The University Of Sheffield.

School of Health And Related Research

# HEALTH ECONOMICS & DECISION SCIENCE

## Discussion Paper Series

**Title: Estimating a Preference-Based Index for mental health from the Recovering Quality of Life (ReQoL) measure: Valuation of ReQoL-UI**

**Authors: Anju Devianee Keetharuth, Donna Rowen, Jakob Bue Björner and John Brazier**

Corresponding author: Dr Anju Devianee Keetharuth, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK. Tel: +44 (0) 114 222 0884,
Email: d.keetharuth@sheffield.ac.uk

Estimating a Preference-Based Index for mental health from the Recovering Quality of Life (ReQoL) measure: Valuation of ReQoL-UI

**Authors:** Anju Devianee Keetharuth, Donna Rowen, Jakob Bue Björner and John Brazier

**Corresponding author:** Anju Devianee Keetharuth. Phd

School of Health of Related Research, University of Sheffield, United Kingdom
Telephone: +44 (0) 114 222 0884
ORCID: 0000-0001-8889-6806

**Affiliation of co-authors**

| | |
|---|---|
| Jakob Bue Björner | Optum Patient Insights, Rhode Island, United States and University of Copenhagen, Denmark<br>ORCID: 0000-0001-7033-8224 |
| John Brazier | School of Health of Related Research, University of Sheffield, United Kingdom<br>ORCID: 0000-0001-8645-4780 |
| Donna Rowen | School of Health of Related Research, University of Sheffield, United Kingdom<br>ORCID: 0000-0003-3018-5109 |


| | |
|---|---|
| **Funding** | This research is funded by the National Institute for Health Research (NIHR) Policy Research Programme (Ref:104/0001). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. |
| **Acknowledgements** | The authors would like to thank all the participants in the study, the healthcare professionals and other staff involved in the recruitment of participants and conduct of the study. |

**Abstract (247 words)**

**Objectives**

There are increasing concerns about the appropriateness of generic preference-based measures to capture health benefits in the area of mental health. This study estimates preference weights for a new measure, Recovering Quality of Life (ReQoL-10), to better capture the benefits of mental health care.

**Methods**

Psychometric analyses of a larger sample of mental health service users (n = 4266) using confirmatory factor analyses and item response theory (IRT) were used to derive a health state classification system and inform the selection of health states for utility assessment. A valuation survey with members of the UK public representative in terms of age, gender and region was conducted using face-to-face interviewer administered time-trade-off (TTO) with props. A series of regression models were fitted to the data and the best performing model selected for the scoring algorithm.

**Results**

The ReQoL-UI classification system comprises six mental health items and one physical health (PH) item. Sixty-four health states were valued by 305 participants. The preferred model was a random effects model, with significant and consistent coefficients and best model fit. Estimated utilities modelled for all health states ranged from -0.195 (state worse than dead) to 1 (best possible state).

**Conclusions**

The development of the ReQoL-UI is based on a novel application of IRT methods for generating the classification system and selecting health states for valuation. Conventional TTO was used to elicit utility values that are modelled to enable the generation of QALYs for use in cost-utility analysis of mental health interventions.

**Declaration of interest:** None.

**Keywords: Mental health; preference-based measure; QALYs; ReQoL-10; ReQoL-20**

## Introduction

Quality adjusted life years (QALYs), a composite measure combining quality of life and length of life, are used in cost-effectiveness analyses. Typically, the quality of life component of the QALY is generated using an off-the-shelf generic or condition-specific preference based measure. The most commonly used generic preference-based measure, EQ-5D, has a focus on physical health (PH) with only one of the five dimensions pertaining to mental health (MH). There is growing evidence that EQ-5D is not well suited for use in certain areas of MH (1-4), raising the question of whether a condition-specific preference-based measure developed would be more appropriate for use in cost-effectiveness analyses in MH. Such a measure may have the advantage of performing better psychometrically, as it may be better able to detect changes in MH over time and differences across treatments. In addition, the measure would be more relevant for, and acceptable to people for inclusion in data collection with MH problems.

The Recovering Quality of Life (ReQoL) measures, ReQoL-10 and ReQoL-20, were developed to be used in routine practice and to evaluate interventions in the area of MH (5). ReQoL-10 contains ten MH items and ReQoL-20, ten additional items. Both versions contain a PH item (Appendix 1). The theoretical framework underpinning the themes for the measures were established from a qualitative literature review and in-depth interviews (1, 6, 7). Six MH themes (activity; belonging; choice, control and autonomy; hope; self-perception and well-being) and a theme relating to PH were identified. Psychometric evidence generated through two studies recruiting over 6,500 service users (8, 9) was combined with the qualitative evidence (10) to select items for the final measures (11).

The aim of this paper is to present the derivation of ReQoL utility index (ReQoL-UI), a recovery-focussed preference-based measure derived from ReQoL-10 and ReQoL-20. It presents the development of a novel approach, since standard methods used to select health states for valuation rely on independence between dimensions which is not present between the MH items in ReQoL-UI.

## Methods

ReQoL-UI was constructed in four stages: 1) the derivation of the classification system of ReQoL -UI; 2) the choice of health states for valuation using item response theory (IRT); 3) the time trade-off valuation (TTO) survey used to elicit values for a selection of ReQoL-UI health states; and 4) the modelling of preference weights that can be used to generate utility values for all health states defined by the ReQoL -UI.

### *ReQoL data for use in stages 1 and 2*

Data were gathered from 4,266 individuals accessing MH services from primary (12%) and secondary care (67%) and from a trial cohort for a depression study (5%). The sample is described in detail elsewhere (5, 9). In summary, 58% of the sample was female, the age range was 16 to 98 and mean (sd) age was 47 (17) years. Respondents self-reported a wide range of diagnoses including common MH disorders (51%) and psychotic disorders (18%).

### *Stage 1: Development of the ReQoL-UI classification system*

The aim of this stage is to generate a health state classification system amenable for valuation. The ten ReQoL-10 MH items and the PH item were therefore considered for the reduced classification system as these items also appear in ReQoL-20. However, the use of all the 11 items to elicit preference weights during the valuation exercise would be cognitively too onerous. To maintain the face validity of the ReQoL-10 measure, we chose one item from each of the six MH themes and the PH item, all being identified as important to services users experiencing MH difficulties. To select the MH items, we adopted the following steps: a) consider the dimensionality of the ReQoL item set; b) exclude any misfitting item(s); c) select items with the best psychometric properties. For step a) Confirmatory factor analysis (CFA) was undertaken and model fit was assessed using root mean square error approximation (RMSEA) and common factor information criteria (CFI). In a bifactor model providing an adequate fit, the negatively (n = 24) and positively worded items (n = 15) loaded onto a 'negative' factor and a 'positive' factor respectively (9). However, the explained common variance of the global factor was 85% suggesting the measure was uni-dimensional.

To undertake steps b) and c), the graded response item response theory (IRT) model was fitted to the 39 items to estimate item parameters and the full results are presented elsewhere (8). The graded response model is a flexible IRT model that supports examination of item and scale properties, estimation of item characteristics, and estimation of mental health-related functioning severity scores based on the responses. Theta (θ) scores represent overall severity of the health states in terms of mental health-related function, where higher scores indicate better functioning. For various levels of theta scores ranging from -2 to 2 in intervals of 0.4, the ReQoL-10 items (as subset of the 39 items) were ranked in order of the item's contribution to measurement precision (Fisher information function) (12). This approach ensured that the most informative items were chosen and that the items covered the range of severity observed among MH service users. The CFA analyses were conducted in MPlus (13), IRT analyses in IRTPRO 3.1 (14).

***Stage 2: Selecting health states***

Standard approaches for selecting health states (e.g. orthogonal arrays) for valuation rely on independence between dimensions, which is not the case in ReQoL-UI. Previous studies where the classification system has a unidimensional component with correlated items have used a Rasch vignette approach (15, 16). The latter approach uses Rasch-based threshold analysis to select commonly observed health states for valuation, and then generates utility values for all possible health states using a regression model that predicts TTO utilities using the Rasch score for the health state. Here, we adapted this approach to IRT methods rather than Rasch analysis, since IRT models have been shown to provide a good description of the ReQoL items (8), and IRT provides more flexibility in modelling than Rasch analysis.

