acknowledged.

Funding/Support: This study is independent research funded by the UK Medical Research Council (Grant number 170620), and the EuroQol Research Foundation, Netherlands.

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 manuscript; or decision to submit the manuscript for publication.

Acknowledgment: The authors would like to thank NICE for highlighting the methodological research need to the MRC that resulted in the funding call entitled "Beyond the QALY," which led to this research being funded. The authors would like to acknowledge the support of the National Institute for Health Research Yorkshire and Humber Applied Research Collaboration (formerly CLAHRC), United Kingdom. The authors acknowledge the invaluable contributions of members of the project steering group, advisory group, and public and patient involvement and engagement groups. We would also like to thank members of the EuroQol Group Association for their input at plenary and academy meetings, especially Bas Janssen. The authors would like to acknowledge the contribution of all the participants who took part in the study. We would also like to thank the interviewers who undertook data collection: Ankita Mishra, Emily McDool, Hannah Hussain, Joy Adewusi, Louisa Robinson, Megan Frith, and Ryan Bramley. Finally, the authors would like to extend our acknowledgments to EuroQol Scientists Aureliano Finch, Bram Roudijk, and Elly Stolk for their input.

REFERENCES

1. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford, United Kingdom: Oxford University Press; 2017.
2. 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.
3. 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.
4. 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.