We selected health states for valuation choosing the response combinations that are most likely to be encountered in practice by estimating the probability of each possible combination of health states according to IRT. We performed such calculations across the entire range of the IRT MH score from -2.18 (worst score on all six items) to 1.85 (best scores on all six items), categorising this range into 15 score groups (15 represented a reasonable trade-off between complexity and detail). Score group 1 through 8 covered the range from -2.18 to 0, while score group 9-15 covered the range from 0 to 1.85. For each score group, the response combinations providing a score within this range were ranked according to their probability and the three most likely response combinations were chosen as health states. For score group 15, only one response combination (555555) provided a score in this range, so this score group only contributed one health state (for a total of 14*3+1=43 health states). To ensure accurate utility assessment of poor MH states, we purposively oversampled response combinations providing a MH score below the average. For each of the eight score groups below 0, we selected two additional response combinations for a total of 59 MH states (43+8*2). These were combined with the PH item by randomly selecting one physical level to be considered together with the MH states (using the random number generator in Excel). Five additional combinations of PH and MH states were added. For mental health, these included the worst possible MH score (111111), the best possible score (555555) and a score indicating "average" MH (333333). This approach was chosen since MH and PH form two separate dimensions, and appear separately in the regressions analyses undertaken in stage 4. All items were scaled from 1 to 5 with level 5 indicating the worst PH or MH (highest level of impact).

***Stage 3: Design and conduct of the valuation study***

People's preferences for the sample of health states previously selected were elicited using TTO, a choice-based technique, in face-to-face interviews with members of the UK public. Respondents were selected to generate a nationally representative sample based on age and gender from postcodes in Scotland, England and Wales. Households in the selected areas received a letter in advance, advising them that an interviewer would call, with an opportunity to opt out. Interviews were managed by a market research agency and were conducted by experienced interviewers trained by the researchers. Interviews were held in the respondents' own home and respondents were offered £10 for their participation.

During the interview, respondents first completed demographic and health questions followed by the ReQoL-10 to familiarise themselves with the health state classification system and response options. Second, respondents undertook a warm up task in the form of a practice TTO question. The interviewer had the discretion to decide whether a second practice question was necessary. Third, respondents undertook TTO valuation of eight different health states. The Measurement and Valuation of Health protocol and its related props were used for states better than dead (17) and lead-time TTO was used for states worse than dead (18). Respondents were first asked whether they would prefer to live in the health state to be valued for 10 years and then die, or to die immediately to establish whether the health state was better, worse or equal to being dead. For health states better than dead, participants were asked to imagine they would be in the health state that was being valued for a period of ten years. They were then asked to consider a number of shorter periods in full health ($x$) to ascertain how many years of full health the respondent was willing to give up to avoid being in the impaired health state that was being valued. At the point where respondents were indifferent between $x$ years in full health and 10 years in the state, the state took the value $x/10$. For states worse than dead, lead-time TTO was used which involves the same approach but adds a lead-time of 10 years to both full health and the impaired health state to allow respondents to trade these 10 years to avoid the impaired health state. Finally, respondents rated how difficult they found the tasks, and interviewers rated how well they thought the respondent had understood and engaged with the task.

***Stage 4 Modelling health state preferences***

The ReQoL-UI MH items form a uni-dimensional MH component, with the PH item constituting a second dimension. Therefore, similar to the modelling approach used in the Rasch vignette approach (19), TTO values were regressed on the IRT-based MH score (estimated through the expected a posteriori (EAP) approach) and dummy variables to represent four of the severity levels of the PH item (with level 1 as the reference case). The EAP estimates (theta (θ)) were rescaled from a range of -2 to 2 to a scale of 0 (best possible mental health) to 1 (worst possible mental health). Different regression models were fitted using mean and individual level data including a simple linear relationship, quadratic and cubic relationships. First, model specifications included mean level Ordinary Least Squares (OLS) where mean scores were regressed on the rescaled θ scores and on dummy variables for the levels of the PH item. To account for multiple observations per individuals we also estimated random effects (RE) models (20) using maximum likelihood estimation. The error term $\varepsilon_{ij} = u_j + e_{ij}$ where $u_j$ is the random effect and $e_{ij}$ represents the random error term for the $i$th health state valuation of the $j$th individual.

$$y^{s}_{ij} = f(\theta, x_{\lambda,}\ \beta) + \varepsilon^{s}_{ij,} \qquad y^{s}_{i,j} = \{1 - \frac{z_{ij}}{w_{ij}} \quad \text{if state better than dead}$$

$$\{1 + \frac{z_{ij}}{10}\} \quad \text{if state worse than dead}$$

Where i = 1,2 … n represents the individual health states and j = 1,2 … m represents the respondents. The dependent variable $y^{s}_{ij}$ is disutility (1-TTO) for health state i valued by respondent j and θ represents IRT scores for the corresponding health state, X is a vector of dummy explanatory variables for each level λ of the PH items with level λ = 1 acting as a baseline. All models excluded a constant because we used full health as defined by ReQoL-UI level 1111111 as our upper anchor for TTO (21). We explored the inclusion of interaction terms that interacted the severity of the MH component, theta, with the PH dimension, where, as health worsens the interaction term increases. We estimated consistent models, where adjacent inconsistent levels of the physical dimension were merged, to ensure that as health worsened the utility value would not increase. All modelling was performed using STATA 15 (22).

***Model performance***

Several criteria were used to evaluate model performance: (i) inconsistencies in parameter estimates and significance of coefficients; (ii) comparing predictive model performance using root mean square error (RMSE), mean absolute error (MAE), difference between actual and predicted values at health state level, percentage of observations with absolute errors (AE) > 0.05 and >0.1; and plots of actual and predicted health state values; (iii) comparing Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for different model specifications within the same types of models.

## Results

### *Stage 1 Health state classification system*

Table 1 reports analyses of the ReQoL-10 items, using the IRT results. One item '*I could do the things I wanted to do*' was excluded as it was misfitting, indicating that standard IRT scoring may not be appropriate for this item. Through ranking the remaining nine items in order of highest information across different severity levels for mental health, five items were selected; items providing the most information at the highest severity levels were: *'I thought my life was not worth living'* and *'I felt unable to cope'* and *'I enjoyed what I did'*; the items providing the most information at the low severity end were: *'I felt happy'* and *'I felt confident in myself'*. To ensure that each theme was represented, an item from the belonging theme, *'I felt lonely'* was chosen even though it was the fifth best item at both the severe and milder ends. The selected items were rephrased to the present tense (Table 2).

Table 1 Evaluation of most informative item by each score level

| | Item Information Functions | | | | | | | | | Most informative item on each score level ranked by iteration | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| theta | ACT1 | ACT5P[a] | BEL2[a] | BEL3P | CHO4[a] | HOP1P | HOP4[a] | SEL2P[a] | WB11P[a] | Iteration 1 | Iteration 2 | Iteration 3 | Iteration 4 | Iteration 5 |
| -2 | 0.703 | 1.034 | 0.700 | 0.648 | 0.923 | 0.578 | 1.305 | 0.410 | 0.749 | **HOP4** | **ACT5P** | CHO4 | WB11P | ACT1 |
| -1.6 | 0.953 | 1.571 | 1.143 | 0.731 | 2.319 | 0.868 | 2.109 | 0.976 | 1.576 | **CHO4** | HOP4 | WB11P | ACT5P | **BEL2** |
| -1.2 | 1.077 | 1.753 | 1.443 | 0.764 | 3.251 | 1.065 | 2.413 | 1.731 | 2.159 | CHO4 | HOP4 | WB11P | ACT5P | SEL2P |
| -0.8 | 1.114 | 1.783 | 1.537 | 0.778 | 3.373 | 1.133 | 2.475 | 2.081 | 2.179 | CHO4 | HOP4 | WB11P | SEL2P | ACT5P |
| -0.4 | 1.124 | 1.745 | 1.560 | 0.784 | 3.370 | 1.154 | 2.455 | 2.140 | 2.277 | CHO4 | HOP4 | WB11P | SEL2P | ACT5P |
| 0 | 1.124 | 1.735 | 1.562 | 0.785 | 3.402 | 1.157 | 2.145 | 2.158 | 2.231 | CHO4 | WB11P | SEL2P | HOP4 | ACT5P |
| 0.4 | 1.120 | 1.821 | 1.519 | 0.779 | 3.271 | 1.166 | 1.330 | 2.187 | 2.304 | CHO4 | WB11P | SEL2P | ACT5P | BEL2 |
| 0.8 | 1.098 | 1.800 | 1.323 | 0.754 | 2.283 | 1.160 | 0.581 | 2.203 | 2.284 | **WB11P** | CHO4 | SEL2P | ACT5P | BEL2 |
| 1.2 | 1.027 | 1.534 | 0.913 | 0.685 | 0.897 | 1.084 | 0.211 | 1.968 | 2.184 | WB11P | **SEL2P** | ACT5P | HOP1P | ACT1 |
| 1.6 | 0.833 | 0.961 | 0.495 | 0.554 | 0.262 | 0.874 | 0.071 | 1.269 | 1.536 | WB11P | SEL2P | ACT5P | HOP1P | ACT1 |
| 2 | 0.554 | 0.458 | 0.231 | 0.393 | 0.070 | 0.579 | 0.023 | 0.581 | 0.716 | WB11P | SEL2P | HOP1P | ACT1 | ACT5P |

Key: [a] Most informative items chosen for the health state classification system: ACT5P *'I enjoyed what I did'* BEL2 '*I felt lonely*' CHO4 *'I felt unable to cope'* HOP4 *'I thought my life was not worth living'* SEL2P *'I felt confident in myself'* WB11P *'I felt happy'*

The following were not selected: ACT1 *'I found it hard to get started with everyday task*' BEL3P '*I felt able to trust others*' HOP1P '*I felt hopeful about my future*'. The remaining tenth item 'I could do the things I wanted to do' was a misfitting item.

Table 2 ReQoL descriptive system

| Theme | Description of health states | Levels |
|---|---|---|
| 1. Activity (act5p: I enjoyed what I did) | I enjoy what I do **most or all of the time** | 1 |
| | I often **enjoy** what I do | 2 |
| | I **sometimes** enjoy what I do | 3 |
| | I **only occasionally** enjoy what I do | 4 |
| | I **never** enjoy what I do | 5 |
| 2. Belonging and relationships (bel2: I felt lonely) | I **never** feel lonely | 1 |
| | I **only occasionally** feel lonely | 2 |
| | I **sometimes** feel lonely | 3 |
| | I **often** feel lonely | 4 |
| | I feel lonely **most or all of the time** | 5 |
| 3. Choice, control and autonomy (cho4: I felt unable to cope) | I **never** feel **unable** to cope | 1 |
| | I **only occasionally** feel unable to cope | 2 |
| | I **sometimes** feel unable to cope | 3 |
| | I **often** feel unable to cope | 4 |
| | I feel unable to cope **most or all of the time** | 5 |
| 4. Hope (hop4: I thought my life was not worth living) | I **never** think that my life is not worth living | 1 |
| | I **only occasionally** think that my life is not worth living | 2 |
| | I **sometimes** think my life is not worth living | 3 |
| | I **often** think my life is not worth living | 4 |
| | **Most or all of the time** I think my life is not worth living | 5 |
| 5. Self-perception (sel2p: I felt confident in myself) | I feel confident in myself **most or all of the time** | 1 |
| | I **often** feel confident in myself | 2 |

| | | |
|---|---|---|
| | I **sometime**s feel confident in myself | 3 |
| | I **only occasionally** feel confident in myself | 4 |
| | I **never** feel confident in myself none of the time | 5 |
| 6. Wellbeing (wb11p: I felt happy) | I feel happy **most or all of the time** | 1 |
| | I **often** feel happy | 2 |
| | I **sometimes** feel happy | 3 |
| | I **only occasionally** feel happy | 4 |
| | I feel **never** happy | 5 |
| 7. Physical health item (Please describe your physical health – problems with pain, mobility, difficulties caring for yourself or feeling physically unwell) | I have **no problems** with physical health | 1 |
| | I have **slight problems** with physical health | 2 |
| | I have **moderate problems** with physical health | 3 |
| | I have **severe problems** with physical health | 4 |
| | I have very **severe problems** with physical health | 5 |

### *Stage 2: Selecting health states*

The method described above yielded 59 health states for valuation (Table A1 in the appendix)). The additional five health states provided the opportunity to add the pits state ‘5555555’ as the best state was already selected. Two other states were selected so that more severe levels of physical impairment were combined with the best MH state to isolate the impact of PH more clearly; and finally, a moderate severity state was added.

### *Stage 3: Design and conduct of the valuation study*

*Valuation survey participants*

Interviews (n = 305) were carried out by 15 experienced and trained interviewers each undertaking between five and 30 interviews. The proportion of total suitable participants answering their door at the time of the interview was 28%. Five participants were excluded from the analysis: four valued all health states as identical and less than 1, implying they did not understand the task; one valued all health states as worse than dead, implying that they thought that no state was worth living. The characteristics of the sample are compared with the population from England and Wales (Table 3).

Table 3 Characteristics of respondents in the valuation survey

| | Mean | SD | Range | England & Wales norms |
|---|---|---|---|---|
| Age | 51.6 | 19.1 | 18 to 96 | 39 [a] |
| Life satisfaction score | 8.0 | 1.8 | 2 to 10 | 7.5[b] |
| Health satisfaction score | 7.7 | 2.0 | 1 to 10 | |
| | | | | |
| | | **n** | **Percentage (%)** | **England & Wales norms[c] (%)** |
| Gender | Male | 135 | 45.0 | 49.1 |
| | Female | 164 | 54.7 | 50.9 |
| | Other | 1 | 0.3 | |
| Marital Status | Single | 67 | 22.3 | 34.6 |
| | Married / Partner | 161 | 53.7 | 46.6 |
| | Separated/Divorced | 26 | 8.7 | 11.6 |
| | Widowed | 45 | 15.0 | 7.0 |
| | Prefer not to say | 1 | 0.3 | |
| Ethnicity | White | 278 | 92.7 | 86.0 |
| | Asian / Asian British | 16 | 5.3 | 7.5 |
| | Black / African / Caribbean / Black British | 3 | 1.0 | 3.3 |
| | Other ethnic group | 3 | 1.0 | 3.2 |
| Degree | Yes | 87 | 29.0 | 27.0 |
| | No | 202 | 67.3 | |
| | Missing | 11 | 3.7 | |
| Main activity | Employed | 146 | 48.7 | 61.7 |
| | Retired | 97 | 32.3 | 13.9 |
| | Housework | 22 | 7.3 | 4.3 |
| | Student | 5 | 1.7 | 9.3 |
| | Unemployed | 16 | 5.3 | 4.4 |
| | Long-term sick | 8 | 2.7 | 4.3 |
| | Other | 6 | 2.0 | 2.2 |
| Overall health | Excellent | 34 | 11.3 | |
| | Very good | 126 | 42.0 | |
| | Good | 96 | 32.0 | |
| | Fair | 31 | 11.4 | |
| | Poor | 10 | 3.3 | |
| | Missing | 1 | 0.3 | |
| Age categories | 16-25 | 20 | 6.7 | 11.9% |
| | 26-64 | 173 | 57.7 | 52.8% |
| | 65 and over | 84 | 28.0 | 16.4% |
| | Missing | 23 | 7.7 | |
| | Yes | 83 | 27.7 | |

| | | | | |
|---|---|---|---|---|
| Experienced serious illness yourself | No | 211 | 70.3 | |
| | Missing | 6 | 2.0 | |
| Experienced serious illness in the family | Yes | 143 | 47.7 | |
| | No | 149 | 49.7 | |
| | Missing | 8 | 2.7 | |
| Experienced serious illness in caring for others | Yes | 77 | 25.7 | |
| | No | 215 | 71.7 | |
| | Missing | 8 | 2.7 | |
| How well interviewer thought the respondent understood and carried out the TTO tasks during the interview? (answered by interviewers) | Understood and performed exercises easily | 192 | 63.37 | |
| | Some problems but seemed to understand the exercises in the end | 106 | 34.98 | |
| | Doubtful whether the respondent understood the exercises | 5 | 1.65 | |
| Level of concentration and effort of the respondent as perceived by the interviewer | Concentrated very hard and put in a great deal of effort | 143 | 47.19 | |
| | Concentrated fairly hard and put in a some effort | 140 | 46.20 | |
| | Didn't concentrate very hard and put in little effort into it | 18 | 5.94 | |
| | Concentrated at the beginning but lost interest/concentration towards the end | 2 | 0.66 | |

Notes: [a] median age only was found [b] ONS Life satisfaction 2016 [c] Statistics for England in the Census 2011. The census includes persons aged 16 and above whereas this study only surveys persons aged 18 and above.