5. Mukuria C, Peasgood T, Brazier J. Applying EuroQol Portable Valuation Technology to the EQ Health and wellbeing Short (EQHWB-S): a pilot study. School of Health and Related Research, University of Sheffield Discussion Paper Series. https://eprints.whiterose.ac.uk/179740/1/Applying%20EuroQol%20Portable%20Valuation%20Technology%20to%20the%20EQ%20Health%20and%20Wellbeing%20Short%20%28EQ-HWB-S%29%20-%20A%20pilot%20study.pdf. Accessed March 13, 2023.
6. Roudijk B, Ludwig K, Devlin N. EQ-5D-5L value set summaries. In: Devlin N, Roudijk B, Ludwig K, eds. *Value Sets for EQ-5D-5L: A Compendium, Comparative Review & User Guide*. New York, NY: Springer International Publishing; 2022:55–212.
7. Jonker MF, Donkers B, de Bekker-Grob E, Stolk EA. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. *Health Econ*. 2019;28(3):350–363.
8. Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. The effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. *Value Health*. 2018;21(7):767–771.
9. Oppe M, Norman R, Yang Z, van Hout B. Experimental design for the valuation of the EQ-5D-5L. In: Devlin N, Roudijk B, Ludwig K, eds. *Value Sets for EQ-5D-5L: A Compendium, Comparative Review & User Guide*. New York, NY: Springer International Publishing; 2022:29–54.
10. Metrics Choice. *Ngene [Software for Experimental Design]*. Sydney: Choice Metrics; 2018.
11. Cook RD, Nachtrheim CJ. A comparison of algorithms for constructing exact D-optimal designs. *Technometrics*. 1980;22(3):315–324.
12. 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.
13. 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.
14. Lipman SA. Time for tele-TTO? Lessons learned from digital interviewer-assisted time trade-off data collection. *Patient*. 2021;14(5):459–469.
15. 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;309:115227.
16. *Microsoft*. PowerPoint. Version Professional Plus. Microsoft Corporation; 2016.
17. Welie AG, Gebretekle GB, Stolk E, et al. Valuing health state: an EQ-5D-5L value set for Ethiopians. *Value Health Reg Issues*. 2020;22:7–14.
18. Yang F, Katumba KR, Roudijk B, et al. Developing the EQ-5D-5L value set for Uganda using the 'Lite'protocol. *Pharmacoeconomics*. 2022;40(3):309–321.
19. 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;40(9):863–882.
20. Tobin J. Estimation of relationships for limited dependent variables. *Econometrica*. 1958;26(1):24–36.
21. Jones A. *Applied Econometrics for Health Economists: A Practical Guide*. Boca Raton, FL: CRC Press; 2007.
22. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Med Decis Making*. 2015;35(3):328–340.
23. Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Med Care*. 2017;55(7):e51–e58.
24. Lawrence LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*. 1989;45(1):255–268.
25. Harvey AC. Estimating regression models with multiplicative heteroscedasticity. *Econometrica*. 1976;44(3):461–465.
26. StataCorp. Stata Statistical Software. Release 17. College Station, TX: StataCorp LLC; 2021.
27. A05 SA: employment, unemployment and economic inactivity by age group (seasonally adjusted). Office for National Statistics. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/employmentunemploymentandeconomicinactivitybyagegroupseasonallyadjusteda05sa. Accessed March 13, 2023.
28. Graduates in the UK labour market. Office for National Statistics. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/graduatesintheuklabourmarket/2017. Accessed March 13, 2023.
29. Zipped population projections data files, England. Office for National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/datasets/z3zippedpopulationprojectionsdatafilesengland. Accessed March 13, 2023.
30. 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.
31. Bennett B, Hill S, Gibson A, Oluboyede Y, Longworth L, Shaw JW. Assessing the comparative feasibility, acceptability and equivalence of videoconference interviews and face-to-face interviews using the time trade-off technique. presented at: 39th EuroQol Plenary 2022; Chicago; 2022.
32. Wong ELY, Shah K, Cheung AWL, Wong AYK, Visser M, Stolk E. Evaluation of split version and feedback module on the improvement of time trade-off data. *Value Health*. 2018;21(6):732–741.
33. Carlton J, Peasgood T, Mukuria C, et al. Generation, selection, and face validation of items for a new generic measure of quality of life: the EQ-HWB. *Value Health*. 2022;25(4):512–524.
34. Peasgood T, Mukuria C, Brazier J, et al. Developing a new generic health and wellbeing measure: psychometric survey results for the EQ-HWB. *Value Health*. 2022;25(4):525–533.
35. Dolan P. Modeling valuations for EuroQol health states. *Med Care*. 1997;35(11):1095–1108.
36. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ*. 2018;27(1):7–22.
37. McDonald R, Mullett TL, Tsuchiya A. Understanding the composite dimensions of the EQ-5D: an experimental approach. *Soc Sci Med*. 2020;265:113323.
38. Mulhern B, Shah K, Janssen MF, Longworth L, Ibbotson R. Valuing health using time trade-off and discrete choice experiment methods: does dimension order impact on health state values? *Value Health*. 2016;19(2):210–217.
39. Roudijk B, Janssen B, Olsen JA. How do EQ-5D-5L value sets differ? In: Devlin N, Roudijk B, Ludwig K, eds. *Value Sets for EQ-5D-5L: A Compendium, Comparative Review & User Guide*. New York, NY: Springer International Publishing; 2022:235–258.
40. Norman R, Mercieca-Bebber R, Rowen D, et al. U.K. utility weights for the EORTC QLU-C10D. *Health Econ*. 2019;28(12):1385–1401.
41. King MT, Viney R, Simon Pickard A, et al. Australian utility weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the cancer-specific quality of life questionnaire, EORTC QLQ-C30. *Pharmacoeconomics*. 2018;36(2):225–238.
42. Mulhern B, Norman R, Lourenco R, Malley J, Street D, Viney R. Investigating the relative value of health and social care related quality of life using a discrete choice experiment. *Soc Sci Med*. 2019;233:28–37.
43. Finch AP, Brazier J, Mukuria C. Selecting bolt-on dimensions for the EQ-5D: testing the impact of hearing, sleep, cognition, energy, and relationships on preferences using pairwise choices. *Med Decis Making*. 2021;41(1):89–99.
44. Mukuria C, Connell J, Carlton J, et al. Qualitative review on domains of quality of life important for patients, social care users, and informal carers to inform the development of the EQ-HWB. *Value Health*. 2022;25(4):492–511.
45. Crocker M, Hutchinson C, Mpundu-Kaambwa C, Walker R, Chen G, Ratcliffe J. Assessing the relative importance of key quality of life dimensions for people with and without a disability: an empirical ranking comparison study. *Health Qual Life Outcomes*. 2021;19(1):264.
46. Netten A, Burge P, Malley J, et al. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol Assess*. 2012;16(16):1–166.
47. Batchelder L, Malley J, Burge P, et al. Carer social care-related quality of life outcomes: estimating English preference weights for the adult social care outcomes toolkit for carers. *Value Health*. 2019;22(12):1427–1440.
48. Cole A, Shah K, Mulhern B, Feng Y, Devlin N. Valuing EQ-5D-5L health states 'in context'using a discrete choice experiment. *Eur J Health Econ*. 2018;19(4):595–605.
49. Ramos-Goñi JM, Oppe M, Stolk E, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics*. 2020;38(7):653–663.