Forty-eight respondents (16%) reported a MH condition, out of whom 35 were receiving treatment. The three most reported MH conditions were anxiety, depression and stress-related (including PTSD). One hundred respondents (34%) reported a physical problem with the three most reported conditions being high blood pressure, tiredness/fatigue and pain. Only five (2%)

and 29 (10%) respondents reported that they found the questions very difficult and quite difficult to understand respectively. Interviewers doubted that five (2%) respondents had not quite understood the questions; that 17 respondents (6%) did not concentrate very hard and had put little effort into the valuation task and that two respondents (<1%) concentrated at the beginning but subsequently lost concentration or interest. Interviews lasted 34 minutes on average (sd = 10).

***Health State Values***

The number of observations per state vary from 27 to 44. The distribution of observed TTO values show that 20% and 6% of observations at 1 and -1 respectively (Appendix Figure A1). The mean observed TTO values by health state range from -0.178 (worst state = 5455555) to 0.966 (best state = 1111111). In the first three states (Table A1), it is evident that as PH severity increases, the mean TTO value falls. The worst health state (5555555) has higher mean TTO (-0.128) than the state 5455555 (-0.178) and state 5553554 (-0.144), but it should be noted that different respondents valued these health states.

***Stage 4: Modelling the health state utility data to generate utility values for all health states***

The best performing mean linear and quadratic OLS and RE models, assessed in terms of MAE, RMSE, AIC, BIC and percentage of observation with AE greater than 0.1 and 0.05, are presented in Table 4. The RE models were preferred to fixed effects models using the Hausman test. There were some inconsistent coefficients in the linear models for levels 2 and 3 of PH compared with level 1. The coefficients for the quadratic models were all in the direction expected, where increasing severity leads to decreases in utility, with the exception of the interaction terms combining level 2 of PH and theta (compared with level 1) for both OLS and RE Model 2, where the coefficient was positive rather than negative. The cubic models are not presented as they do not provide monotonous decreasing utility scores for worse MH

(Appendix Table A4). The best performing mean level OLS model (Model 2) and RE models (Model 6) consist of a quadratic specification of $\theta$ with interaction terms for $\theta$ and levels 3, 4 and 5 of PH. They have the lowest RMSE, lowest AIC and BIC and lowest percentage of observations with AE < 0.1 and 0.05. The interaction terms in all models are negative. As shown in Figure 1 and Figure Appendix A2), neither of the models exhibit systematic bias in the predictions by severity for the majority of health states except for the most severe states where larger prediction errors were observed.

RE Model 6 is the overall preferred model because it had better predictive ability, albeit only slightly better than the OLS model when comparing the lowest proportion of absolute errors greater than 0.05 and 0. 1. The estimates for the best health state and worst states are 1 and -0.195 respectively. Depicting the mean TTO predicted by Model 6 for levels 2 to 5 of PH indicate that the decrements for the first two levels of the PH item are very similar, with by far the largest gap being between levels 3 and 4 of physical functioning (Figure 2).

Figure 1 Plot of predicted versus observed utility values for the Random Effects Model 6

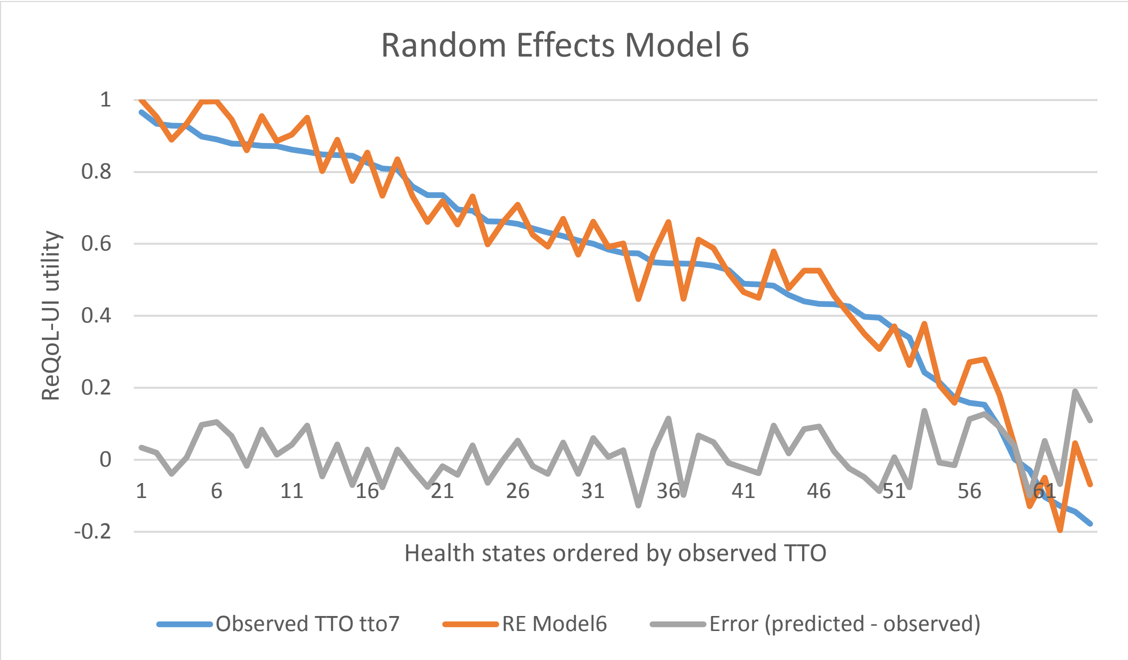


Figure 2 Mean predicted TTO and for each level of physical health.

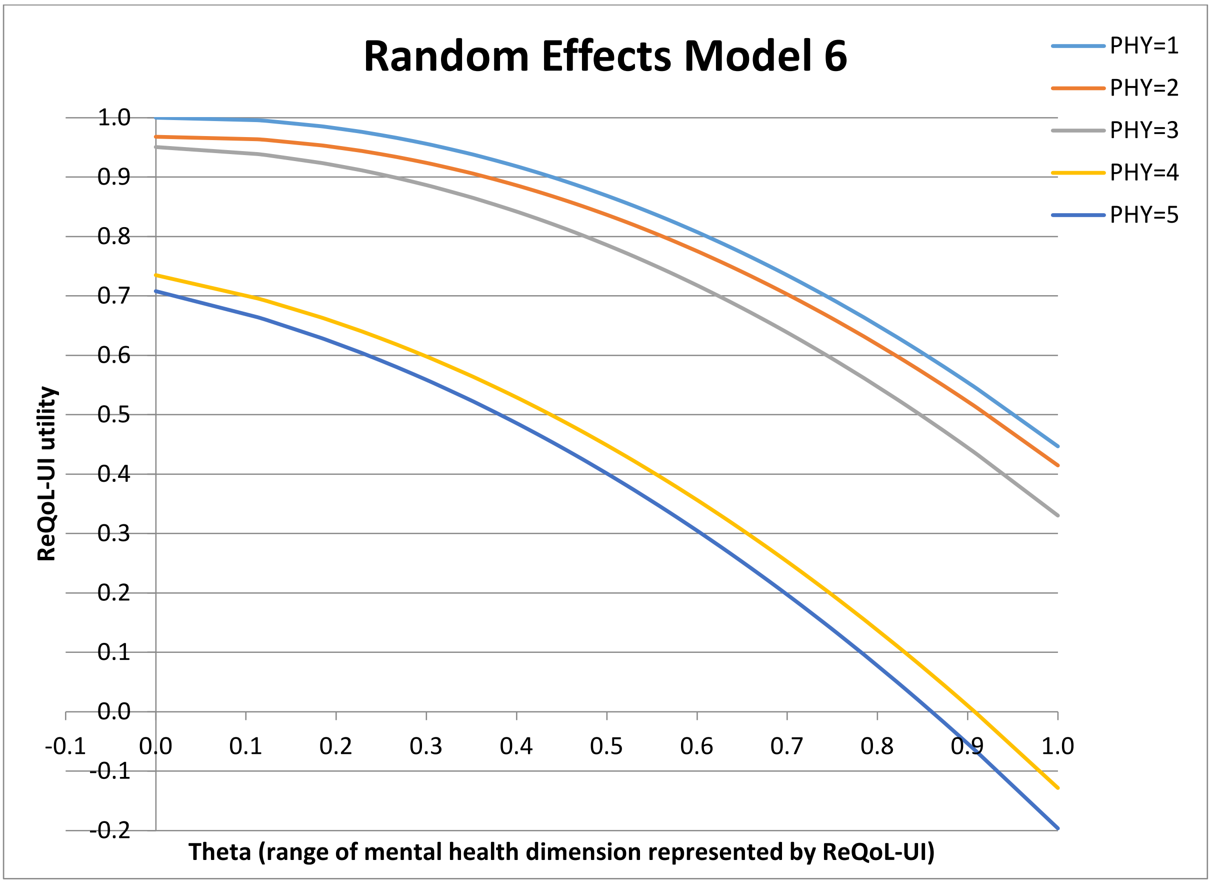

Table 4 Regression Results for Estimating Health Preference Scores

| | OLS Mean models | | | Random effects models | | | |
|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 (preferred) | Model 7 |
| | Linear with interactions | Quadratic with interactions | Quadratic with interactions, only significant coefficients | Linear with interactions | Quadratic with interactions | Quadratic with interactions | Quadratic with interactions, only significant coefficients |
| θ (newtheta) | -0.433*** | 0.01 | -0.053 | -0.441*** | 0.028 | 0.028 | -0.015 |
| θ2 (newthetasq) | | -0.572*** | -0.517*** | | -0.582*** | -0.581*** | -0.558*** |
| phy2 | 0.059 | -0.069 | | 0.089 | -0.033 | -0.032 | |
| phy3 | 0.001 | -0.073 | -0.084*** | 0.027 | -0.050 | -0.049 | -0.076*** |
| phy4 | -0.140** | -0.284*** | -0.270*** | -0.141*** | -0.265*** | -0.265*** | -0.261*** |
| phy5 | -0.189*** | -0.294*** | -0.284*** | -0.201*** | -0.292*** | -0.292*** | -0.288*** |
| inter2 | -0.099 | 0.066 | | -0.151 | 0.002 | | |
| inter3 | -0.135 | -0.037 | | -0.165 | -0.067 | -0.067 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| inter4 | -0.503*** | -0.292*** | -0.293*** | -0.492*** | -0.310*** | -0.310*** | -0.292*** |
| inter5 | -0.501*** | -0.362*** | -0.356*** | -0.465*** | -0.350*** | -0.351*** | -0.330*** |
| Constant[a] | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Observations | 64 | 64 | 64 | 2303 | 2303 | 2303 | 2303 |
| Adjusted R-squared | 0.974 | 0.982 | 0.982 | | | | |
| RMSE | 0.082 | 0.067 | 0.069 | 0.082 | 0.069 | 0.069 | 0.070 |
| MAE | 0.069 | 0.056 | 0.058 | 0.069 | 0.057 | 0.057 | 0.057 |
| AIC | -121 | -144 | -147 | 3451 | 3430 | 3428 | 3426 |
| BIC | -102 | -122 | -132 | 3514 | 3499 | 3492 | 3477 |
| Number of observations with AE >0.1 | 15 | 9 | 8 | 13 | 8 | 8 | 10 |
| Percentage of observations with AE >0.1 | 23% | 14% | 13% | 20% | 13% | 13% | 16% |
| Number of observations with AE >0.05 | 42 | 32 | 33 | 39 | 29 | 29 | 32 |
| Percentage of observations with AE >0.05 | 66% | 50% | 52% | 61% | 45% | 45% | 50% |

KEY: θ = IRT theta rescaled to 0 (best possible mental health score) and 1 (worst possible mental health scores);
[a] The models did not have a constant but a constant 1 is presented here so the coefficients can be presented as utility decrement.
AE – absolute error; AIC – Akaike Information Criterion; BIC – Bayesian Information Criterion; inter4 = θ * phy4 inter5 = θ * phy5 ; MAE – mean absolute error; phy2 - Level 2 physical health (phy1 – best physical health and phy5 – worst); phy3 - Level 3 physical health; phy4 - Level 4 physical health phy5 - Level 5 physical health; RMSE – root mean square error *** p<0.01, ** p<0.05, * p<0.1

## Discussion

We developed the ReQoL-UI health classification which comprises six MH and one PH item from ReQoL-10 (and ReQoL-20) and have produced a set of preference weights. An algorithm has been estimated to generate the ReQoL-UI scores, available in STATA, SPSS and Excel, using the predictions from the preferred RE model with the corresponding θ for all the possible combinations for the seven items. The preference weights enable utility values to be generated from the ReQoL measures for use in cost-effectiveness analyses across the full severity range of MH conditions. ReQoL-10 and ReQoL-20 were specifically developed with considerable input from service users and have high face and content validity (10, 11, 23). Therefore, the corresponding utilities are likely to be more appropriate for use to evaluate mental healthcare interventions than those generated from generic measures. Future research will empirically test the use of the ReQoL-UI in trials and studies including comparison with preference-based generic measures including EQ-5D and SF-6D to compare their relative psychometric performance.

Whilst MAE is higher than some other TTO valuation studies where error is in the region of 0.05 (24, 25), it is possible that this is due to the different model specification estimated here that does not differentiate between the different MH items in terms of their differential impact on utility. Both the mean OLS models and the RE models have good predictive ability across the range of utility values, with predictive performance lowest for all models for the very severe states. The poorer predictive performance for the models for the more severe states may have been observed due to the inconsistencies in the TTO utility values for some of the more severe states, where worst state had a similar but slightly higher mean TTO value than two other severe health states.

Unlike the EQ-5D, six of the items form a unidimensional component in the ReQoL-UI classification system related to MH, with one dimension for PH. From the regression results, the importance attributed to both PH and MH is clear. In the preferred RE model, over 50% of the utility decrement is attributed to the severity of the MH condition compared with 23% for the worst level (Level 3) of anxiety and depression in the EQ-5Q-3L preference weights (17). Nonetheless, the PH item has a large utility decrement of 0.29 for the most severe level. One key advantage of including the PH item is that a utility decrement is generated for PH as well as for mental health, and as our qualitative research showed PH should not be ignored for people with MH problems. In all the models, the signs of the interaction terms are negative and highly significant. This finding means that association between poor MH (theta) and low utility values is stronger when there is also a moderate to severe PH problem.

An added advantage of the PH item is that it minimise any focussing effects upon MH in the health states valued. One concern raised with condition-specific valuation is that if the classification system is focussed on a particular set of symptoms then respondents can exaggerate their impact on utility as these have not been placed within the context of other symptoms or more generic aspects of health (26, 27). However, in this study respondents considered PH problems alongside MH problems, which would have minimised any concerns of focussing effects arising due to the MH focus of the remaining items.

This paper provides an innovative use of IRT to select items and health states for a preference-based classification system. It improves the credibility of the states selected for valuation compared to the use of statistical designs like an orthogonal array that can generate states with unlikely combinations of levels across dimensions. Several papers have used item threshold based on Rasch analysis to construct such a health classification system (15, 16) but, to our knowledge, none has used analyses of response combination probabilities. This approach allowed us to choose the health states that are more likely to be observed in real life. We selected 59 MH states and this constitutes a clear advantage since the larger number of health states included in this valuation study provides for a more robust regression model compared to previous applications of this approach. We analysed response combination based on the graded response IRT model, but the approach could also be applied with the Rasch model. While the models are very similar, IRT models may fit a broader set of scales.

The ReQoL-UI can be used in cost effectiveness analyses to capture the utility impact of problems in MH. The use of condition-specific preference-based measures to inform policy is one that is debated, as many reimbursement agencies recommend the use of a generic preference-based measure (27). For example, in the UK, the National Institute for Health and Care Excellence (NICE) recommends the use of one particular measure, the EQ-5D for use in cost-effectiveness analyses for health technology assessment (28). However, alternative preference-based measures can be used in sensitivity analyses and where it can be evidenced that EQ-5D is not valid for the condition or patient population of interest. ReQoL-UI has the advantage over generic preference-based measures for use in people with MH problems that it was developed with considerable input from MH service users and has five items capturing mental health.

## Compliance with Ethical standards

*Disclosure of potential conflicts of interests*

The authors have no conflicts of interest to declare.

*Research involving Human Participants*

Ethical approval for Stages 1 and 2 was granted by the Edgbaston National Research Ethics Services committee, West Midlands (14/WM/1062). Ethical approval for the valuation survey was obtained from the School of Health and Related Research via the University of Sheffield Research Ethics Committee process. The reference number is: 009307.

Informed consent was obtained from all respondents in the study.

## References


1. Brazier J, Connell J, Papaioannou D, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technology Assessessment. 2014; 18
2. Papaioannou D, Brazier J, Parry G. How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in schizophrenia? A systematic review. Value Health. 2011; 14: 907-20.
3. Saarni SI, Viertiö S, Perälä J, et al. Quality of life of people with schizophrenia, bipolar disorder and other psychotic disorders. The British Journal of Psychiatry. 2010; 197: 386-94.
4. Mulhern B, Mukuria C, Barkham M, et al. Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. The British Journal of Psychiatry. 2014; 205: 236-43.
5. Keetharuth AD, Brazier J, Connell J, et al. Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. The British Journal of Psychiatry. 2018; 212: 42-49.
6. Connell J, Brazier J, O'Cathain A, et al. Quality of life of people with mental health problems: a synthesis of qualitative research. Health and Quality of Life Outcomes. 2012; 10: 138.
7. Connell J, O'Cathain A, Brazier J. Measuring quality of life in mental health: Are we asking the right questions? Soc Sci Med. 2014; 120: 12-20.
8. Keetharuth A.D B, M., Bjorner, J., Browne J. , Brazier J.,. An item reponse theory analysis of the Recovering Quality of Life (ReQoL). HEDS Discussion Paper 2018 University of Sheffield 2018 https://www.sheffield.ac.uk/scharr/sections/heds/discussion-papers.
9. Keetharuth AD, Bjorner JB, Barkham M, et al. Exploring the item sets of the Recovering Quality of Life (ReQoL) measures using factor analysis. Qual Life Res. 2019; 28: 1005-15.
10. Connell J, Carlton J, Grundy A, et al. The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). Qual Life Res. 2018; 27: 1893-902.
11. Keetharuth A.D., Taylor Buck E., Conway K., et al. Integrating Qualitative and Quantitative Data in the Development of Outcome Measures: The Case of the Recovering Quality of Life (ReQoL) Measures in Mental Health Populations. Int J Environ Res Public Health. 2018 15.
12. Samejima F. A general model for free-response data. Psychometrika Monograph Supplement. 1972.
13. Muthen LK, Muthen BO. Mplus User's Guide. Sixth ed. Los Angeles, CA: Muthen & Muthen., 1998-2011.
14. Cai L, Du Toit S, Thissen D. IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software]. Chicago, IL: Scientific Software International. 2017.
15. Mavranezouli I, Brazier, J. E., Young, T. A., & Barkham, M. . Using Rasch analysis to form plausible health states amenable to valuation: the development of CORE-6D from a measure of common mental health problems (CORE-OM). Qual Life Res. 2011; 20: 321-33.
16. Young TA, Rowen D, Norquist J, et al. Developing preference-based health measures: using Rasch analysis to generate health state values. 2010; 19: 907-17.
17. Dolan P. Modeling valuations for EuroQol health states. Medical care. 1997: 1095-108.
18. Devlin NJ, Tsuchiya A, Buckingham K, et al. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time'approach. Health Economics. 2011; 20: 348-61.
19. Mavranezouli I, Brazier JE, Rowen D, et al. Estimating a preference-based index from the Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM) valuation of CORE-6D. Med Decis Making. 2013; 33: 381-95.
20. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ. 2002; 21: 271-92.

21. Yang Y, Brazier JE, Tsuchiya A, et al. Estimating a preference-based index for a 5-dimensional health state classification for asthma derived from the asthma quality of life questionnaire. Medical Decision Making. 2011; 31: 281-91.
22. StataCorp S. Statistical software: release 14. College Station: StataCorp LP. 2015.
23. Grundy A, Keetharuth D, Barber R, et al. Public Involvement in health outcomes research: lessons learnt from the development of the Recovering Quality of Life (ReQoL) measures. Health and Quality of Life Outcomes. 2019; 17.
24. Rowen D, Brazier J, Young T, et al. Deriving a preference-based measure for cancer using the EORTC QLQ-C30. Value Health. 2011; 14: 721-31.
25. Rowen D, Mulhern B, Banerjee S, et al. Estimating preference-based single index measures for dementia using DEMQOL and DEMQOL-Proxy. Value Health. 2012; 15: 346-56.
26. Brazier J, Tsuchiya A. Preference-based condition-specific measures of health: what happens to cross programme comparability? Health economics. 2010; 19: 125-29.
27. Rowen D, Zouraq IA, Chevrou-Severac H, et al. International regulations and recommendations for utility data for health technology assessment. Pharmacoeconomics. 2017; 35: 11-19.
28. NICE. Guide to the Methods of Technology Appraisal 2013. London: National Institute for Health and Care Excellence, 2013.

APPENDIX

Appendix 1 ReQoL items

ReQoL-20 items

| | | Response options |
|---|---|---|
| 1. | I found it difficult to get started with everyday tasks * | 5 frequency-based options ranging from: none of the time, only occasionally, sometimes, often, most or all of the time. |
| 2. | I felt able to trust others * | |
| 3. | I felt unable to cope * | |
| 4. | I could do the things I wanted to do * | |
| 5. | I felt happy * | |
| 6. | I thought my life was  not worth living * | |
| 7. | I enjoyed what I did * | |
| 8. | I felt hopeful about my future * | |
| 9. | I felt lonely * | |
| 10. | I felt confident in myself * | |
| 11. | I did things I found rewarding | |
| 12. | I avoided things I needed to do | |
| 13. | I felt irritated | |
| 14. | I felt like a failure | |
| 15. | I felt in control of my life | |
| 16. | I felt terrified | |
| 17. | I felt anxious | |
| 18. | I had problems with my sleep | |
| 19. | I felt calm | |
| 20. | I found it hard to concentrate | |
| Physical health question | Please describe your physical health (problems with pain, mobility, difficulties caring for yourself or feeling physically unwell) over the last week | 5 severity-based options ranging from: no problems, slight problems, moderate problems, severe problems and very severe problems. |

*ReQoL -10 consists of the first ten items and ReQoL-20 is made up of the 20 items. Both versions contain the last physical item which is not included in the summative score.
Further details, sample copies, scoring guides and requests for permissions to use the ReQoL measures are available from: The Clinical Outcomes team at Oxford University Innovation Ltd at: [http://innovation.ox.ac.uk/outcome-measures/recovering-quality-life-reqol-questionnaire/](http://innovation.ox.ac.uk/outcome-measures/recovering-quality-life-reqol-questionnaire/)

Figure A1 Histogram of observed TTO values for ReQoL-UI

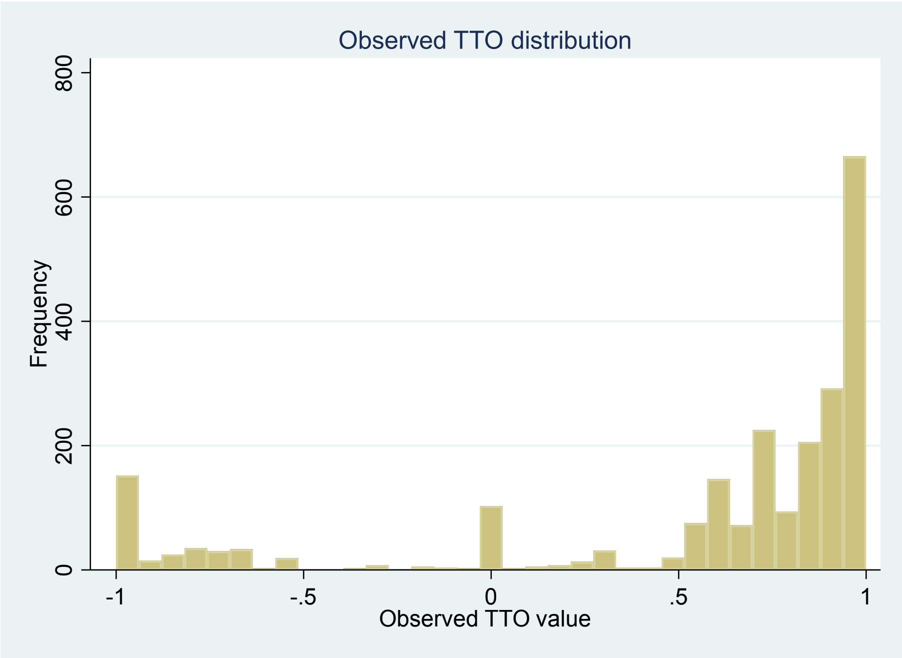


Figure A2 Plot of predicted versus observed utility values for the best mean OLS model (Model

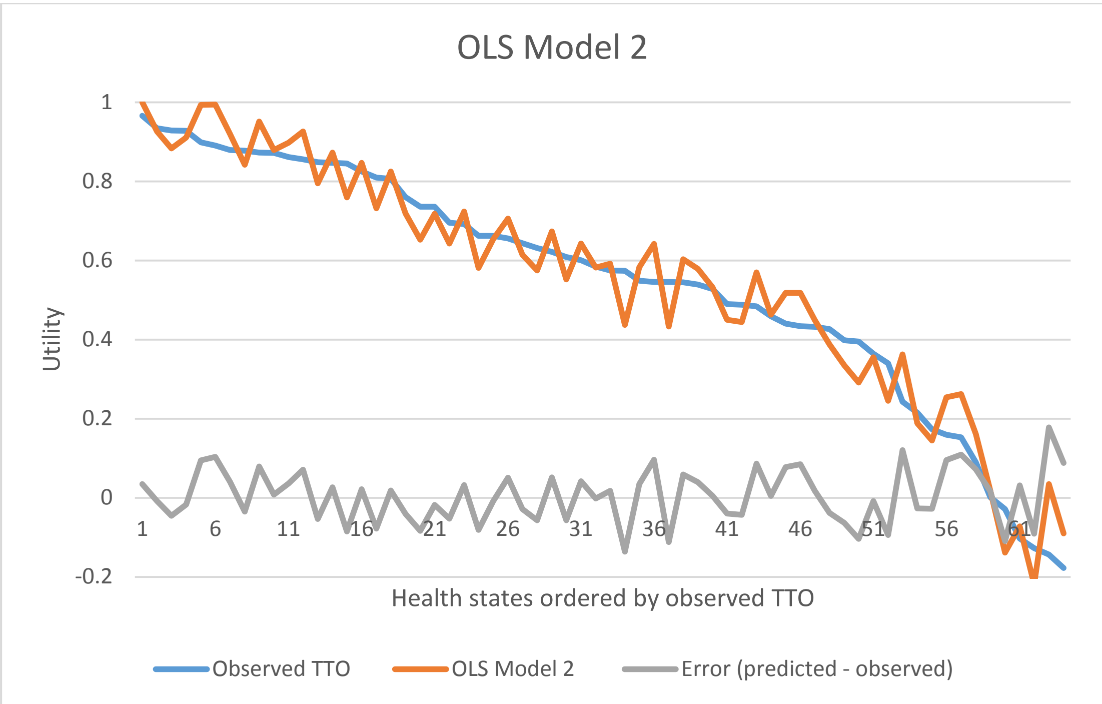

**Table A1 Random effects linear models (no constant)**

RE models without constant; Dependent variable : distto , independent variable newtheta

| VARIABLES | Model1a_NC | Model1b_NC | Model1c_NC | Model1d_NC | Model1e_NC | Model1f_NC |
|---|---|---|---|---|---|---|
| θ (newtheta) | -0.619*** | -0.593*** | -0.602*** | -0.441*** | -0.455*** | -0.472*** |
| phy2 | 0.0783** | | | 0.0889 | | |
| phy3 | 0.00232 | -0.0262 | | 0.0266 | 0.0181 | -0.0712** |
| phy4 | -0.335*** | -0.357*** | -0.349*** | -0.141*** | -0.148*** | -0.148*** |
| phy5 | -0.398*** | -0.424*** | -0.416*** | -0.201*** | -0.205*** | -0.208*** |
| inter2 | | | | -0.151 | | |
| inter3 | | | | -0.165 | -0.147 | |
| inter4 | | | | -0.492*** | -0.478*** | -0.469*** |
| inter5 | | | | -0.465*** | -0.454*** | -0.440*** |
| Constant | 1 | 1 | 1 | 1 | 1 | 1 |
| Observations | 2303 | 2303 | 2303 | 2303 | 2303 | 2303 |
| RMSE | 0.1143 | 0.1211 | 0.1207 | 0.0822 | 0.0828 | 0.0843 |
| MAE | 0.0945 | 0.1037 | 0.1034 | 0.0693 | 0.0712 | 0.0730 |
| AIC | 3486 | 3491 | 3490 | 3451 | 3449 | 3449 |
| BIC | 3527 | 3525 | 3518 | 3514 | 3500 | 3495 |
| Number of health states with errors >0.1 (abs) | 29 | 33 | 32 | 13 | 14 | 18 |
| Percentage of health states with errors >0.1 (abs) | 45% | 52% | 50% | 20% | 22% | 28% |
| Number of health states with errors >0.05 (abs) | 46 | 49 | 50 | 39 | 42 | 45 |
| Percentage of health states with errors >0.05 (abs) | 72% | 77% | 78% | 61% | 66% | 70% |
| Min utility for the 64 states | -0.017 | -0.016 | -0.019 | -0.108 | -0.114 | -0.12 |
| Max utility for the 64 states | 1.002 | 1 | 1 | 1.027 | 1.018 | 1 |

**Table A2 Random Effect quadratic models with no constant**

RE models without constant

Dependent variable : distto , independent variable newtheta

| | Quadratic models | | | | | | |
|---|---|---|---|---|---|---|---|
| VARIABLES | Model2a_NC | Model2b_NC | Model2c_NC | Model2d_NC | Model2e_NC | Model2f_NC | Model2g_NC |
| θ (newtheta) | 0.0701 | 0.0612 | 0.0280 | 0.0278 | 0.0278 | 0.0148 | |
| θ2 (newthetasq) | -0.747*** | -0.739*** | -0.582*** | -0.581*** | -0.591*** | -0.558*** | -0.574*** |
| phy2 | -0.00599 | | -0.0330 | -0.0321 | -0.0306 | | |
| phy3 | -0.0533* | -0.0508* | -0.0495 | -0.0494 | -0.0895*** | -0.0762*** | -0.0770*** |
| phy4 | -0.402*** | -0.399*** | -0.265*** | -0.265*** | -0.267*** | -0.261*** | -0.264*** |
| phy5 | -0.463*** | -0.461*** | -0.292*** | -0.292*** | -0.294*** | -0.288*** | -0.290*** |
| inter2 | | | 0.00171 | | | | |
| inter3 | | | -0.0668 | -0.0671 | | | |
| inter4 | | | -0.310*** | -0.310*** | -0.302*** | -0.292*** | -0.288*** |
| inter5 | | | -0.350*** | -0.351*** | -0.341*** | -0.330*** | -0.329*** |
| Constant | 1 | 1 | | 1 | 1 | 1 | 1 |
| Observations | 2303 | 2303 | 2303 | 2303 | 2303 | 2303 | 2303 |
| RMSE | 0.0834 | 0.0835 | 0.0688 | 0.0688 | 0.0687 | 0.0697 | 0.0700 |
| MAE | 0.0648 | 0.0649 | 0.0566 | 0.0566 | 0.0567 | 0.0574 | 0.0576 |
| AIC | 3443 | 3442 | 3430 | 3428 | 3427 | 3426 | 3424 |
| BIC | 3489 | 3482 | 3499 | 3492 | 3484 | 3477 | 3470 |
| Number of health states with errors >0.1 (abs) | 14 | 14 | 8 | 8 | 10 | 10 | 9 |
| Percentage of health states with errors >0.1 (abs) | 22% | 22% | 13% | 13% | 16% | 16% | 14% |
| Number of health states with errors >0.05 (abs) | 32 | 33 | 29 | 29 | 30 | 32 | 32 |
| Percentage of health states with errors >0.05 (abs) | 50% | 52% | 45% | 45% | 47% | 50% | 50% |
| Min utility for the 64 states | -0.14 | -0.139 | -0.196 | -0.195 | -0.199 | -0.191 | -0.193 |
| max utility for the 64 states | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

**Table A3 Random Effect cubic models with no constant**

Dependent variable : distto , independent variable newtheta

| VARIABLES | Model3a_NC | Model3b_NC | Model3c_NC | Model3d_NC | Model3e_NC | Model3f_NC | Model3g_NC |
|---|---|---|---|---|---|---|---|
| θ (newtheta) | 0.399* | 0.364 | 0.280 | 0.217 | 0.227 | 0.147 | |
| θ2 (newthetasq) | -1.688*** | -1.627*** | -1.304* | -1.122* | -1.158* | -1.030 | -0.647*** |
| θ3 (newthetacube) | 0.658 | 0.629 | 0.486 | -0.374 | 0.392 | 0.329 | 0.0809 |
| phy2 | -0.0142 | | -0.0689 | -0.0359 | -0.0347 | | |
| phy3 | -0.0514* | -0.0459 | -0.0567 | -0.0515 | -0.0873*** | -0.0729** | -0.0741*** |
| phy4 | -0.418*** | -0.412*** | -0.294*** | -0.282*** | -0.285*** | -0.275*** | -0.260*** |
| phy5 | -0.468*** | -0.463*** | -0.303*** | -0.297*** | -0.300*** | -0.292*** | -0.286*** |
| inter2 | | | 0.0629 | | | | |
| inter3 | | | -0.0455 | -0.0602 | | | |
| inter4 | | | -0.272** | -0.293*** | -0.285*** | -0.276*** | -0.293*** |
| inter5 | | | -0.330*** | -0.344*** | -0.335*** | -0.324*** | -0.331*** |
| Constant | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Observations | 2303 | 2303 | 2303 | 2303 | 2303 | 2303 | 2303 |
| RMSE | 0.0824 | 0.0827 | 0.0686 | 0.0692 | 0.0690 | 0.0702 | 0.0693 |
| MAE | 0.0649 | 0.0650 | 0.0568 | 0.0576 | 0.0576 | 0.0584 | 0.0574 |
| AIC | 3443 | 3441 | 3431 | 3430 | 3428 | 3427 | 3425 |
| BIC | 3495 | 3487 | 3506 | 3499 | 3491 | 3484 | 3477 |
| Number of health states with errors >0.1 (abs) | 17 | 16 | 10 | 10 | 11 | 11 | 10 |
| Percentage of health states with errors >0.1 (abs) | 27% | 25% | 16% | 16% | 17% | 17% | 16% |
| Number of health states with errors >0.05 (abs) | 36 | 33 | 29 | 30 | 31 | 32 | 33 |
| Percentage of health states with errors >0.05 (abs) | 56% | 52% | 45% | 47% | 48% | 50% | 52% |
| Min utility for the 64 states | -0.098 | -0.097 | -0.17 | -0.172 | -0.175 | -0.17 | -0.183 |
| max utility for the 64 states | 1.025 | 1.021 | 1.016 | 1.011 | 1.011 | 1.004 | 1 |

Table A4 OLS models

| VARIABLES | Model1_NC | Model 1c_NC | Model 2_NC | Model 2c_NC | Model2e_NC | Model 2f_NC | Model 3_NC | Model 3a_NC | Model 3b_NC | Model 3c_NC | Model 3d_NC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| θ (newtheta) | -0.579*** | -0.433*** | 0.100 | 0.00978 | -0.052932 | | 0.383 | 0.251 | 0.122 | 0.025 | -0.428* |
| θ2 (newthetasq) | | | -0.767*** | -0.572*** | -0.5165521*** | -0.5795743*** | -1.605** | -1.275* | -0.909 | -0.753 | 0.339 |
| θ3 (newthetacube) | | | | | | | 0.594 | 0.474 | 0.249 | 0.171 | -0.440 |
| phy2 | 0.085** | 0.059 | -0.014 | -0.0688 | | | -0.022 | -0.103 | -0.038 | | |
| phy3 | 0.009 | 0.001 | -0.062* | -0.0737 | -0.0836655*** | -0.0876205*** | -0.062* | -0.083 | -0.074 | -0.067 | -0.031 |
| phy4 | -0.323*** | -0.140** | -0.415*** | -0.284*** | -0.2701098*** | -0.2859731*** | -0.430*** | -0.312*** | -0.291*** | -0.276*** | |
| phy5 | -0.394*** | -0.189*** | -0.476*** | -0.294*** | -0.2839871*** | -0.2955383*** | -0.482*** | -0.307*** | -0.295*** | -0.286*** | -0.234*** |
| inter2 | | -0.099 | | 0.0655 | | | | 0.123 | | | |
| inter3 | | -0.135 | | -0.0365 | | | | -0.013 | -0.040 | -0.025 | -0.089 |
| inter4 | | -0.503*** | | -0.292*** | -0.2931935*** | -0.2741532*** | | -0.254** | -0.295*** | -0.289*** | -0.705*** |
| inter5 | | -0.501*** | | -0.362*** | -0.3559337*** | -0.3447323*** | | -0.339*** | -0.368*** | -0.355*** | -0.432*** |
| Constant | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Observations | 64 | 64 | 64 | 64 | 64 | 64 | 64 | 64 | 64 | 64 | 64 |
| Adjusted R-squared | 0.955 | 0.974 | 0.974 | 0.982 | 0.982 | 0.982 | 0.974 | 0.982 | 0.982 | 0.982 | 0.974 |
| RMSE | 0.1103 | 0.0817 | 0.0833 | 0.0673 | 0.0685 | 0.0688 | 0.0821 | 0.0666 | 0.0673 | 0.0683 | 0.0813 |
| MAE | 0.0911 | 0.0690 | 0.0651 | 0.0561 | 0.0575 | 0.0575 | 0.0656 | 0.0559 | 0.0570 | 0.0579 | 0.0641 |
| AIC | -91 | -121 | -125 | -144 | -147 | -149 | -124 | -143 | -144 | -144 | -124 |
| BIC | -80 | -102 | -112 | -122 | -132 | -136 | -109 | -119 | -122 | -124 | -106 |
| Number of observations with AE >0.1 | 26 | 15 | 13 | 9 | 8 | 8 | 16 | 10 | 10 | 11 | 14 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Percentage of observations with AE >0.1 | 41% | 23% | 20% | 14% | 13% | 13% | 25% | 16% | 16% | 17% | 22% |
| Number of observations with AE >0.05 | 42 | 42 | 32 | 32 | 33 | 32 | 36 | 29 | 31 | 33 | 36 |
| Percentage of observations with AE >0.06 | 66% | 66% | 50% | 50% | 52% | 50% | 56% | 45% | 48% | 52% | 56% |
| Min utility for the 64 states | 0.027 | -0.124 | -0.143 | -0.219 | -0.209 | -0.22 | -0.111 | -0.196 | -0.202 | -0.199 | -0.235 |
| max utility for the 64 states | 1.009 | 1.001 | 1.002 | 1 | 1 | 1 | 1.024 | 1.013 | 1.003 | 1 | 1 |